MONASH University

# Early developmental biology of the spiny mouse (*Acomys cahirinus*)

by
Jared Mamrot
BSc(Hons), MClinEmbryology

Supervised by
Dr. Hayley Dickinson
Professor David W. Walker
Professor David K. Gardner
Dr. Mulyoto Pangestu

at
The Ritchie Centre
Hudson Institute of Medical Research

A thesis submitted for the degree of Doctor of Philosophy at
Monash University in 2018
Faculty of Medicine, Nursing and Health Sciences

## Copyright notice

# Table of contents

**Abstract** ..... **5**
**Declaration** ..... **6**
**Publications during enrolment** ..... **7**
**Accepted conference abstracts** ..... **7**
**Thesis including published works declaration** ..... **8**
**Acknowledgements** ..... **10**
**Literature review** ..... **12**
**Introduction** ..... **12**
**Embryogenesis** ..... **13**
Fertilisation ..... 16
The oocyte ..... 17
The zona pellucida ..... 19
Syngamy ..... 21
Embryo division ..... 21
**Epigenetics and the maternal-to-zygote transition** ..... **24**
Histone modification and DNA methylation ..... 25
DNA demethylation ..... 27
Imprinting ..... 29
**Embryonic genome activation** ..... **31**
Recruitment of maternal transcripts ..... 32
Degradation of maternal transcripts ..... 34
Activation of the zygotic genome ..... 35
Markers of activation ..... 37
Conventional investigative approaches ..... 37
Modern investigative approaches ..... 38
Modelling the EGA ..... 39
**The spiny mouse (Acomys cahirinus)** ..... **39**
Ecology and Phylogeny ..... 40
Divergent physiology from typical rodents ..... 41
Embryogenesis ..... 43
**References** ..... **46**
**Chapter 1: Methods development** ..... **69**
**1.1 Overview** ..... **69**
**1.2 Rationale** ..... **69**
**1.3 Aims and hypotheses** ..... **79**
**1.4 Methods** ..... **80**
1.4.1 Animal housing and ethics ..... 80
1.4.2 Embryo collection and culture ..... 80
1.4.3 ICM transfer ..... 82

1.4.4 Immunofluorochemistry ..... 88
1.4.5 Embryo transfer ..... 88
1.4.6 Vasectomy in the spiny mouse ..... 89
1.4.7 Pseudopregnancy induction in the spiny mouse ..... 90
**1.5 Results ..... 91**
1.5.1 Embryo collection and culture ..... 91
1.5.2 ICM transfer ..... 93
1.5.3 Embryo Immunofluorochemistry ..... 95
1.5.4 Embryo transfer ..... 97
1.5.5 Vasectomy in the spiny mouse ..... 98
1.5.6 Pseudopregnancy induction in the spiny mouse ..... 99
**1.6 Discussion ..... 100**
**1.7 References ..... 103**

**Chapter 2: Embryo transfer ..... 113**
**2.1 Overview ..... 113**

**Chapter 3: Transcriptome assembly ..... 120**
**3.1 Overview ..... 120**
**3.2 Supplementary figures ..... 136**
**3.3 Detailed protocol ..... 139**

**Chapter 4: Embryonic genome activation ..... 169**
**4.1 Overview ..... 169**
**4.2 Supplementary figures ..... 189**

**General Discussion ..... 202**
**Embryology ..... 202**
**Genetics ..... 206**
**Limitations ..... 208**
**Future directions ..... 209**
**Conclusion ..... 209**
**References ..... 211**

## Abstract

The spiny mouse (*Acomys cahirinus*) is a species of rodent used in medical research. Displaying rare and unique physiological characteristics it provides advantages over the common mouse for modelling various aspects of human physiology. The aim of this PhD project was to investigate the utility of the spiny mouse for modelling early human development. This thesis has two primary themes: novel characterization of spiny mouse anatomy and physiology, and development of novel methods to evaluate the spiny mouse for use as an animal model. Each chapter addresses one or both of these themes, describing (1) the development of a novel methods to generate interspecies and intersex inner cell mass (ICM) / trophectoderm (TE) chimeric embryos, (2) development of methods to increase the efficacy of embryo transfer with a view to implement this technique in the spiny mouse, (3) the assembly of a reference transcriptome for the spiny mouse, and (4) characterization of gene transcription in the spiny mouse embryo and comparison of transcription profiles between mouse, spiny mouse and human embryos. The studies outlined in this thesis provide a foundation for evaluating the utility of the spiny mouse as an animal model of early human development.

## Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature: 

Print Name: Jared Mamrot

Date: 10/03/2018

## Publications during enrolment

**Mamrot, J.**, Legaie, R., Ellery, S. J., Wilson, T., Seemann, T., Powell, D.R., Gardner, D.K., Walker, D.W., Temple-Smith, P. & Dickinson, H. (2017). De novo transcriptome assembly for the spiny mouse (Acomys cahirinus). *Scientific Reports*, *7*(1), 8996.

Bellofiore, N., Ellery, S. J., **Mamrot, J.**, Walker, D. W., Temple-Smith, P., & Dickinson, H. (2017). First evidence of a menstruating rodent: the spiny mouse (Acomys cahirinus). *American Journal of Obstetrics & Gynecology*, *216*(1), 40-e1.

**Mamrot, J.**, Pangestu, M., Walker, D., Gardner, D. K., & Dickinson, H. (2015). Confirmed dioestrus in pseudopregnant mice using vaginal exfoliative cytology improves embryo transfer implantation rate. *Reproductive biomedicine online*, *31*(4), 538-543.

## Accepted conference abstracts

**Mamrot, J.**, Mamrot, S., Weybury, M., Wiadrowski, L., Temple-Smith, P., Dickinson, H. (2017) Anatomy and physiology of the male reproductive tract in the spiny mouse (*Acomys cahirinus*) across the lifespan. ESA/SRB 2017, Perth, Australia.

**Mamrot, J.**, Papenfuss, T., Legaie, R., Ellery, S., Gardner, D. K., Walker, D., Temple-Smith, P. & Dickinson, H. (2016). Draft *de novo* transcriptome assembly for the spiny mouse (*Acomys cahirinus*). ASMR 2016, Melbourne, Australia.

Bellofiore, N., Ellery, S.J., **Mamrot, J.**, Walker, D.W., Temple-Smith, P. & Dickinson, H. (2016) First evidence of a menstruating rodent: the spiny mouse (*Acomys cahirinus*). SRF 2016, Winchester, UK.

**Mamrot, J.**, Pangestu, M., Walker, D., Gardner, D. K., & Dickinson, H. (2015) Quantitative assessment of uterine receptivity prior to embryo transfer increases implantation rate to >95%. ESA/SRB 2015, Adelaide, Australia.

Makanji, Y., Al-Muhana, D., **Mamrot, J.**, Temple-Smith, P., Harrison, C. A., Walker, D., & Dickinson, H. (2014) Characterisation of Spiny Mouse (*Acomys cahirinus*) Folliculogenesis Using a Three-Dimensional *In Vitro* Culture Model. ENDO 2014, Chicago, USA.

## Thesis including published works declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes 2 original manuscripts published in peer-reviewed journals and 1 submitted publication. The core theme of the thesis is the developmental biology of the spiny mouse (*Acomys cahirinus*). The ideas, development and writing up of all papers in the thesis were the principal responsibility of myself, the student, working within the Faculty of Medicine, Nursing and Health Sciences under the supervision of Hayley Dickinson, Mulyoto Pangestu, David Gardner and David Walker.

In the case of Chapters 2, 3 and 4 my contribution to the work involved the following:

| **Thesis Chapter** | **Publication Title** | **Status** *(published, in press, accepted or returned for revision, submitted)* | **Nature and % of student contribution** | **Co-author name(s) Nature and % of Co-author's contribution*** | **Co-author(s), Monash student Y/N*** |
|---|---|---|---|---|---|
| 2 | Confirmed dioestrus in pseudopregnant mice using vaginal exfoliative cytology improves embryo transfer implantation rate. | Accepted | 80%. Conceived, designed and conducted the experiment and analysis. Prepared all figures and drafted the manuscript. | David K. Gardner, Mulyoto Pangestu, David W. Walker, Hayley Dickinson.<br><br>Conceptualisation, experimental design, supervisory direction, funding, and contributions to the drafted manuscript (20%) | No |
| 3 | De novo transcriptome assembly for the spiny mouse (Acomys cahirinus). | Accepted | 80%. Designed and conducted the assembly, analysis and validation, prepared all | Roxane Legaie, Stacey J. Ellery, Trevor Wilson, Torsten Seemann, David R. Powell, David K. Gardner, David W. Walker, | No |

| | | | | | |
|---|---|---|---|---|---|
| | | | figures and drafted the manuscript. | Peter Temple-Smith, Anthony T. Papenfuss & Hayley Dickinson.<br><br>Conceptualisation and experimental design, supervisory direction, RNA sequencing, funding, and contributions to the drafted manuscript (20%) | |
| 4 | Gene transcription during early embryonic development in the spiny mouse (Acomys cahirinus) | Submitted | 80%. Conceived and designed the experiment. Collected and prepared samples, conducted the analysis and validation, prepared all figures and wrote the manuscript. | David K. Gardner, Peter Temple-Smith, Hayley Dickinson<br><br>Conceptualisation and experimental design, supervisory direction, funding, and contributions to the drafted manuscript (20%) | No |

I have renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

**Student signature:** [redacted] **Date:** 10/03/2018

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

**Main Supervisor signature:** [redacted] **Date:** 10/03/2018

## Acknowledgements

Thank you Hayley for your support throughout my PhD. I am incredibly grateful for your advice, guidance, effort and kindness, and for teaching me about the good and the bad aspects of a career in science. Thank you for believing in me and for being a wonderfully positive influence. Thank you for providing me with opportunities that have led to a career I love and for being an important part of my life the last few years. I can't wait to see the amazing things you will achieve in the future. Thank you David Walker, David Gardner, and Mulyoto Pangestu for co-supervising my PhD. I truly appreciate your time and effort and all the help you provided me. Thank you Tony Papenfuss for the opportunity to work in your extraordinary bioinformatics lab. Thank you Lois Salamanson, Justin St John, and Eva Dimitriadis for your guidance and input and for not discontinuing my candidature at each of my review milestones. Thank you to past and present members of the Dickinson Embryology and Placental Biology lab group, the Walker group and the Gardner group who assisted me along the way, and to members of the MHTP medical genomics sequencing facility for your contributions, especially Roxane Legaie for teaching me what the command line was and opening up a world of possibilities. Thank you Sally Catt and Peter Temple-Smith for your mentorship, for the use of your lab space and equipment, and for hiring me as a marker / demonstrator in EPRD. Thank you to the Masters and GRS students who chose me to supervise your minor research projects. They were always challenging, and things rarely went according to plan, but overall it was an incredibly positive experience and I very much enjoyed the role of 'teacher'. Thank you to the

Monash University Faculty of Medicine, Nursing and Health Studies for supporting me with a scholarship, and to the ARC and NHMRC for funding our research.

Most importantly, thank you to my family for your love and support (and a special mention for Aunt Sophie for her many significant contributions over the years). Lastly, and most deserving, thank you to my amazing wife Sarah: I love you with all my heart, forever and always. Thank you for supporting me through this incredibly challenging journey. This thesis is dedicated to you.

# Literature review

## Introduction

Development in all living organisms is regulated by DNA, and propagation of a species requires the successful transmission of DNA from parents to progeny. Plants and animals have evolved remarkably diverse reproductive strategies to facilitate this transfer, with varying degrees of parental control over genetic inheritance. Following the fusion of sperm and egg, early development is predominantly under maternal control, with maternal mRNA, proteins, and other molecules inherited by the oocyte governing survival during the early stages of embryogenesis (Paynton et al., 1988; Zheng et al., 2010). Maternal control subsides as development proceeds, with embryo survival increasingly reliant on transcription of the incipient embryonic genome; this process is termed the ‘maternal to zygotic transition’ (MZT) and it marks the beginning of autonomic control of development in offspring (Harrison & Eisen, 2015; Schier, 2007; Tadros & Lipshitz, 2009).

Complex biological mechanisms drive this transition. Factors such as chromatin state (Flyamer et al., 2017; Santos et al., 2002; Smith et al., 2012), intracellular calcium signaling (Ducibella et al., 2006; O’Neill et al., 2012, 2015) and mRNA methylation (Zhao et al., 2017) promote degradation and clearance of inherited maternal products, and facilitate transcription of embryonic genes. Our understanding of events implicated in the MZT has evolved over time, with technological advances repeatedly providing new insights into the mechanisms regulating this ephemeral process. The onset of gene transcription has historically been termed the 'embryonic genome activation' (EGA) due to detection of a major burst of transcription at the 2-cell stage in mice (Knowland & Graham, 1972; Sawicki et al., 1981) and 4-cell stage in humans (Braude et al., 1988). The development and

application of scientific techniques with greater sensitivity have since revealed the genome activation begins earlier, at the 1-cell 'zygote' stage, and the process of genome activation in the embryo is now commonly referred to as the 'zygotic genome activation' (ZGA) (Aoki et al., 1997; Latham et al., 1992).

Research into the MZT and the ZGA/EGA has traditionally been conducted on model organisms such as plants (*Arabidopsis*), nematodes (*C. elegans*), zebrafish (*Danio rerio*) and mice (*Mus musculus*), as human embryos are a valuable and scarce resource. Comparisons between animal/plant models and humans have revealed similarities in the pattern of early development, especially during major physiological events required for life such as fertilisation, however important species-specific differences have also been identified, limiting translation of findings from traditional animal models to humans in many instances. The search for a more appropriate model of early human development is ongoing.

## Embryogenesis

Life in animals and plants begins with a single cell. This cell is created when male and female gametes fuse, inheriting nuclei from both parents, and cytoplasm from the egg. Within this newly formed cell, haploid chromosomes are paired to form a new genome in a process termed syngamy. This marks the creation of a genetically-independent organism, and comprises the first stage of embryogenesis. Interestingly, embryogenesis has evolved independently in animals and flowering plants, with striking similarities between these evolutionarily divergent kingdoms (Meyerowitz, 2002). Common to all animals and plants, reprogramming of the newly formed zygote into totipotent cells is necessary for continued development, with complex processes such as cell proliferation, morphogenesis, and

organogenesis culminating in a diverse array of complex organisms. This reprogramming event has been exploited for several scientific applications, such as genetic modifications, cloning, and stem cell therapies.

Our understanding of this process is largely derived from animal and plant models, with the majority of available literature describing embryogenesis in traditional model organisms such as *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Xenopus laevis* and *Mus musculus*. The MZT has also been investigated in human embryos, but in a more limited capacity. Typically, research using human embryos has been conducted with the primary aim of characterising the process, rather than investigating the underlying mechanisms, due to the relative scarcity and ethical significance of these cells. These data have enabled comparison of the pattern of embryogenesis between humans and other organisms, with mammalian models providing the closest approximation of human embryogenesis and early development Figure 1). This section will focus on several key events in early embryogenesis, which begin with the process of fertilisation.

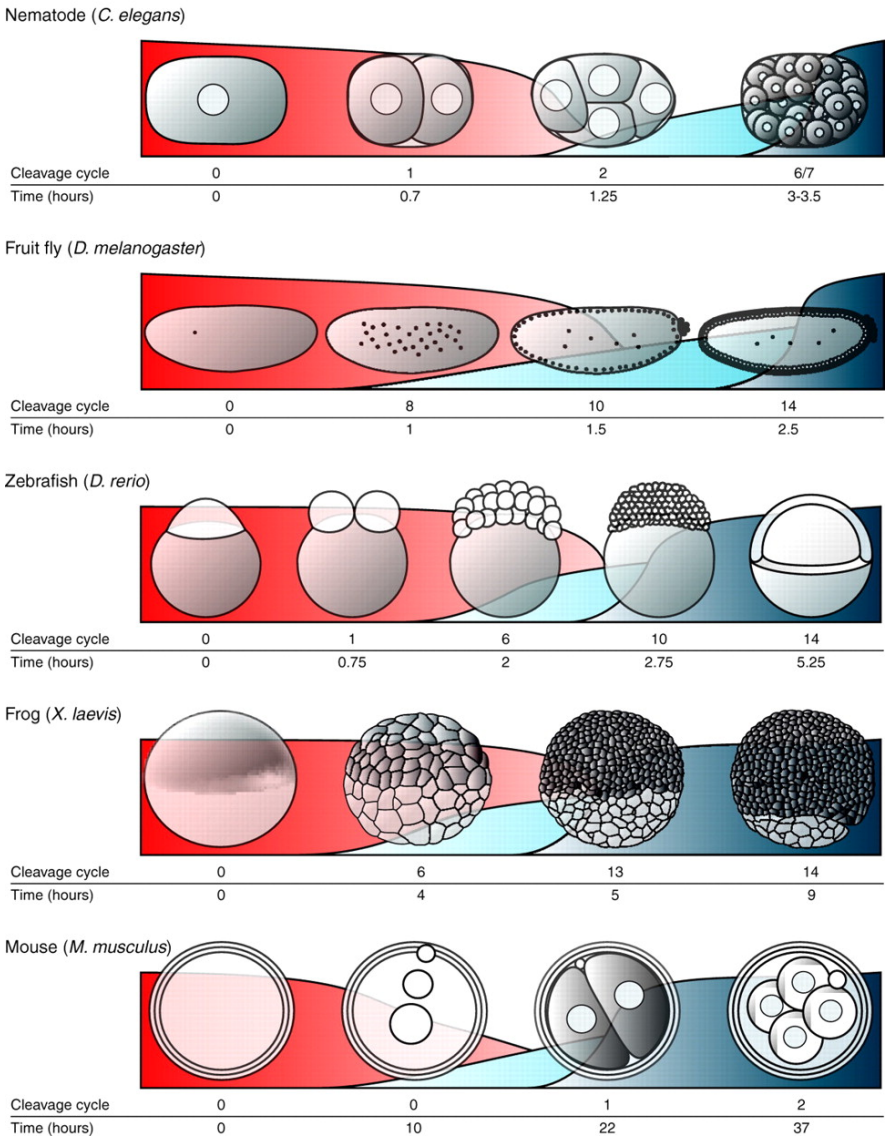



Figure 1: A comparative overview of the maternal-to-zygotic transition (MZT) in several model organisms. Key embryonic stages for each model organism are depicted schematically above the corresponding cleavage cycle and time after fertilization. The red curves represent the degradation profiles of destabilized maternal transcripts in each species. The light and dark blue curves illustrate the minor and major waves, respectively, of zygotic genome activation. From Tadros W and Lipshitz HD, "The maternal-to-zygotic transition: a play in two acts". Development, 136(18): 3033-3042. Copyright © 2009 by The Company of Biologists Ltd. Reprinted with permission from the copyright holder.

## Fertilisation

Fertilisation is defined as the union of mature male and female gametes, which results in the creation of genetically distinct offspring. This fundamental event is governed by a plethora of intercellular and intracellular mechanisms, which occur in a species-specific sequence of physiological processes. The process of fertilisation can differ strikingly between species, for instance it can take place internally (e.g. mammals) or externally (e.g. amphibians), and some plant species have the capacity for self-fertilisation (e.g. *Arabidopsis*), however many of the underlying cellular mechanisms are conserved in nature (Busa & Nuccitelli, 1985; Lord & Russell, 2002; Tang et al., 2007). Common physiological events occurring in all sexually reproducing species include the maturation and binding of gametes, and the successful establishment of a new genome. The significance of these events in early development, and the essential role fertilisation plays in propagating a species, has been recognized in the scientific literature for over 140 years.

The first accurate scientific report of fertilisation in mammals was published in 1875 by Édouard Van Beneden (Alexandre, 2001). Commonly known for discovering meiosis (Van Beneden, 1883; reviewed in Hamoir, 1992), Van Beneden asserted that contact between sperm and ova was required for diffusion of the 'spermatic component' through the membrane, forming a zygote with equal inheritance from haploid gametes. This finding was derived from earlier research conducted using nematodes (*Ascaris megalocephala*), and from observations made in his detailed monograph on the physiology of the mammalian oocyte (Van Beneden, 1870). Understanding the structure and function of the oocyte was, and still remains, necessary for understanding the process of fertilisation, and the first stages of embryogenesis.

## The oocyte

First identified in 1827 in a publication by Karl Ernst von Baer (Von Baer, 1827), the mammalian oocyte is a highly differentiated cell type defined by several distinguishing characteristics (Telfer & McLaughlin, 2007). These include the ability to undergo meiosis, the presence of DNA- / RNA-binding proteins H1foo (Izzo et al., 2008; Tanaka et al., 2001) and Msy2 (Yu et al., 2003), and genes such as *Zp1*, *Zp2*, and *Zp3*, which encode proteins for oocyte-specific functions (Lefievre, 2004). Within the ovary, immature oocytes are predominantly generated *in utero* (and during the perinatal period in altricial species, such as mice and rats), and then lie dormant in the diplotene stage of the first meiotic prophase, surrounded by a thin layer of flattened non-proliferating somatic cells; together these 'balls' of cells are termed primordial follicles, and mammals are typically born with a complete complement of up to several million per ovary (Edson et al., 2009). Development of primordial follicles resumes after puberty, with a small number of follicles periodically recruited to enter the growth phase.

Recruitment is characterised by proliferation of the somatic granulosa cells, with follicular growth categorised into distinct phases: primary, secondary, small antral cavity, and large antral cavity (McGee & Hsueh, 2000). These phases can be demarcated morphologically (Adams & Hertig, 1964), and at the molecular level, with distinct transcriptomic profiles reported for each stage (Pan et al., 2006). As follicles grow they become transcriptionally quiescent, with maternally inherited transcripts exceptionally stable during the growth phase, exhibiting half-lives of ~10 days, compared to hours, or minutes, in somatic cells (Brower et al., 1981; De Leon et al., 1983; Jahn et al., 1976; Svoboda et al., 2015). This unique characteristic is afforded by the RNA-binding protein Msy2 (Yu, 2001; Yu et al., 2003), however the eventual acquisition of meiotic and developmental competence results

in a ‘switch’ from mRNA stability in the oocyte to mRNA instability in the embryo, with extensive degradation of maternal transcripts during the MZT (Svoboda et al., 2015; Zhao et al., 2017). The ultimate objective for follicle development is for the oocyte to acquire the capacity for fertilisation and meiotic cleavage, and to reach the stage of ovulation, where the oocyte is released from the follicle into the reproductive tract for insemination (Figure 2A) (McGee & Hsueh, 2000). The union of oocyte and sperm at this juncture marks the initiation of embryogenesis.

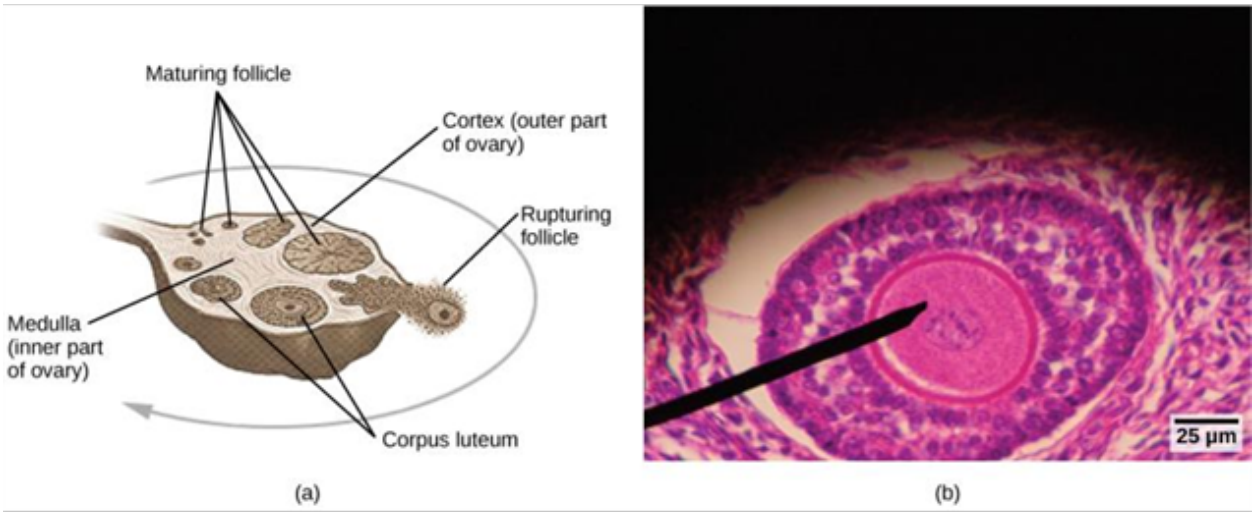



Figure 2: (A) The stages of oogenesis within the ovary. The developing follicle matures through defined stages - from Primordial, to Primary, to Secondary, to Graafian - facilitating development and ultimate expulsion of an oocyte into the reproductive tract (ovulation). (B) A light micrograph of a follicle with a maturing oocyte at its centre. Images by Edmund Atkins are licensed CC-BY 4.0 from Concepts of Biology, Chapter 24: Animal Reproduction and Development by Charles Molnar and Jane Gair.

**The zona pellucida**

The first point of contact between sperm and the oocyte occurs at the zona pellucida (ZP), which is an extracellular protein matrix surrounding the oocyte. In humans, ZP proteins are encoded by the maternal genes *Zp1, Zp2, Zp3, Zp4* (Gupta, 2015; Lefievre, 2004), however integrity of the zona pellucida is species-specific, for instance only Zp1, Zp2, and Zp3 proteins are found in mice (Bleil & Wassarman, 1980). The importance of each protein has been thoroughly investigated, with *Zp1*-null mice displaying reduced fecundity, and *Zp2*-null and *Zp3*-null phenotypes completely infertile (Rankin et al., 1996, 1999, 2001). Other proteins and factors have been implicated in sperm-ZP interactions, such as CRISP1 (Maldera et al., 2013), with characterisation of the molecular basis for ZP penetration, acrosome reaction, and entry of sperm DNA into the oocyte, an ongoing process (Kaji & Kudo, 2004). The physiological role of the ZP is to facilitate sperm-oocyte recognition, grant sperm entry into the perivitelline space, provide a post-fertilisation block to polyspermy in some species, and to ensure safe passage of the embryo through the oviduct into the uterus (Li et al., 2013). These cell-to-matrix and cell-to-cell functions are supported by molecules present in both gametes (Ikawa et al., 2010), however if ZP function is impaired, embryogenesis is unlikely to proceed.

Contact between sperm and oocyte initiates a cascade of physiological events required for successful fertilisation. In sperm, the key mechanism triggered by sperm-oocyte signalling is the acrosome reaction. The acrosome is a large Golgi-derived lysosome-like organelle, which is positioned over the nucleus in the apical region of the sperm head (Kaji & Kudo, 2004). This organelle secretes enzymes, such as acrosin, which aid penetration through the protein matrix (Abou-Haila & Tulsiani, 2000). After penetration, sperm enter the perivitelline space and are able to fuse with the oocyte membrane (Figure 3). If this occurs

with more than one sperm, this is termed polyspermy. In many species polyspermy is embryo-lethal, with several mechanisms evolving to 'block' additional sperm from binding after fertilisation. The underlying mechanisms regulating this process are not well characterised, and species-specific differences in sperm/egg interactions have been identified e.g. Izumo1:Juno interactions (Bianchi et al., 2014; Bianchi & Wright, 2015).

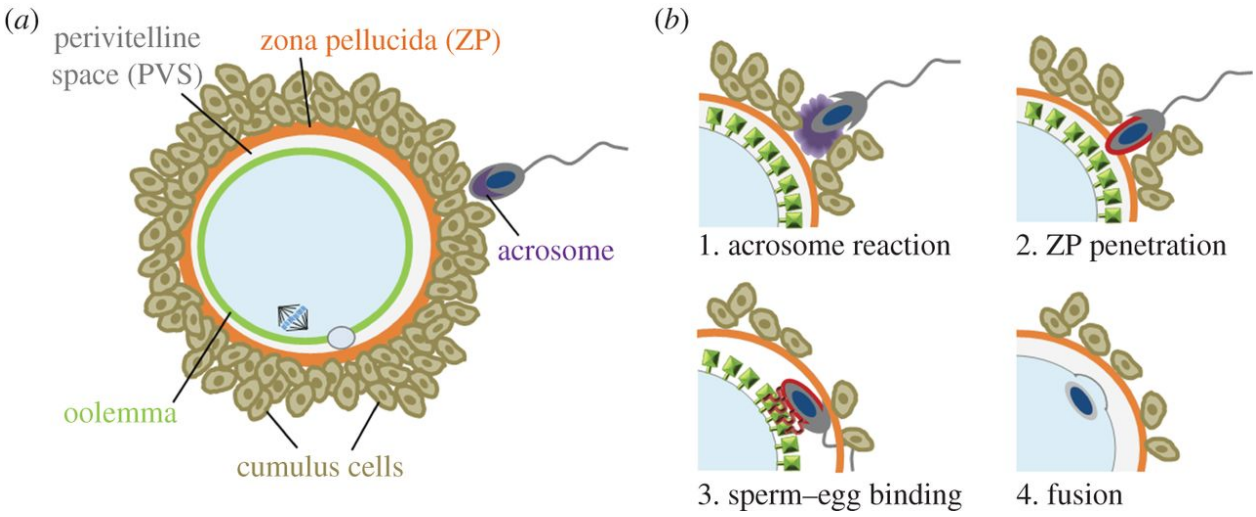



Fig 3. Fertilization in mammals: (a) anatomy of mature germ cells. (b) There are four key steps required for successful fertilization: (1) a sperm undergoes the acrosome reaction and passes through the cumulus cells surrounding the oocyte to bind to the zona pellucida, (2) the sperm penetrates the zona, (3) enters the perivitelline space, and (4) fuses with the plasma membrane. The fused gametes now form a zygote combining haploid genomes from each parent. Images by Enrica Bianchi and Gavin J. Wright are licensed CC-BY 4.0 from "Cross-species fertilization: the hamster egg receptor, Juno, binds the human sperm ligand, Izumo1. (2014) Phil. Trans. R. Soc. B, 370(1661)". DOI: 10.1098/rstb.2014.0101.

## Syngamy

Fusion between mature sperm and oocyte membranes initiates a further physiological cascade necessary for continued development. After binding to the membrane, sperm DNA is released from its protaminated state and undergoes decondensation, then it is then repackaged with maternal histones to form the male pronucleus (Becker, 2005; Mayer et al., 2000). The male pronucleus migrates toward the female pronucleus, and the two pronuclei combine, with the first mitotic spindle assembly forming around the newly assembled genome, which is composed of haploid genomes from both parents (Clift & Schuh, 2013). This process is termed syngamy, and it marks the creation of a genetically distinct organism. Syngamy is indicative of successful fertilisation, however several physiological events must take place before initiation of gene transcription and mitotic cleavage can occur, which are necessary for continued development of the newly formed zygote.

## Embryo division

Within the oocyte, events triggered by sperm-oocyte binding include cell-cycle resumption, cortical granule exocytosis, recruitment of maternal transcripts, decrease in MPF and MAP kinase activity, and a transient increase in PAF, with major events driven by intracellular calcium flux (Ducibella et al., 2002; Wells & O'Neill, 1994; O'Neill et al., 2015). These events are critically important for normal fertilisation and cell division, as blocking or altering calcium signalling at this stage negates the EGA and results in embryonic arrest (Ducibella et al., 2006; Ozil et al., 2006). The mechanisms involved in this cascade have not yet been fully characterised, however key regulators such as Paf (1-*o*-alkyl-2-acetyl-*sn*-glycero-3-phosphocholine) have been identified, which drive development in a species-

specific manner (Jin & O'Neill, 2011; O'Neill et al., 2015). Dysfunctional activation of transcription at the zygote stage (during the ZGA) or later embryonic stages (e.g. during the major burst of transcription known as the EGA) will culminate in embryonic arrest, with embryos failing to proliferate and develop after the MZT.

In normal embryonic development, cell division occurs shortly after fertilisation. The timeline for this process is largely species-specific and dependent on multiple factors (Baroux et al., 2008; Tadros & Lipshitz, 2009). Interestingly, maternal factors alone are sufficient for the embryo to cleave prior to the EGA in most plants and animals (Baroux et al., 2008). This has been demonstrated in many species after the administration of α-amanitin, an inhibitor of RNA polymerase II and III (Lindell et al., 1970), which prevents *de novo* transcription of the embryonic genome (Davis et al., 1996; Rajjou et al., 2004). Comparison of embryo cleavage patterns between model organisms shows commonalities, such as the timing of cell cleavages commonly regulated by transcription (O'Farrell, 2004). Embryo cleavage prior to the EGA is regulated by transcription of maternal mRNAs. Successful transition from maternal transcripts to embryonic transcripts during the MZT is required for successful progression to the morphogenesis stage of development (Figure 4).

The onset of morphogenesis marks the end of the MZT. This occurs between 1 and 15 cell cleavage cycles after fertilisation in all species investigated to date (Baroux & Grossniklaus, 2015). It takes place at the morula stage (16- to 32-cells) in mammals (Ducibella & Anderson, 1975), the blastula stage (~30,000 cells) in *Xenopus* and (Newport & Kirschner, 1982a, 1982b), the syncytial blastoderm stage (~6000 nuclei) in *Drosophila* (Wessel et al., 2015), and the globular-stage (~100 cells) in *Arabidopsis* (Olsen, 2001). Cell lineage differentiation and cellular migration marks the beginning of gastrulation and

organogenesis, with continued development after this stage no longer maintained by maternal factors (Li et al., 2013). As with early cell cleavage, divisions after the MZT are also primarily governed by gene transcription; maternal transcripts govern cleavage prior to the MZT, and embryonic transcripts govern cleavage after the MZT. Gene transcription at all stages of early development is largely determined by epigenetic mechanisms established prior to the EGA (Reik, 2001; Stelzer et al., 2015).

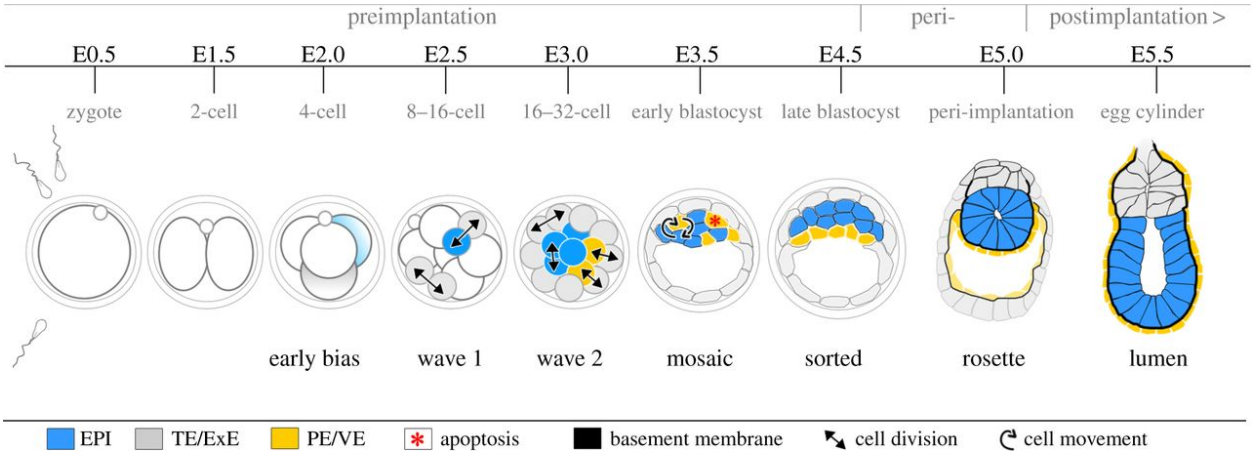



Fig 4. Morphogenesis of the mouse embryo. The timing of cell cleavages and the onset of morphogenesis (the morula stage; E3.0) are shown. The pluripotent inner cell mass (ICM) and epiblast (EPI) shown in blue will ultimately give rise to all tissues of the body. The primitive endoderm (PE) and visceral endoderm (VE) shown in yellow provide nutritional and structural support to the pluripotent ICM, aiding establishment of the anterior–posterior axis. The extraembryonic lineage is comprised of the trophectoderm (TE) and extraembryonic ectoderm (ExE) and will ultimately form the placenta and placental membranes. Image by Ivan Bedzhov, Sarah J. L. Graham, Chuen Yan Leung, and Magdalena Zernicka-Goetz is licensed CC-BY 3.0 from "Developmental plasticity, cell fate specification and morphogenesis in the early mouse embryo (2014). Phil Trans R Soc B 369(1657): 20130538." https://doi.org/10.1098/rstb.2013.0538

## Epigenetics and the maternal-to-zygote transition

Complex multicellular organisms exhibit a diverse range of cellular phenotypes despite all cells within an organism containing virtually identical genomic DNA. The physiological mechanisms responsible for this phenomenon constitute the field of epigenetics (Berger et al., 2009; Goldberg et al., 2007). The term 'epigenetics' is commonly defined as heritable patterns of gene expression that cannot be explained by changes in DNA sequence alone (Bird, 2007; Russo et al, 1996). Epigenetic mechanisms determine cell phenotype via interrelated biological processes such as histone modifications, DNA methylation, non-coding RNA interaction (such as micro RNAs, small interfering RNAs, and PIWI-interacting RNAs) and DNA binding factors, all of which act to dynamically regulate DNA transcription within the cell (Lawrence et al., 2016; Rabani et al., 2014; Reik et al., 2001; Whitelaw & Garrick, 2005).

Epigenetic regulation of transcription is necessary for life in eukaryotes. Abnormal or disrupted epigenetic activity can result in impaired or arrested transcription and cell death, with defects typically manifesting very early in development (Morgan et al., 2005). Epigenetic marks are first established during gametogenesis in a sex-specific manner, with patterns of DNA methylation established globally and in specific 'imprinting' regions (Kafri et al., 1992; Monk et al., 1987, Reik et al., 2001). In mature gametes, the male and female genomes are transcriptionally quiescent until fertilisation when both genomes undergo extensive epigenetic modification in order to initiate transcription of DNA and reprogram embryonic cells to the totipotent state: a fundamental step in the maternal-to-zygotic transition (Chavez et al., 2014; Reik et al., 2001; Reik, 2007). Epigenetic control and initiation of transcription during the MZT is highly complex; despite the importance for successful development, many of the processes involved are not fully understood.

**Histone modification and DNA methylation**

Epigenetic control of gene expression relies on the structural conformation of DNA within the nucleus. Each human cell contains approximately 2m of DNA, which is compacted and organised into the macromolecular complex chromatin (Bloom & Joglekar, 2010). Chromatin is comprised of DNA, RNA and proteins arranged in a conserved structure consisting of 145-147bp of DNA tightly wound around protein octamers containing two of each histone H2A, H2B, H3, and H4 (Luger et al., 1997). These DNA-encircled histone units are termed 'nucleosomes' and they occur approximately every 200bp of DNA in an arrangement resembling beads on a string (Kornberg, 1974; Lawrence et al., 2016). First reported in the early 1960s (e.g. Allfrey et al., 1964), histone tails extending from the nucleosome can undergo covalent post-translational modifications such as methylation to regulate DNA accessibility and transcription, for example methylation of the 4th lysine residue of the H3 tail (H3K4) is a common modification associated with gene activation whereas H3K27 methylation typically correlates with gene repression (Barski et al., 2007).

Various histone modifications have been identified and characterised, such as methylation, acetylation, phosphorylation, ubiquitinylation, sumoylation, ADP ribosylation, deamination, propionylation and butyrylation, which all act to facilitate or inhibit access of the transcriptional machinery to regions of genomic DNA (Kebede et al., 2015; Kouzarides, 2007; Lorch et al., 1987). Many of these modifications are established shortly after conception and the effects of epigenetic marks on gene expression are essential for establishing cell identity and function in the early embryo (Barski et al., 2007). Recent evidence suggests that the number of free nucleosome-forming histones within the nucleus also plays a determining role in initiating the ZGA (Joseph et al., 2017). In addition to histone modifications (e.g. methylation of the histone tail), DNA can be directly

methylated at specific sites. DNA methylation is relatively stable and heritable, with daughter cells inheriting epigenetic information from parent cells throughout life. Dynamic DNA methylation and demethylation is essential for successful activation of the incipient embryonic genome.

The process of DNA methylation involves the addition of a methyl group to the fifth carbon of cytosine residues in CpG dinucleotides, thereby creating 5-methylcytosine (5meC), which alters DNA structural fidelity and can alter transcription (Holliday & Pugh, 1975, 1996). CpG dinucleotides (when a cytosine nucleotide is followed by a guanine nucleotide) are often found at the beginning of transcription start sites; regions with a high frequency of CpG sites are termed 'CpG islands'. DNA methylation at CpG islands is a vital mechanism for many cellular processes, such as X chromosome inactivation in females (Robertson & Wolffe, 2000), ageing (Zampieri et al., 2015), and silencing parasitic retrotransposons (Yoder et al., 1997). DNA methylation is mediated by DNA methyltransferases (DNMTs), with different DNMTs providing different functions (Law & Jacobson, 2010). These functions can be broadly characterised as either *de novo* methylation (eg DNMT3a, DNMT3b, DNMT3L), or maintenance of methylation (eg DNMT1) (Fuks et al., 2000; Hermann et al., 2004; Okano et al., 1999).

DNA hypermethylation typically correlates with transcription repression due to the actions of methyl-CpG binding proteins and recruitment of histone deacetylase complexes, and DNA hypomethylation typically correlates with active transcription (Bogdanović & Gómez-Skarmeta, 2014). Assessing methylation status in the embryo can be performed with relatively high accuracy using scientific techniques such as ChIP-Seq and bisulfite sequencing. Prior to the ZGA (before the zygote's genome is able to be transcribed) embryonic cells undergo a global wave of demethylation that does not activate or repress

gene expression, but instead removes a substantial proportion of inherited epigenetic modifications establishing an epigenetic 'ground-state' (Iurlaro et al., 2017; Jones, 2012).

**DNA demethylation**

When the sperm and the oocyte fuse and initiate fertilisation, DNA within the oocyte is packaged loosely with maternal histones, and it remains in the metaphase stage of the second meiotic division (MII). Conversely, sperm DNA is highly protaminated, and before fertilisation can proceed, protamine removal and decondensation of the sperm head must take place. This is accompanied by a wave of global DNA demethylation within the zygote. Subsequent repackaging of sperm DNA with maternal histones forms the male pronucleus (Figure 5). Following syngamy (the union of the two pronuclei), the zygotic genome is passively demethylated in a DNA replication-dependent manner until the morula stage, when active *de novo* methylation is established via enzymes such as DNMT3b, facilitating cell lineage differentiation (Lee et al., 2014; Rougier et al., 1998; Watanabe et al., 2002). DNA demethylation of the paternal genome is initiated rapidly after sperm-oocyte fusion (~4-6 h after first contact), beginning prior to the first round of DNA replication, indicating that it is an active process (Santos et al., 2002). Active demethylation of DNA within the paternal pronucleus is necessary for the first events of the ZGA (Ambrosi et al., 2017).

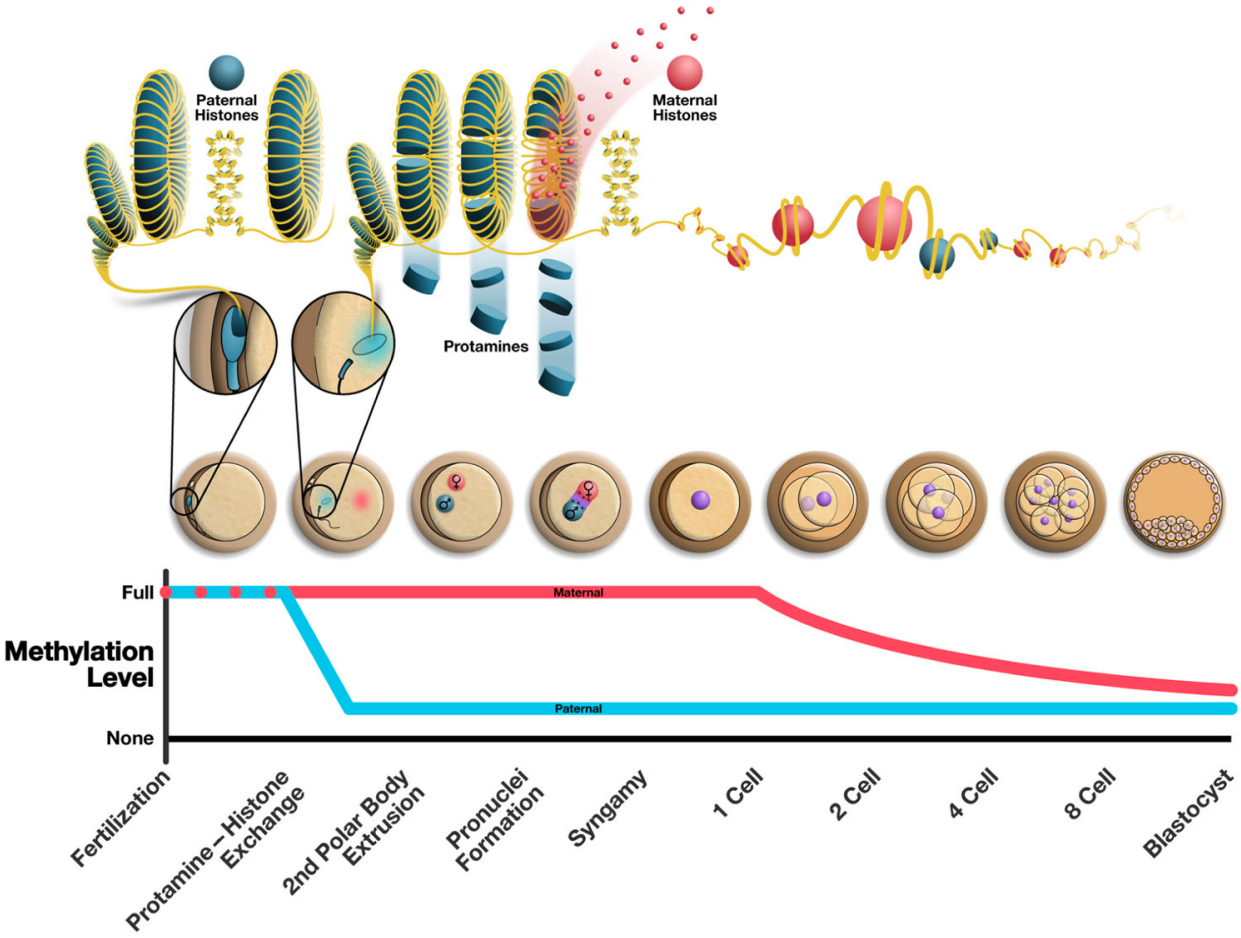



Figure 5. Epigenomic processes during fertilization. Chromatin structure of the mature sperm (highly protaminated and tightly packed) transitions, with maternally derived histones replacing protamines, resulting in decondensation of the sperm head. Methylation level over time, for both the maternal and paternal pronucleus, is shown, where the paternal pronucleus undergoes active demethylation and the maternal DNA is demethylated passively in a DNA replication dependent manner. The approximate chronology of major events in the early embryo is outlined along the bottom of the figure. Image by Timothy G. Jenkins and Douglas T. Carrell licensed CC-BY 3.0 from Dynamic alterations in the paternal epigenetic landscape following fertilization. Front. Gene. 3:143. doi:10.3389/fgene.2012.00143.

The concept of active DNA demethylation in early development was first established in plants, with the underlying mechanisms regulating this process unidentified in mammals

until relatively recently (Wu & Zhang, 2010). Active demethylation of paternal DNA is yet to be identified in many non-mammalian vertebrates and this is an active area of research (Beaujean et al., 2004; Lei et al., 2015). Implicating paternal DNA demethylation in gene transcription initiation is the recent discovery of the ten-eleven translocation (TET) enzymes (Kohli & Zhang, 2013). These enzymes, specifically TET3 in the mammalian zygote, act to oxidize 5-methylcytosine (5meC) to 5-hydroxymethylcytosine (5hmC), facilitating active and passive DNA demethylation to globally reprogram the epigenetic landscape of the zygote in preparation for transcription of the embryonic genome (Guo et al., 2014). Analysing the relative expression of DNA methyltransferases and the activity of TET enzymes can aid in identifying abnormal patterns of DNA methylation, and provide an indication of developmental potential of the embryo. Although transient DNA demethylation is crucial for continued development, this process must occur in a controlled fashion, as DNA demethylation affecting *all* genes would very likely be detrimental to development and impede the MZT (Iurlaro et al., 2017).

**Imprinting**

Global demethylation of the zygotic genome does not remove all epigenetic marks. Interestingly, a small number of genes, termed ‘imprinting’ genes, are protected from DNA demethylation by incompletely understood molecular mechanisms (Li et al., 2008; Messerschmidt et al., 2012; Nakamura et al., 2012). These genes retain their methylation state throughout embryonic development, which has significant functional implications for development of the embryo (Reik et al., 2001; Santos & Dean, 2004). Imprinting genes were discovered ~30 years ago, with separate experiments creating ‘same sex’ zygotes i.e. transplanting pronuclei to create a zygote with two male, or two female, pronuclei (McGrath & Solter 1984; Surani et al., 1984). These experiments proved both male and

female genomes were required for development, demonstrating that inherited parental genomes are not functionally equal in mammals: a dogma arising from the work of Édouard Van Beneden on fertilisation and meiosis in the late 1800s (Section 2.1).

Imprinting in mammals gives rise to parent-specific monoallelic expression of ~150 genes including well-known genes *H19* and *Igf2r*, which are paternally inherited, and *Igf2*, which is maternally inherited (Barlow et al., 1991; Bartolomei et al., 1991; Dechiara et al., 1991). The expression pattern of imprinted genes appears to be regulated by the parent of origin, with DNA methylation and histone modifications in *cis*-acting regulatory elements activating or silencing either the maternal or paternal allele (Massah et al., 2015; Murrell et al., 2004). If this process is dysfunctional, it can lead to impaired embryonic development and undesirable consequences for the organism. Imprinting-related conditions in humans include Prader-Willi syndrome and Beckwith-Wiedemann syndrome: both are caused by disrupted imprinting in single genes and illustrate severe phenotypic outcomes of aberrant methylation state (Cassidy, 1997; Weksberg et al., 1993).

The effect of dysfunctional imprinting during embryonic development has also recently been identified in patients derived from assisted reproductive techniques e.g. *in vitro* fertilisation (Debaun et al., 2003; Lazaraviciute et al., 2014; Turan et al., 2010). Within this patient cohort there is an increased incidence of imprinting-related disorders, however, as with most epidemiological observations, the underlying mechanisms are difficult to determine with confidence (Loke & Craig, 2017; Song et al., 2015). Dysfunctional imprinting and global methylation patterns are also evident in SCNT and cloned embryos and are likely implicated in the limited success rate of these techniques. These examples illustrate the effects of impaired epigenetic reprogramming in the early embryo. We have an incomplete understanding of epigenetic mechanisms during development and their

control of gene expression during the MZT. Given the historical and clinical relevance of this field of research, further investigation into epigenetic regulation of embryonic transcription initiation is warranted.

## Embryonic genome activation

The embryonic genome activation (EGA) was initially defined as the point in time when a large increase in *de novo* RNA synthesis could be measured via incorporation of radio-labelled uridine or adenosine in the embryo (Zalokar 1976, Clegg & Piko, 1977). Subsequent investigation utilising more sensitive techniques has since revealed that *de novo* transcription of the embryonic genome (i.e. mRNAs of non-maternal origin) first occurs earlier than previously established, with microinjected reporter genes in *Xenopus* and mouse embryos demonstrating the capacity for transcription during the zygote stage of development (Aoki et al., 1997; Latham et al., 1992; Newport & Kirschner, 1982b; Telford et al., 1990). A number of RNAs have since been identified as non-maternal in origin (i.e. they have come from the newly formed embryonic genome), with this early 'wave' of *de novo* transcription termed the 'minor' ZGA. Products from the ZGA are then involved in the major burst of transcription traditionally referred to as the EGA, which plays a key role in the transition from maternal to embryonic control of development (Baroux et al., 2008; Leichsenring et al., 2013; Tadros & Lipshitz, 2009). Although many ZGA-specific genes have been successfully identified, fully characterising this process in humans and model organisms using modern scientific techniques has the capacity to improve our understanding of the biological mechanisms governing early development.

Maternal transcripts are far more abundant in the early-stage embryo than *de novo* transcripts, with dynamic turnover of maternal mRNAs taking place in the newly formed

embryo throughout the MZT (Piko & Clegg, 1982). This process is regulated by different methods, including $m^6A$ mRNA labelling and clearance, post-translational deadenylation of the 3' poly(A) tail and decapping (Houseley & Tollervey, 2009; De Renzis et al., 2007). This creates a challenge for identifying transcripts derived from the zygotic genome as the transcript origin is masked when traditional scientific techniques are applied. Modern methods used successfully to differentiate between maternal / zygotic transcripts in model organisms include depleting maternal mRNAs via subtractive hybridization (Wu et al., 2003), blocking *de novo* synthesis in the zygote using an RNA polymerase inhibitor such as α-Amanitin (Zeng & Schultz, 2005), inducing selective loss of the zygotic contribution via ablation of whole chromosome arms (De Renzis et al., 2007), and using specialised bioinformatics to identify transcript origin based on the timing and regulation of transcription within single-cell RNA-Seq data (Ge, 2017; Xue et al., 2013). Recently, the application of next-generation platform technologies, and chemical treatments to alter either maternal or zygotic RNAs, have enabled significant progress in characterising the maternal and zygotic contributions to the genome activation.

**Recruitment of maternal transcripts**

Identifying the mechanisms that govern the ZGA/EGA in mammals has proven unexpectedly challenging, even with the application of modern scientific techniques. Our current understanding of gene transcription during mammalian development, and the mechanisms involved in this process, begins with the oocyte. As mentioned previously (Section 2.2), transcription during oocyte maturation and fertilisation is a dynamic process. Msy2-dependent RNA binding curtails mRNA degradation prior to fertilisation, with these mRNAs conserved in the transcriptionally quiescent oocyte, lying dormant until they are recruited for translation in the transcriptionally active embryo (Brower et al., 1981; De Leon

et al., 1983; Jahn et al., 1976; Svoboda et al., 2015). The molecular processes of recruitment and translation of these mRNAs are relatively well-defined (Groppo & Richter, 2009), however the functions of proteins generated in this fashion are not well characterised, with identification of maternally-derived transcripts traditionally a serendipitous consequence of pursuing unrelated research aims (Svoboda et al., 2015).

The means to identify patterns of recruited maternal transcripts was established with the development of microarray technologies. Microarray experiments that use oligo(dT) priming to generate cDNA are able to identify maternal-origin mRNAs, as well as the regulatory sequences within the 3'-untranslated region (3'UTR) that govern the recruitment process (Peaston et al., 2004; Zeng et al., 2004). Recruitment of maternal mRNAs is thought to rely on the elongation of the poly(A) tail (Thélie et al., 2007), which results in more efficient oligo(dT) priming (Ma et al., 2012; Mehlmann, 2012) and can be used to identify transcripts that increase in relative abundance between the oocyte and zygote stages (indicative of recruitment of maternal mRNA). Although the functional roles of many recruited transcripts remain unknown, some transcripts have been implicated in key developmental processes, significantly advancing our understanding of how life begins.

Oocyte activation is dependant on calcium release from internal stores, which is regulated by inositol 1,4,5-triphosphate ($IP_3$) (Ducibella et al., 2006; Ridgway et al., 1977; Steinhardt et al., 1977). This calcium signalling mechanism is critical for embryo survival and cell division (Kaneuchi et al., 2015), (briefly discussed in Section 2.6), and reliant on the presence of the $IP_3$ receptor (ITPR1) (Xu et al., 2003). Oocyte maturation correlates with an increase in ITPR1 translation, with ITPR1 protein produced from recruited maternally-derived transcripts. This process is known to implicate CAMK2G, also encoded by maternal mRNA, with low levels of ITPR1 / CAMK2G acting to prevent fertilisation of

immature oocytes. The concomitant increase in ITPR1 / CAMK2G has been shown to facilitate fertilisation in activated mature oocytes (Backs et al., 2009). It is this activation event that marks the first detected degradation of maternal mRNAs from the embryo, and the onset of the MZT (Paynton et al., 1988).

**Degradation of maternal transcripts**

Inherited maternal transcripts are progressively degraded in the embryo in a selective manner. A recent report by Zhao et al. (2017) provides the predominant theory for how maternal mRNAs are selected for clearance. In the zebrafish (*Danio rerio*), the epitranscriptomic modification N6-methyladenosine ($m^6A$) was found in over a third of maternal mRNAs, and clearance of these 'marked' mRNAs was facilitated by an $m^6A$-binding protein, Ythdf2. Silencing Ythdf2 in embryos impedes maternal mRNA clearance, and negates the MZT in almost all animal subjects examined. The ability to reach adulthood in a small fraction of Ythdf2 KO mutants indicates the existence of parallel pathways for selective clearance of maternal mRNAs.

Another candidate for selecting mRNAs for clearance is miR-430, which appears to overlap with Ythdf2 in function: the decay of common targets occurs earliest, followed by the specific targets of Ythdf2, then specific targets of miR-430 (Zhao et al., 2017). The authors postulate that "Ythdf2 is a maternally supplied factor acting to degrade maternal mRNAs in the zygote, which is then followed by zygotically supplied miR-430 for further maternal RNA degradation". This model of $m^6A$-based selection of mRNAs for clearance has not yet been validated or investigated in other species, however alternative theories have very little experimental support. The process of degrading and clearing mRNAs once they are selected is far better understood.

The primary mechanisms of mRNA degradation is comprised of deadenylation of the 3' poly(A) tail and decapping (Houseley & Tollervey, 2009; De Renzis et al., 2007; Su et al., 2007). The RNA- / DNA-binding protein Msy2 (described earlier as stabilising transcripts in the oocyte), is integral to the process of mRNA degradation, with CDK1-mediated phosphorylation of Msy2 'triggering a switch' from mRNA stabilisation to mRNA destabilisation, following oocyte activation (Medvedev et al., 2008). Species-specific differences have been identified in this process: up to 50% of total RNA is depleted in mammalian embryos, however far less RNA clearance takes place in *Xenopus* and *Danio rerio* (zebrafish) embryos, with total RNA (including ribosomal RNA) remaining relatively constant throughout the degradation process (Piko & Clegg, 1982). The functional role of species-specific total RNA maintenance or depletion is not fully understood (Alizadeh et al., 2005).

**Activation of the zygotic genome**

As maternal control of development subsides, the transition to autonomic control of development begins. The timing and onset of the ZGA are well characterised in model species (Baroux et al., 2008; O'Farrell et al., 2004), however the mechanisms driving the process are still being investigated. Species-specific differences partially account for this 'gap' in our knowledge, as different mechanisms have been found to play different roles across species (Heyn et al., 2014). This variation is apparent in the timing of the ZGA; transcription occurs remarkably quickly in mice, taking place during the 1-cell zygote stage (~7 h post-pronucleus formation), whereas in the cow it is delayed until 8-16 cell stage, taking days longer to reach this developmental milestone (Barnes & Eyestone, 1990; Frei et al., 1989; Meirelles et al., 2004).

A conventional approach for investigating the ZGA is to expose early stage embryos to α-amanitin, a potent inhibitor of RNA polymerase II and III, then assess the developmental stage reached (Warner & Versteegh, 1974). In mice, early zygote-stage embryos exposed to α-amanitin progress to the 2-cell stage, before succumbing to cleavage arrest. Interestingly, when exposed to suboptimal *in vitro* culture environments, mouse embryos fail to progress past the 2-cell stage. This phenomenon is termed the '2-cell block', and it has been reported to reflect impaired activation of the zygotic genome (Goddard & Pratt, 1983; Schultz, 1993). Overcoming this developmental block, by providing embryos with a more supportive *in vitro* environment, has been shown to ameliorate the abnormal ZGA and failed development.

A key mechanism identified as driving the EGA in mammals is the Paf signalling pathway. Paf is a potent ether phospholipid implicated in many aspects of pregnancy in many animal species (Fenelon et al., 2013; Tiemann, 2008). The primary action of Paf is to trigger the CREB family of transcription factors (CREB / ATF1), which bind CRE response elements, initiating a 'chain reaction' of transcription of ~4000 genes (Jin and O'Neill, 2011). Activation of the CRE system also leads to activation of other transcription factors, notably Jun, Jund, Junb, Egr1, Fosl2, Cebpb, Sp4, Sp1, Atf3, Btg2, Fosb, and Klf5, from which the entire transcriptome is generated (O'Neill et al., 2015). The expression pattern of CREB transcription factors are not typically used as markers of the ZGA/EGA, as the small transient increase in expression is not easily discernible against background noise. The majority of these markers do not require a large increase in transcription to be effective as the signalling pathway amplifies the effect. Genes downstream of CREB transcription factors have traditionally been more effective as markers of gene activation when characterising the ZGA.

**Markers of activation**

Genes traditionally used as markers of the ZGA / EGA are those involved in the molecular machinery governing translation. A well-known example is eukaryotic initiation factor 1 alpha (eIF-1α) (Davis et al., 1996). The role of eIF-1α in development is to assemble the 40S ribosome complex required for protein production. This event occurs during the ZGA, with transcription of eIF-1α beginning at the early 2-cell stage in humans. The assembled 40S ribosome complex is required for rapid increase in transcription which characterises the EGA. Once the 40S ribosome complex is assembled protein levels can be used as markers of the EGA in place of mRNA levels. Heat shock proteins, such as Hsp70.1, have traditionally been used for this purpose, with a rapid increase in protein levels during the EGA (Christians et al., 1995). Other markers include 'housekeeping genes', such as ß-actin and Gapdh (Bunnell et al., 2011), which are defined as having a relatively consistent, unchanging pattern of gene expression throughout development, and are often used for normalising gene expression data in qPCR-based experiments. Other markers of the ZGA / EGA include maternal factors Ago2 (*Eif2c2*) (Lykke-Andersen et al., 2008), Mater (*Nlpr5*) (Peng et al., 2012), Sox2 (Pan & Schultz, 2011), and Nanog (Lee et al., 2013), all of which play critical roles in development.

**Conventional investigative approaches**

Conventional scientific techniques employed to investigate gene activity in early development include primarily RT-PCR and Northern blot analysis. Techniques such as expressed sequence tag (EST) analysis, and subtractive hybridization, have also been applied, but to a lesser extent (Ko et al., 2000). Technical limitations such as the small number of cells and ephemeral nature of the event being investigated have impeded

progress in this field. For example, extraction and reverse transcription yields must be extraordinarily high when conducting PCR on embryos, and whole genomic amplification introduces substantial bias in the results. Normalisation is also difficult, as genes once thought to be stable at this early time point are unreliable for this purpose (Gu et al., 2014). In addition, the small amounts of RNA/cDNA generated restrict the number of genes / samples that can be investigated with a PCR-based approach, increasing the number of animals required to conduct experiments. There is also a large difference in the absolute quantity of mRNA in the embryo before / after the EGA, requiring customised PCR cycles, and restricting the ability to draw comparisons between groups. Modern technologies are able to overcome many of these technical limitations, however the choice of technique and the methodology employed need to be properly evaluated for this application.

**Modern investigative approaches**

Massively parallel RNA sequencing (RNA-Seq) is a useful technology for investigating transcript abundance in early development. This technology is continuously improving, with the use of single-cell and long-read RNA-Seq increasingly employed in research labs around the world, however challenges and limitations remain. One limitation of RNA-Seq is the requirement for a genomic reference when mapping reads; this can restrict the use of this technology in species for which a genome assembly is not yet available. An alternative to genome-based mapping exists, whereby RNA-Seq data can be assembled *de novo* into a transcriptome, then the RNA-Seq reads can be mapped to this reference to generate transcript counts in place of a reference genome (Birol et al., 2009). *De novo* transcriptome assembly is an appealing approach for studying development in non-model organisms, as it is more cost-effective than sequencing a genome and the data produced can be of similar quality and utility (Yandell & Ence, 2012).

### Modelling the EGA

Recent studies comparing molecular, genetic, epigenetic and morphological profiles of human embryos and mice at different stages of development have identified important physiological differences that have restricted translation of findings between mice and humans (Chavez et al., 2012; Plachta et al., 2011; Vassena et al., 2011; Wong et al., 2010; Wossidlo et al., 2011; Xue et al., 2013). The development and continued refinement of molecular techniques such as CRISPR gene editing (Cong et al., 2013) is reducing reliance on the mouse as an animal model, with research now more amenable to non-traditional organisms for investigation and modelling of human physiological processes. Popular models such as pigs, sheep and cows provide benefits over the mouse in terms of their pattern and timing of development, however there are also significant disadvantages to using large animals in research. We propose the spiny mouse (*Acomys cahirinus*) as a candidate for modelling early human development, as preliminary evidence suggests it closely matches the developmental profile of the human embryo, more so than the mouse.

## The spiny mouse (Acomys cahirinus)

The Egyptian, or 'Common', spiny mouse (*Acomys cahirinus*) is a precocial murid species native to the arid semi-deserts of Africa and the Middle East (Osborn & Helmy, 1980; Wilson & Reeder, 2005). Derived from the Greeke "*akn*", meaning a sharp point, and "*mys*", from mouse, *Acomys* are named for the 'hair-like' spines comprising their dorsal pelage (Montandon et al., 2014; Osborn & Helmy, 1980; Simpson, 1945). Spiny mice were first employed in biomedical research as a model of diet-induced type 2 diabetes (Gonet et al., 1966); this spontaneous condition was likely an unfortunate consequence of being fed an energy-dense diet under conditions of captivity (Shafrir & Adler, 1983), as this condition

is not commonly seen in captive acomys populations maintained on a standard mouse chow diet (Dickinson & Walker, 2007, Haughton et al., 2016). Following use as a model of diabetic nephropathy, characterisation of basic physiology and behavior was conducted (Brunjes, 1990). More recently, the spiny mouse has been used to model perinatal development and events at birth (Dickinson et al., 2005; 2007; Ellery et al., 2013). Spiny mice follow a precocial pattern of development, with key developmental milestones occurring prenatally at a relative rate closely approximating human development (D'Udine & Alleva 1987; Dickinson et al., 2005). However, embryogenesis and early development have not been fully investigated in this species.

**Ecology and Phylogeny**

Prominent throughout much of Africa, spiny mice are distributed across coastal south-western and south-central Asia. They live in small social groups, and their primary habitat is dry, sandy, rocky desert, where they hide and forage amongst boulder screes (Jones & Dayan, 2000). They are not a nest-building species, but they have been known to occupy burrows belonging to other animals, such as gerbils (Walker, 1968). Classified as a member of the Subfamily *Muroidae*, numerically the most diverse mammalian group with over 1,300 named species (Flynn et al., 1985), classification of *Acomys* was initially based on molar morphology (Denys et al., 1992). The presence of a $T_3$ cusp instead of a $T_1$ cusp on the third upper molar is unlike all extant murine species, correlating to only a small number of extinct species within this Subfamily (Skinner and Chimimba, 2005). This finding suggests a weak taxonomical relationship between the spiny mouse and traditional rodents, which is supported by albumin immunoreactivity studies comparing spiny mice and other muroids (Chevret et al., 1993).

Modern phylogenetic analysis suggests spiny mice may be significantly removed from 'Old World' mice and rats (Murinae). Karyotyping and chromosome banding suggest a closer relationship to the Mongolian Gerbil (Meriones unguiculatus, subfamily Gerbillinae) (Agulnik & Silver, 1996; Martin et al., 2000), and preliminary transcriptomic analysis suggests the musk shrew (Suncus murinus, family Soricidae) to be the closest relative (Fushan et al., 2015). The current estimated point of evolutionary divergence from *Mus musculus* is ~35.6 million years ago, and divergence from *Suncus murinus* took place ~97.5 million years ago (Hedges et al., 2006, 2015), however these estimates are likely to change when genomic DNA sequencing and analysis is conducted for *Acomys* (Wray, 2001).

**Divergent physiology from typical rodents**

As a likely consequence of their atypical phylogeny, spiny mice display physiological characteristics not found in other rodents (Brunjes, 1990). Development in *Acomys* is relatively unique, and occurs in a precocial fashion, with key developmental milestones occurring prenatally at a relative rate closely approximating human development (D'Udine & Alleva 1987; Dickinson et al., 2005). They also display an extraordinary capacity for skin regeneration (Seifert et al., 2012), a unique hormonal milieu (Lamers et al., 1986; Quinn et al., 2013, 2014), and a menstrual cycle (Bellofiore et al., 2017); characteristics of considerable interest within the biomedical research community. The spiny mouse has proven to be an effective model of human physiology and pathophysiology, especially during fetal development, with anatomical and physiological characterisation ongoing.

A key characteristic supporting the use of *Acomys* in biomedical research is their developmental profile. Development in *Acomys* was first described at the beginning of the

20th century, with fetal anatomy and placental structure characterised as significantly differing to that of the mouse and rat (Assherton, 1905). Spiny mice begin *ex utero* life comparatively well developed (Assherton, 1905; Dieterlen, 1962; Porter & Doane, 1976; Porter & Etscorn, 1974). Gestation is ~38-39 days post-conception, and pups are born covered with fur, with their eyes and ears open, and displaying sophisticated locomotive, thermoregulatory and sensory capabilities shortly after birth (Brunjes, 1990; Dickinson et al., 2007; Young, 1976). The development of major organs occurs largely *in utero*, with the lung (Oosterhuis et al., 1984), liver (Lamers et al., 1985), kidney (Dickinson et al., 2005), small intestine and pancreatic enzymes (Lamers et al., 1999), muscle (Cannata et al., 2010) and immune system (Dickinson, unpublished) nearly completely developed by the time of birth. This bears a closer resemblance to the pattern of development in humans than to other rodents, which typically complete organogenesis in the days following birth (Dickinson et al., 2005). This human-like pattern of growth is also reflected in spiny mouse brain development (Brunjes, 1985; Brunjes, 1989; Fleiss et al., 2011).

In addition to their long gestation, and predominantly *in utero* organogenesis, spiny mice have an endocrine profile not typical of conventional rodents. Cortisol, rather than corticosterone, is the primary circulating glucocorticoid in the spiny mouse (Lamers et al., 1986; Quinn et al., 2013), and both the adrenal gland and the brain synthesise dehydroepiandrosterone (DHEA) (Quinn et al., 2014); this is unlike other rodents, and all other animals excluding primates, however it relatively similar to human endocrinology. This characteristic has been exploited in recent studies, investigating the fetal effects of maternal stress (Dickinson et al., 2007: O'Connell et al., 2011), birth asphyxia (Hutton et al., 2009) and maternal infection (Ratnayake et al., 2012). Identifying the most

physiologically relevant species for investigating human physiology is paramount for translation of findings to humans.

**Embryogenesis**

The female spiny mouse is known to exhibit a postpartum estrus within 24 h of delivery of her litter (Dickinson et al., 2007). This correlates with a pre-ovulatory surge of progesterone from 6-12h post-delivery, followed by low progesterone until 10 days gestational age (Dickinson; Figure 6), accommodating conception of the next litter in a relatively short time-frame, typically 12 h after giving birth. This rapid postpartum copulation has clear benefits for investigating embryo development, as it allows accurate staging of preimplantation embryonic development, although the timing of *in vivo* embryonic cell divisions has not yet been reported.

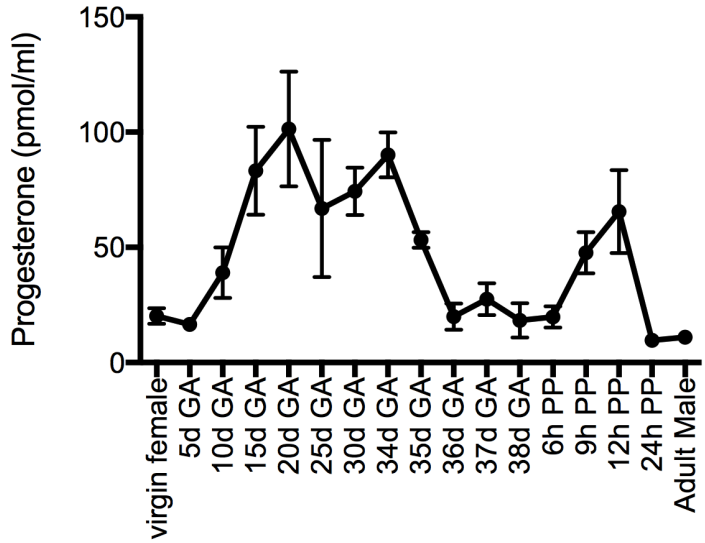



Figure 6: Progesterone profile during and after pregnancy in the spiny mouse (Hayley Dickinson, unpublished).

Embryonic development from the 2-cell stage to the morula stage of development typically occurs over a ~62 h period in humans (Niakan et al., 2012), however, this growth period is significantly shorter in the common mouse, taking only ~42h to reach the same developmental milestone (Bowman and McLaren, 1970). Differences in the timing of implantation are also evident, with human embryos typically implanting ~5.5 days after conception, compared to ~4.5 days in the mouse (Harlow & Quinn, 1982). Preliminary data suggest the rate of embryo development in the spiny mouse more closely reflects the timing of human embryos, when compared to mouse embryos (Figure. 7).

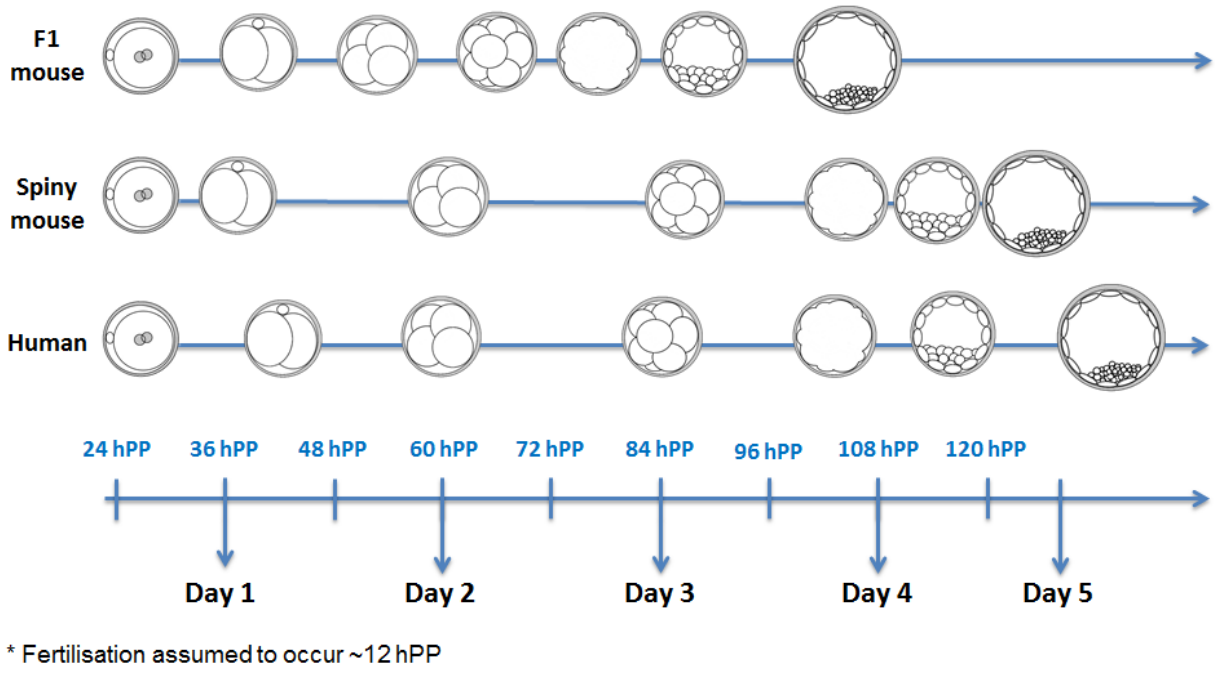



Figure 7: time taken to reach each cell division in mouse, human and spiny mouse.

There is currently a lack of information in the literature regarding key developmental process, such as when the ZGA/EGA and MZT occurs, the timing and process of trophoblastic invasion, and the morphology of the spiny mouse embryo after implantation until mid-gestation.

The aims of this thesis are to characterise early development in the spiny mouse (*Acomys cahirinus*) and to assess the utility of this species as a model of early human development.

## References


Aanes H, Winata CL, Lin CH, Chen JP, Srinivasan KG, Lee SG, Lim AY, Hajan HS, Collas P, Bourque G, Gong Z, Korzh V, Alestrom P, Mathavan S (2011). Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome research*, 21(8): 1328–1338.

Abou-Haila A, Tulsiani DR (2000). Mammalian Sperm Acrosome: Formation, Contents, and Function. *Archives of biochemistry and biophysics*, 379(2): 173–182.

Adams EC, Hertig AT (1964). Studies on guinea pig oocytes: I. Electron Microscopic Observations on the Development of Cytoplasmic Organelles in Oocytes of Primordial and Primary Follicles. *The journal of cell biology*, 21(3): 397–427.

Adenot PG, Mercier Y, Renard JP, Thompson EM (1997). Differential H4 acetylation of paternal and maternal chromatin precedes DNA replication and differential transcriptional activity in pronuclei of 1-cell mouse embryos. *Development*, 124(22): 4615–25.

Agulnik SI, Silver LM (1996). The Cairo spiny mouse Acomys cahirinus shows a strong affinity to the Mongolian gerbil Meriones unguiculatus. *Molecular biology and evolution*, 13(1): 3-6.

Akkers RC, Heeringen SJV, Jacobi UG, Janssen-Megens EM, Françoijs K-J, Stunnenberg HG, Veenstra GJC (2009). A Hierarchy of H3K4me3 and H3K27me3 Acquisition in Spatial Gene Regulation in Xenopus Embryos. *Developmental cell*, 17(3): 425–434.

Alexandre H (2001). A history of mammalian embryological research. *International journal of developmental biology*, 45(3), 457–468.

Alizadeh Z, Kageyama SI, Aoki F (2005). Degradation of maternal mRNA in mouse embryos: Selective degradation of specific mRNAs after fertilization. *Molecular reproduction and development*, 72(3): 281–290.

Allfrey VG, Faulkner R, Mirsky AE (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proceedings of the National Academy of Sciences*. 51(5): 786-94.

Ambrosi C, Manzo M, Baubec T (2017). Dynamics and context-dependent roles of DNA methylation. *Journal of molecular biology*, 429(10): 1459-75.

Aoki F, Hara KT, Schultz RM (2003). Acquisition of transcriptional competence in the 1-cell mouse embryo: Requirement for recruitment of maternal mRNAs. *Molecular reproduction and development*, 64(3): 270–274.

Aoki F, Worrad DM, Schultz RM (1997). Regulation of Transcriptional Activity during the First and Second Cell Cycles in the Preimplantation Mouse Embryo. *Developmental biology,* 181(2): 296–307.

Assheton R (1905). On the Foetus and Placenta of the Spiny Mouse (Acomys cahirinus). *Journal of zoology*, 75(3): 280-8.

Backs J, Stein P, Backs T, Duncan FE, Grueter CE, McAnally J, Qi X, Schultz RM, Olson EN (2009). The γ isoform of CaM kinase II controls mouse egg activation by regulating cell cycle resumption. *Proceedings of the National Academy of Sciences*, 107(1), 81–86.

Barlow DP, Stöger R, Herrmann BG, Saito K, Schweifer N (1991). The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus, *Nature* 349(6304): 84–87.

Barnes F, Eyestone W (1990). Early cleavage and the maternal zygotic transition in bovine embryos. *Theriogenology*, 33(1): 141–152.

Baroux C, Autran D, Gillmor C, Grimanelli D, Grossniklaus U (2008). The Maternal to Zygotic Transition in Animals and Plants. *Cold Spring Harbor Symposia on quantitative biology*, 73: 89–100.

Baroux C, Grossniklaus U (2015). The Maternal-to-Zygotic Transition in Flowering Plants. *Current topics in developmental biology*, 113: 351–371.

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007). High-resolution profiling of histone methylations in the human genome. *Cell*. 129(4): 823-37.

Bartolomei MS, Zemel S, Tilghman SM (1991). Parental imprinting of the mouse H19 gene. *Nature*, 351(6322): 153–155.

Bazzini AA, Lee MT, Giraldez AJ (2012). Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science*, 336(6078): 233–37.

Beaujean N, Hartshorne G, Cavilla J, Taylor J, Gardner J, Wilmut I, Meehan R, Young L (2004). Non-conservation of mammalian preimplantation methylation dynamics. *Current biology,* 14(7): R266–R267.

Becker M (2005). Differential in vivo binding dynamics of somatic and oocyte-specific linker histones in oocytes and during ES cell nuclear transfer. *Molecular biology of the cell*, 16(8): 3887–3895.

Bellofiore N, Ellery SJ, Mamrot J, Walker DW, Temple-Smith P, Dickinson H (2017). First evidence of a menstruating rodent: the spiny mouse (*Acomys cahirinus*). *American journal of obstetrics and gynecology*. 216(1): 40-e1.

Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A (2009). An operational definition of epigenetics. *Genes & development,* 23(7): 781–783.

Bianchi E, Doe B, Goulding D, Wright GJ (2014). Juno is the egg Izumo receptor and is essential for mammalian fertilization. *Nature*. 508(7497): 483-7.

Bianchi E, Wright GJ (2015). Cross-species fertilization: the hamster egg receptor, Juno, binds the human sperm ligand, Izumo1. *Philosophical transactions of the Royal Society B: biological sciences*, 370(1661): 20140101.

Bird A (2007). Perceptions of epigenetics. *Nature*. 447(7143): 396-8.

Birol I, Jackman SD, Nielsen CB, Qian JQ, Varhol R, Stazyk G, Morin RD, Zhao Y, Hirst M, Schein JE, Horsman DE, Connors JM, Gascoyne RD, Marra MA, Jones SJM (2009). De novo transcriptome assembly with ABySS. *Bioinformatics,* 25(21): 2872–2877.

Bleil JD, Wassarman PM (1980). Structure and function of the zona pellucida: Identification and characterization of the proteins of the mouse oocyte's zona pellucida. *Developmental biology* 76(1), 185–202.

Bloom K, Joglekar A (2010). Towards building a chromosome segregation machine. *Nature*. 463(7280): 446-56.

Bogdanović O, Gómez-Skarmeta JL (2014). Embryonic DNA methylation: insights from the genomics era. *Briefings in functional genomics*. 13(2): 121-30.

Bowman P, McLaren A (1970). Cleavage rate of mouse embryos in vivo and in vitro. *Development*, 24(1): 203-7.

Braude P, Bolton V, Moore S (1988). Human gene expression first occurs between the four-and eight-cell stages of preimplantation development. *Nature*. 332(6163): 459-61.

Brower PT, Gizang E, Boreen SM, Schultz RM (1981) Biochemical studies of mammalian oogenesis: Synthesis and stability of various classes of RNA during growth of the mouse oocyte in

vitro. *Developmental biology*, 86(2): 373–383.

Brunjes PC (1989). A comparative study of prenatal development in the olfactory bulb, neocortex and hippocampal region of the precocial mouse Acomys cahirinus and rat. *Developmental Brain Research*, 49(1):7-25.

Brunjes PC (1990). The precocial mouse, Acomys cahirinus. *Psychobiology*, 18(3):339-50.

Bultman SJ (2006). Maternal BRG1 regulates zygotic genome activation in the mouse. *Genes & development*, 20(13): 1744–1754.

Bunnell TM, Burbach BJ, Shimizu Y, Ervasti JM (2011). β-Actin specifically controls cell growth, migration, and the G-actin pool. *Molecular biology of the cell*, 22(21): 4047-58.

Burton A, Torres-Padilla ME (2014). Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. *Nature reviews molecular cell biology,* 15(11): 723–735.

Busa WB, Nuccitelli R (1985). An elevated free cytosolic Ca2  wave follows fertilization in eggs of the frog, Xenopus laevis. *The journal of cell biology*, 100(4): 1325–1329.

Cannata DJ, Ireland Z, Dickinson H, Snow RJ, Russell AP, West JM, Walker DW (2010). Maternal creatine supplementation from mid-pregnancy protects the diaphragm of the newborn spiny mouse from intrapartum hypoxia-induced damage. *Pediatric research*, 68:393-8.

Cassidy SB (1997). Prader-Willi syndrome. *Journal of medical genetics*, 34(11): 917–923.

Chavez SL, Loewke KE, Han J, Moussavi F, Colls P, Munne S, Behr B, Pera RAR (2012). Dynamic blastomere behaviour reflects human embryo ploidy by the four-cell stage. *Nature communications,* 3: 1251.

Chavez SL, McElroy SL, Bossert NL, De Jonge CJ, Rodriguez MV, Leong DE, Behr B, Westphal LM, Reijo Pera RA (2014). Comparison of epigenetic mediator expression and function in mouse and human embryonic blastomeres. *Human molecular genetics*, 23(18): 4970-84.

Chevret P, Denys C, Jaeger JJ, Michaux J, Catzeflis FM (1993). Molecular evidence that the spiny mouse (Acomys) is more closely related to gerbils (Gerbillinae) than to true mice (Murinae). *Proceedings of the National Academy of Sciences*, 90(8): 3433-6.

Christians E, Campion E, Thompson EM, Renard JP (1995). Expression of the HSP 70.1 gene, a landmark of early zygotic activity in the mouse embryo, is restricted to the first burst of transcription. *Development*, 121(1): 113-22.

Clark GF (2013). The role of carbohydrate recognition during human sperm-egg binding. *Human Reproduction*, 28(3): 566–577.

Clegg KB, Pikó L (1977). Size and specific activity of the UTP pool and overall rates of RNA synthesis in early mouse embryos. *Developmental biology*, 58(1): 76–95.

Clift D, Schuh M (2013). Restarting life: fertilization and the transition from meiosis to mitosis. *Nature Reviews Molecular Cell Biology*, 14(9): 549–562.

Collart C, Owens NDL, Bhaw-Rosun L, Cooper B, Domenico ED, Patrushev I, Sesay AK, Smith JN, Smith JC, Gilchrist MJ (2014). High-resolution analysis of gene activity during the Xenopus mid-blastula transition. *Development*, 141(9): 1927–1939.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science,* 339(6121): 819–823.

Davis JW, Sousa PAD, Schultz RM (1996). Transient Expression of Translation Initiation Factor eIF-4C during the 2-Cell Stage of the Preimplantation Mouse Embryo: Identification by mRNA Differential Display and the Role of DNA Replication in Zygotic Gene Activation. *Developmental Biology*, 174(2): 190–201.

Debaun MR, Niemitz EL, Feinberg AP (2003). Association of In Vitro Fertilization with Beckwith-Wiedemann Syndrome and Epigenetic Alterations of LIT1 and H19. *The American Journal of Human Genetics*, 72(1): 156–160.

Dechiara TM, Robertson EJ, Efstratiadis A (1991). Parental imprinting of the mouse insulin-like growth factor II gene. *Cell*, 64(4): 849–859.

De Leon VD, Johnson A, Bachvarova R (1983). Half-lives and relative amounts of stored and polysomal ribosomes and poly(A) RNA in mouse oocytes. *Developmental Biology*, 98(2): 400–408.

Denys C, Michaux J, Petter F, Aguilar JP, Jaeger JJ (1992). Molar morphology as a clue to the phylogenetic relationship of Acomys to the Murinae. *Israel journal of zoology*, 38(3-4): 253-62.

De Renzis S, Elemento O, Tavazoie S, Wieschaus EF (2007). Unmasking activation of the zygotic genome using chromosomal deletions in the Drosophila embryo. *PLoS Biology*, 5(5): e117.

Dickinson H, Walker DW, Cullen-McEwen L, Wintour, EM, Moritz K (2005). The spiny mouse

(*Acomys cahirinus*) completes nephrogenesis before birth. *American Journal of Physiology-Renal Physiology*, 289(2): F273–F279.

Dickinson H, Walker DW, Wintour EM, Moritz K (2007). Maternal dexamethasone treatment at midgestation reduces nephron number and alters renal gene expression in the fetal spiny mouse. *AJP: Regulatory, Integrative and Comparative Physiology*, 292(1): R453–R461.

Dieterlen F (1062). Geburt und Geburtshilfe bei der Stachelmaus, Acomys cahirinus. *Ethology*, 19(2): 191-222.

Ducibella T, Anderson E (1975). Cell shape and membrane changes in the eight-cell mouse embryo: Prerequisites for morphogenesis of the blastocyst. *Developmental Biology*, 47(1): 45–58.

Ducibella T, Huneau D, Angelichio E, Xu Z, Schultz RM, Kopf GS, Fissore R, Madoux S, Ozil J-P (2002). Egg-to-Embryo Transition Is Driven by Differential Responses to $Ca^{2+}$ Oscillation Number. *Developmental Biology*, 250(2): 280–291.

Ducibella T, Schultz R, Ozil J (2006). Role of calcium signals in early development. *Seminars in Cell & Developmental Biology*, 17(2): 324–332.

D'Udine B, Alleva E (1988). The Acomys cahirinus (spiny mouse) as a new model for biological and neurobehavioural studies. *Polish journal of pharmacology and pharmacy*, 40(5): 525-34.

Edgar B, Schubiger G (1986). Parameters controlling transcriptional activation during early drosophila development. *Cell*, 44(6): 871–877.

Edson MA, Nagaraja AK, Matzuk MM (2009). The Mammalian Ovary from Genesis to Revelation. *Endocrine Reviews*, 30(6): 624–712.

Ellery, S. J., Ireland, Z., Kett, M. M., Snow, R., Walker, D. W., & Dickinson, H (2013). Creatine pretreatment prevents birth asphyxia–induced injury of the newborn spiny mouse kidney. *Pediatric research*, 73(2): 201.

Fenelon JC, Shaw G, O'Neill C, Frankenberg S, Renfree MB (2013). Paf receptor expression in the marsupial embryo and endometrium during embryonic diapause. *Reproduction*, 147(1): 21–31.

Fleiss B, Coleman HA, Castillo-Melendez M, Ireland Z, Walker DW, Parkington HC (2011). Effects of birth asphyxia on neonatal hippocampal structure and function in the spiny mouse. *International Journal of Developmental Neuroscience*, 29(7):757-66.

Flyamer IM, Gassler J, Imakaev M, Brandão HB, Ulianov SV, Abdennur N, Razin SV, Mirny LA, Tachibana-Konwalski K (2017). Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature*, 544(7648): 110-4.

Flynn LJ, Jacobs LL, Lindsay EH (1985). Problems in Muroid Phylogeny: Relationship to Other Rodents and Origin of Major Groups. *Evolutionary Relationships among Rodents*, 589–616.

Frei RE, Schultz GA, Church RB (1989). Qualitative and quantitative changes in protein synthesis occur at the 8-16-cell stage of embryogenesis in the cow. *Reproduction*, 86(2): 637–641.

Fuks F, Burgers WA, Brehm A, Hughes-Davies L, Kouzarides T (2000). DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nature genetics*, 24(1): 88–91.

Fushan AA, Turanov AA, Lee SG, Kim EB, Lobanov AV, Yim SH, Buffenstein R, Lee SR, Chang KT, Rhee H, Kim JS (2015). Gene expression defines natural changes in mammalian lifespan. *Aging cell*, 14(3): 352-65.

Ge SX. (2017). Exploratory bioinformatics investigation reveals importance of “junk” DNA in early embryo development. *BMC genomics*. 18(1): 200.

Goddard MJ, Pratt HP (1983). Control of events during early cleavage of the mouse embryo: an analysis of the ‘2-cell block’. *Journal of embryology and experimental morphology*, 73(1): 111–133.

Golbus MS, Calarco PG, Epstein CJ (1973). The effects of inhibitors of RNA synthesis (α-amanitin and actinomycin D) on preimplantation mouse embryogenesis. *Journal of Experimental Zoology*, 186(2): 207–216.

Goldberg AD, Allis CD, Bernstein E (2007). Epigenetics: A Landscape Takes Shape. *Cell*, 128(4): 635–638.

Gonet AE, Stauffacher W, Pictet R, Renold AE (1966). Obesity and diabetes mellitus with striking congenital hyperplasia of the islets of langerhans in spiny mice (Acomys Cahirinus). *Diabetologia*, 1: 162–171.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Palma FD, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol Nature Biotechnology,* 29(7): 644–652.

Groppo R, Richter JD (2009). Translational control from head to tail. *Current Opinion in Cell Biology*, 21(3): 444–451.

Gu Y, Shen X, Zhou D, Wang Z, Zhang N, Shan Z, Jin L, Lei L, (2014). Selection and expression profiles of reference genes in mouse preimplantation embryos of different ploidies at various developmental stages. *PloS One*, 9(6), p.e98956.

Guo H, Ping Z, Liying Y, Rong L, Boqiang H, Ying L, Jie Y , Boqiang H, Lian Y, Yan J, Ren X, Lin S, Li J, Jin X, Shi X, Liu P, Wang X, Wang W, Wei Y, Li X, Guo F, Wu X, Fan X, Yong J, Wen L, Xie SX, Tang F, Qiao J (2014). The DNA methylation landscape of human early embryos. *Nature*, 511(7511): 606–610.

Gupta SK (2015). Role of zona pellucida glycoproteins during fertilization in humans. *Journal of Reproductive Immunology,* 108: 90–97.

Haberle V, Li N, Hadzhiev Y, Plessy C, Previti C, Nepal C, Gehrig J, Dong X, Akalin A, Suzuki AM, Ijcken WFJV, Armant O, Ferg M, Strähle U, Carninci P, Müller F, Lenhard B (2014). Two independent transcription initiation codes overlap on vertebrate core promoters. *Nature*, 507(7492): 381–385.

Hamatani T, Carter MG, Sharov AA, Ko MS (2004). Dynamics of Global Gene Expression Changes during Mouse Preimplantation Development. *Developmental Cell*, 6(1): 117–131.

Hamoir G (1992). The discovery of meiosis by E. Van Beneden, a breakthrough in the morphological phase of heredity. *The International Journal of Developmental Biology,* 36(1): 9–15.

Harlow GM, Quinn P (1982). Development of preimplantation mouse embryos in vivo and in vitro. *Australian journal of biological sciences*, 35(2): 187–194.

Harrison MM, Eisen MB (2015). Transcriptional Activation of the Zygotic Genome in Drosophila. *Current Topics in Developmental Biology*. 113: 85-112.

Harrison MM, Li X-Y, Kaplan T, Botchan MR, Eisen MB (2011). Zelda Binding in the Early Drosophila melanogaster Embryo Marks Regions Subsequently Activated at the Maternal-to-Zygotic Transition. *PLoS Genetics*, 7(10): e1002266.

Harvey SA, Sealy I, Kettleborough R, Fenyes F, White R, Stemple D, Smith JC (2013). Identification of the zebrafish maternal and paternal transcriptomes. *Development*, 140(13): 2703–2710.

Haughton CL, Gawriluk TR, Seifert AW (2016). The biology and husbandry of the African spiny mouse (Acomys cahirinus) and the research uses of a laboratory colony. *Journal of the American Association for Laboratory Animal Science*. 55(1): 9-17.

Hedges SB, Dudley J, Kumar S (2006). TimeTree: a public knowledge-base of divergence times among organisms, *Bioinformatics* 22(3): 2971–2972.

Hedges SB, Marin J, Suleski M, Paymer M, Kumar S (2015). Tree of Life Reveals Clock-Like Speciation and Diversification. *Molecular Biology and Evolution*, 32(4): 835–845.

Hermann A, Goyal R, Jeltsch A (2004). The Dnmt1 DNA - (cytosine-C5) - methyltransferase Methylates DNA Processively with High Preference for Hemimethylated Target Sites. Journal of Biological Chemistry, 279(46): 48350–48359.

Heyn P, Kircher M, Dahl A, Kelso J, Tomancak P, Kalinka AT, Neugebauer KM (2014). The Earliest Transcribed Zygotic Genes Are Short, Newly Evolved, and Different across Species. *Cell Reports*, 6(2): 285–292.

Holliday R, Pugh JE (1975). DNA modification mechanisms and gene activity during development. *Science* 187, 226–232.

Holliday R, Pugh JE (1996). DNA modification mechanisms and gene activity during development. *Cold Spring Harbor Monograph Archive*, 32: 639–645.

Houseley J, Tollervey D (2009). The Many Pathways of RNA Degradation. *Cell*, 136(4): 763–776.

Hutton LC, Ratnayake U, Shields A, Walker DW (2009). Neuropathology and functional deficits in a model of birth asphyxia in the precocial spiny mouse (*Acomys cahirinus*). *Developmental neuroscience*, 31(6): 523–35.

Ikawa M, Inoue N, Benham AM, Okabe M (2010). Fertilization: a sperm's journey to and interaction with the oocyte. *Journal of Clinical Investigation*, 120(4), 984–994.

Iurlaro M, von Meyenn F, Reik W (2017). DNA methylation homeostasis in human and mouse development. *Current Opinion in Genetics & Development*, 43: 101-9.

Izzo A, Kamieniarz K, Schneider, R (2008). The histone H1 family: specific members, specific functions? *Biological chemistry*, 389(4), 333-343.

Jahn CL, Baran MM, Bachvarova R (1976). Stability of RNA synthesized by the mouse oocyte during its major growth phase. *Journal of Experimental Zoology*, 197(2): 161–171.

Jiang L, Zhang J, Wang J-J, Wang L, Zhang L, Li G, Yang X, Ma X, Sun X, Cai J, Zhang J, Huang X, Yu M, Wang X, Liu F, Wu C-I, He C, Zhang B, Ci W, Liu J (2013). Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Cell*, 153(4): 773–784.

Jin XL, O'Neill C (2011). Regulation of the Expression of Proto-Oncogenes by Autocrine Embryotropins in the Early Mouse Embryo. *Biology of Reproduction*, 84(6): 1216–1224.

Jones PA (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, 13(7): 484-92.

Jones M, Dayan T (2000). Foraging behavior and microhabitat use by spiny mice, Acomys cahirinus and A. russatus, in the presence of Blanford's fox (Vulpes cana) odor. *Journal of Chemical Ecology*, 26(2): 455–469.

Joseph SR, Pálfy M, Hilbert L, Kumar M, Karschau J, Zaburdaev V, Shevchenko A, Vastenhouw NL (2017). Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos. *eLife*. 6: e23326.

Kafri T, Ariel M, Brandeis M, Shemer R, Urven L, Mccarrey J, Cedar H, Razin A (1992). Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes & Development*, 6(5): 705–714.

Kaji K, Kudo A (2004). The mechanism of sperm-oocyte fusion in mammals. *Reproduction*, 127(4): 423–429.

Kaneuchi T, Sartain CV, Takeo S, Horner VL, Buehner NA, Aigaki T, Wolfner MF (2015). Calcium waves occur as Drosophila oocytes activate. *Proceedings of the National Academy of Sciences*, 112(3): 791–796.

Kebede AF, Schneider R, Daujat S (2015). Novel types and sites of histone modifications emerge as players in the transcriptional regulation contest. *FEBS journal*. 282(9): 1658-74.

Kimelman D, Kirschner M, Scherson T (1987). The events of the midblastula transition in Xenopus are regulated by changes in the cell cycle. *Cell*, 48(3): 399–407.

Knowland J, Graham C (1972). RNA synthesis at the two-cell stage of mouse development. *Development*. 27(1): 167-76.

Ko MS, Kitchen JR, Wang XT, Wang XH, Sun T, Grahovac MJ, Kargul GJ, Lim MK, Cui Y, Sano Y, Tanaka T, Liang Y, Mason S, Paonessa PD, Sauls AD, DePalma GE, Sharara R, Rowe LB,

Eppig J, Morrell C, Doi H (2000). Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development. *Development*, 127(8): 1737–1749.

Kohli RM, Zhang Y (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*, 502(7472): 472–479.

Kornberg RD (1974). Chromatin structure: a repeating unit of histones and DNA. *Science*. 184(4139): 868-71.

Kouzarides T (2007). Chromatin modifications and their function. *Cell*. 128(4): 693-705.

Kurimoto K, Yabuta Y, Hayashi K, Ohta H, Kiyonari H, Mitani T, Moritoki Y, Kohri K, Kimura H, Yamamoto T, Katou Y, Shirahige K, Saitou M (2015). Quantitative Dynamics of Chromatin Remodeling during Germ Cell Specification from Mouse Embryonic Stem Cells. *Cell Stem Cell,* 16(5): 517–532.

Lamers WH, Boon L, van Hemert FJ, Labruyere WT, de Jong P, Ruijter JM, Moorman AF (1999). Glutamine synthetase expression in perinatal spiny mouse liver. *The FEBS Journal*, 262(3):803-9.

Lamers WH, Mooren PG, Graaf A, Charles R (1985). Perinatal development of the liver in rat and spiny mouse. *European Journal of Biochemistry*, 146(2): 475–80.

Lamers WH, Mooren PG, Griep HA, Endert ER, Degenhart HJ, Charles RO (1986). Hormones in perinatal rat and spiny mouse: relation to altricial and precocial timing of birth. *American Journal of Physiology-Endocrinology and Metabolism*, 251(1): E78–85.

Latham KE, Schultz RM (2001). Embryonic genome activation. *Frontiers in Bioscience*. 6: D748-759.

Latham KE, Solter D, Schultz RM (1992). Acquisition of a transcriptionally permissive state during the 1-cell stage of mouse embryogenesis. *Developmental biology*. 149(2): 457-62.

Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, 11(3): 204–220.

Lawrence M, Daujat S, Schneider R (2016). Lateral thinking: how histone modifications regulate gene expression. *Trends in Genetics*. 32(1): 42-56.

Lazaraviciute G, Kauser M, Bhattacharya S, Haggarty P, Bhattacharya S (2014) A systematic review and meta-analysis of DNA methylation levels and imprinting disorders in children conceived

by IVF/ICSI compared with children conceived spontaneously. *Human Reproduction Update,* 20(6): 840–852.

Lee MT, Bonneau AR, Giraldez AJ (2014) Zygotic Genome Activation During the Maternal-to-Zygotic Transition. *Annual Review of Cell and Developmental Biology*, 30: 581–613.

Lee MT, Bonneau AR, Takacs CM, Bazzini AA, Divito KR, Fleming ES, Giraldez AJ (2013) Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature*, 503(7476): 360–364.

Lefievre L (2004) Four zona pellucida glycoproteins are expressed in the human. *Human Reproduction,* 19(7): 1580–1586.

Lei M, Zhang H, Julian R, Tang K, Xie S, Zhu J-K (2015) Regulatory link between DNA methylation and active demethylation in Arabidopsis. *Proceedings of the National Academy of Sciences*, 112(11): 3553–3557.

Leichsenring M, Maes J, Mossner R, Driever W, Onichtchouk D (2013) Pou5f1 Transcription Factor Controls Zygotic Gene Activation In Vertebrates. *Science*, 341(6149): 1005–1009.

Lewis WH, Gregory PW (1929) Cinematographs Of Living Developing Rabbit-Eggs. *Science,* 69(1782): 226–229.

Li R, Albertini DF (2014). The road to maturation: somatic cell interaction and self-organization of the mammalian oocyte. *Nature reviews Molecular cell biology*. 14(3): 141.

Li X, Ito M, Zhou F, Youngson N, Zuo X, Leder P, Ferguson-Smith AC (2008) A Maternal-Zygotic Effect Gene, Zfp57, Maintains Both Maternal and Paternal Imprints. *Developmental Cell*, 15(4): 547–557.

Lindell TJ, Weinberg F, Morris PW, Roeder RG, Rutter WJ (1970) Specific Inhibition of Nuclear RNA Polymerase II by agr-Amanitin. *Science*, 170(3956): 447–449.

Loke YJ, Craig JM (2017). Are the effects of IVF on DNA methylation driven by intracytoplasmic sperm injection and male infertility? *Epigenomics*, 8(7): 881-884.

Lorch Y, LaPointe JW, Kornberg RD (1987). Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell*. 49(2): 203-10.

Lord EM, Russell SD (2002). The mechanisms of pollination and fertilization in plants. *Annual Review of Cell and Developmental Biology*, 18(1): 81–105.

Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*. 389(6648): 251-60.

Lykke-Andersen K, Gilchrist MJ, Grabarek JB, Das P, Miska E, Zernicka-Goetz M (2008) Maternal Argonaute 2 Is Essential for Early Mouse Development at the Maternal-Zygotic Transition. *Molecular Biology of the Cell*, 19(10): 4383–4392.

Ma J, Flemr M, Strnad H, Svoboda P, Schultz RM (2012) Maternally Recruited DCP1A and DCP2 Contribute to Messenger RNA Degradation During Oocyte Maturation and Genome Activation in Mouse. *Biology of Reproduction*, 88(1): 11–11.

Maldera JA, Munoz MW, Chirinos M, Busso D, Raffo FG, Battistone MA, Blaquier JA, Larrea F, Cuasnicu PS (2013) Human fertilization: epididymal hCRISP1 mediates sperm-zona pellucida binding through its interaction with ZP3. *Molecular Human Reproduction,* 20(4): 341–349.

Martin Y, Gerlach G, Schlötterer C, Meyer A (2000). Molecular phylogeny of European muroid rodents based on complete cytochrome b sequences. *Molecular phylogenetics and evolution*, 16(1):37-47.

Massah S, Beischlag, TV, Prefontaine GG (2015) Epigenetic events regulating monoallelic gene expression. *Critical reviews in biochemistry and molecular biology*, 50(4): 337–358.

Mayer W, Niveleau A, Walter J, Fundele R, Haaf T (2000) Embryogenesis: Demethylation of the zygotic paternal genome. *Nature*, 403(6769): 501–502.

Medvedev S, Yang J, Hecht NB, Schultz RM (2008) CDC2A (CDK1)-mediated phosphorylation of MSY2 triggers maternal mRNA degradation during mouse oocyte maturation. *Developmental Biology*, 321(1): 205–215.

Mehlmann LM (2012) Losing Mom's Message: Requirement for DCP1A and DCP2 in the Degradation of Maternal Transcripts During Oocyte Maturation. *Biology of Reproduction*, 88(1): 10–10.

Meirelles F, Caetano A, Watanabe Y, Ripamonte P, Carambula S, Merighe G, Garcia S (2004) Genome activation and developmental block in bovine embryos. *Animal Reproduction Science* 82-83, 13–20.

Messerschmidt DM, Vries WD, Ito M, Solter D, Ferguson-Smith A, Knowles BB (2012) Trim28 Is Required for Epigenetic Stability During Mouse Oocyte to Embryo Transition. *Science*, 335(6075): 1499–1502.

Meyerowitz EM (2002) Plants Compared to Animals: The Broadest Comparative Study of Development, *Science* 295(5559): 1482–1485.

McGee EA, Hsueh AJ (2000) Initial and cyclic recruitment of ovarian follicles 1. Endocrine reviews, 21(2): 200–214.

Mcgrath J, Solter D (1984) Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell*, 37(1): 179–183.

Monk M, Boubelik M, Lehnert S (1987) Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development*, 99(3): 371–382.

Montandon SA, Tzika AC, Martins AF, Chopard B, Milinkovitch MC (2014). Two waves of anisotropic growth generate enlarged follicles in the spiny mouse. *EvoDevo*, 5(1): 33.

Morgan HD, Santos F, Green K, Dean W, Reik W (2005). Epigenetic reprogramming in mammals. *Human molecular genetics*. 14(suppl 1): R47-58.

Murrell A, Heeson S, Reik W (2004) Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nature genetics*, 36(8): 889–893.

Nakamura T, Liu Y-J, Nakashima H, Umehara H, Inoue K, Matoba S, Tachibana M, Ogura A, Shinkai Y, Nakano T (2012) PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature*, 486(7403): 415–419.

Newport J, Kirschner M (1982a) A major developmental transition in early xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. *Cell*, 30(3): 675–686.

Newport J, Kirschner M (1982b) A major developmental transition in early xenopus embryos: II. control of the onset of transcription. *Cell,* 30(3): 687–696.

Niakan KK, Han J, Pedersen RA, Simon C, Pera RA (2012). Human pre-implantation embryo development. *Development*, 139(5): 829-41.

O'Connell BA, Moritz KM, Roberts CT, Walker DW, Dickinson H (2011) The placental response to excess maternal glucocorticoid exposure differs between the male and female conceptus in spiny mice. *Biology of reproduction*, 85(5): 1040–7.

O'Farrell P (2004) Embryonic Cleavage Cycles: How Is a Mouse Like a Fly? *Current Biology*,

14(1): R35–R45.

Okano M, Bell DW, Haber DA, Li E (1999) DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell,* 99(3): 247–257.

Olsen OA (2001) Endosperm development: cellularization and cell fate specification. *Annual Review of Plant Physiology and Plant Molecular Biology,* 52(1): 233–267.

O'Neill C, Li Y, Jin X (2012) Survival signaling in the preimplantation embryo. *Theriogenology,* 77(4): 773–784.

O'Neill C, Li Y, Jin X (2015) Survival Signalling in the Preimplantation Embryo. In *Cell Signaling During Mammalian Early Embryo Development*, (pp. 129–149), Springer, New York.

Oosterhuis WP, Mooren PG, Charles R, Lamers WH (1984). Perinatal development of the lung in rat and spiny mouse: its relation to altricial and precocial timing of birth. *Neonatology*, 45(5): 236-43.

Osborn DJ, Helmy I (1980) *The contemporary land mammals of Egypt (including Sinai)* Field Museum of Natural History, Chicago, IL.

Oswald J, Engemann S, Lane N, Mayer W, Olek A, Fundele R, Dean W, Reik W, Walter J (2000) Active demethylation of the paternal genome in the mouse zygote. *Current Biology*, 10(8): 475–478.

Ozil JP, Banrezes B, Tóth S, Pan H, Schultz RM (2006) $Ca^{2+}$ oscillatory pattern in fertilized mouse eggs affects gene expression and development to term. *Developmental Biology*, 300(2): 534–544.

Pan H, O'Brien MJ, Wigglesworth K, Eppig JJ, Schultz RM (2005) Transcript profiling during mouse oocyte development and the effect of gonadotropin priming and development in vitro. *Developmental Biology*, 286(2): 493–506.

Pan H, Schultz RM (2011) SOX2 Modulates Reprogramming of Gene Expression in Two-Cell Mouse Embryos. *Biology of Reproduction*, 85(2): 409–416.

Paranjpe SS, Jacobi UG, van Heeringen SJ, Veenstra GJC (2013) A genome-wide survey of maternal and embryonic transcripts during Xenopus tropicalis development. *BMC Genomics*, 14(1): 762.

Paynton BV, Rempel R, Bachvarova R (1988) Changes in state of adenylation and time course of degradation of maternal mRNAs during oocyte maturation and early embryonic development in the mouse. *Developmental Biology,* 129(2): 304–314.

Peaston AE, Evsikov AV, Graber JH, Vries WND, Holbrook AE, Solter D, Knowles BB (2004) Retrotransposons Regulate Host Genes in Mouse Oocytes and Preimplantation Embryos. *Developmental Cell*, 7(4): 597–606.

Peng H, Chang B, Lu C, Su J, Wu Y, Lv P, Wang Y, Liu J, Zhang B, Quan F, Guo Z Zhang Y (2012) Nlrp2, a Maternal Effect Gene Required for Early Embryonic Development in the Mouse. *PLoS ONE*, 7(1): e30344.

Piko L, Clegg KB (1982) Quantitative changes in total RNA, total poly(A), and ribosomes in early mouse embryos. *Developmental Biology*, 89(2): 362–378.

Pincus G (1930) Observations on the Living Eggs of the Rabbit. *Proceedings of the Royal Society B: Biological Sciences,* 107(749): 132–167.

Plachta N, Bollenbach T, Pease S, Fraser SE, Pantazis P (2011) Oct4 kinetics predict cell lineage patterning in the early mammalian embryo. *Nature Cell Biology,* 13(2): 117–123.

Porter RH, Doane HM (1976). Maternal pheromone in the spiny mouse (Acomys cahirinus). *Physiology & behavior*. 16(1): 75-8.

Potok ME, Nix DA, Parnell TJ, Cairns BR (2013) Reprogramming the Maternal Zebrafish Genome after Fertilization to Match the Paternal Methylation Pattern. *Cell*, 153(4): 759–772.

Puschendorf M, Stein P, Oakeley EJ, Schultz RM, Peters AH, Svoboda P (2006) Abundant transcripts from retrotransposons are unstable in fully grown mouse oocytes. *Biochemical and Biophysical Research Communications*, 347(1): 36–43.

Quinn TA, Ratnayake U, Castillo-Melendez M, Moritz KM, Dickinson H, Walker DW (2014). Adrenal steroidogenesis following prenatal dexamethasone exposure in the spiny mouse. *Journal of endocrinology*, 221(2): 347-62.

Quinn TA, Ratnayake U, Dickinson H, Nguyen TH, McIntosh M, Castillo-Melendez M, Conley AJ, Walker DW (2013). Ontogeny of the adrenal gland in the spiny mouse, with particular reference to production of the steroids cortisol and dehydroepiandrosterone. *Endocrinology,*154(3): 1190-201.

Rabani M, Raychowdhury R, Jovanovic M, Rooney M, Stumpo DJ, Pauli A, Hacohen N, Schier AF, Blackshear PJ, Friedman N, Amit I, Regev A (2014) High-Resolution Sequencing and Modeling

Identifies Distinct Dynamic RNA Regulatory Strategies. *Cell*, 159(7): 1698–1710.

Rajjou L, Gallardo K, Debeaujon I, Vandekerckhove J, Job C, Job D (2004) The Effect of α-Amanitin on the Arabidopsis Seed Proteome Highlights the Distinct Roles of Stored and Neosynthesized mRNAs during Germination. *Plant Physiology*, 134(4): 1598–1613.

Rankin T, Familari M, Lee E, Ginsberg A, Dwyer N, Blanchette-Mackie J, Drago J, Westphal H, Dean J (1996) Mice homozygous for an insertional mutation in the Zp3 gene lack a zona pellucida and are infertile. *Development*, 122(9): 2903–2910.

Rankin T, O'Brien M, Lee E, Wigglesworth KEJJ, Eppig J, Dean J (2001) Defective zonae pellucidae in Zp2-null mice disrupt folliculogenesis, fertility and development. *Development*, 128(7): 1119–1126.

Rankin T, Talbot P, Lee E, Dean J (1999). Abnormal zonae pellucidae in mice lacking ZP1 result in early embryonic loss. *Development*, 126(17): 3847–3855.

Ratnayake U, Quinn TA, Castillo-Melendez M, Dickinson H, Walker DW (2012). Behaviour and hippocampus-specific changes in spiny mouse neonates after treatment of the mother with the viral-mimetic Poly I: C at mid-pregnancy. *Brain, behavior, and immunity*, 26(8): 1288–99.

Reik W, Dean W, Walter J (2001). Epigenetic reprogramming in mammalian development. *Science*, 293(5532): 1089-93.

Reik W (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature,* 447(7143): 425–432.

Ridgway EB, Gilkey JC, Jaffe LF (1977). Free calcium increases explosively in activating medaka eggs. *Proceedings of the National Academy of Sciences*, 74(2): 623–627.

Robertson KD, Wolffe AP (2000). DNA methylation in health and disease. *Nature Reviews Genetics*, 1(1): 11–19.

Rougier N, Bourc'his D, Gomes DM, Niveleau A, Plachot M, Pàldi A, Viegas-Péquignot E (1998). Chromosome methylation patterns during mammalian preimplantation development. Genes & development, 12(14): 2108-13.

Russo VEA, Martienssen RA, Riggs AD (1996). *Epigenetic mechanisms of gene regulation.* Cold Spring Harbor Laboratory Press, Plainview, NY.

Santos F, Dean W (2004). Epigenetic reprogramming during early development in mammals.

*Reproduction*, 127(6): 643–651.

Santos F, Hendrich B, Reik W, Dean W (2002). Dynamic Reprogramming of DNA Methylation in the Early Mouse Embryo. *Developmental Biology*, 241(1): 172–182.

Sarich VM (1985). Rodent Macromolecular Systematics. *Evolutionary relationships among rodents*, (pp. 423–452), Springer, US.

Sawicki JA, Magnuson T, Epstein CJ (1981). Evidence for expression of the paternal genome in the two-cell mouse embryo. *Nature*. 294: 450-451.

Schier AF (2007). The maternal-zygotic transition: death and birth of RNAs. *Science*. 20, 316(5823): 406-7.

Schultz RM (1993). Regulation of zygotic gene activation in the mouse. *Bioessays*, 15(8): 531–538.

Seifert AW, Kiama SG, Seifert MG, Goheen JR, Palmer TM, Maden M (2012). Skin shedding and tissue regeneration in African spiny mice (Acomys). *Nature*, 489(7417): 561.

Shafrir E, Adler JH (1983). Enzymatic and metabolic responses to affluent diet of two diabetes-prone species of spiny mice: Acomys cahirinus and Acomys russatus. *International Journal of Biochemistry*, 15(12): 1439–1446.

Simpson GG (1945). The principles of classification and a classification of mammals. *Bull. Amer. Museum Nat. History:* 85.

Skinner JD and Chimimba CT (2005). *The Mammals of the Southern African Sub-region* Cambridge University Press, New York.

Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, Regev A, Meissner A (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* 484(7394): 339–344.

Song S, Ghosh J, Mainigi M, Turan N, Weinerman R, Truongcao M, Coutifaris C, Sapienza C (2015). DNA methylation differences between in vitro- and in vivo-conceived children are associated with ART procedures rather than infertility. *Clinical Epigenetics,* 7(1): 41.

Steinhardt R, Zucker R, Schatten G (1977). Intracellular calcium release at fertilization in the sea urchin egg. *Developmental Biology*, 58(1): 185–196.

Stelzer Y, Shivalila, CS, Soldner, F, Markoulaki, S, Jaenisch, R (2015). Tracing dynamic changes of DNA methylation at single-cell resolution. *Cell*, 163(1): 218–229.

Su Y-Q, Sugiura K, Woo Y, Wigglesworth K, Kamdar S, Affourtit J, Eppig JJ (2007). Selective degradation of transcripts during meiotic maturation of mouse oocytes. *Developmental Biology*, 302(1): 104–117.

Surani MAH, Barton, Norris ML (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature*, 308(5959): 548–550.

Svoboda P, Franke V, Schultz RM (2015). Sculpting the Transcriptome During the Oocyte-to-Embryo Transition in Mouse. *Current Topics in Developmental Biology*, 113: 305–349.

Tadros W, Lipshitz HD (2009). The maternal-to-zygotic transition: a play in two acts. *Development*, 136(18): 3033-3042.

Tanaka M, Hennebold JD, Macfarlane J, Adashi EY (2001). A mammalian oocyte-specific linker histone gene H1oo: homology with the genes for the oocyte-specific cleavage stage histone (cs-H1) of sea urchin and the B4/H1M histone of the frog. *Development*, 128(5), 655–664.

Tang C, Toomajian C, Sherman-Broyles S, Plagnol V, Guo Y-L, Hu TT, Clark RM, Nasrallah JB, Weigel D, Nordborg M (2007). The Evolution of Selfing in Arabidopsis thaliana. *Science*, 317(5841): 1070–1072.

Telfer EE, Mclaughlin M (2007). Natural history of the mammalian oocyte. *Reproductive BioMedicine Online*, 15(3): 288–295.

Telford NA, Watson AJ, Schultz GA (1990). Transition from maternal to embryonic control in early mammalian development: a comparison of several species. *Molecular Reproduction and Development*, 26(1): 90–100.

Thélie A, Papillier P, Pennetier S, Perreau C, Traverso J, Uzbekova S, Mermillod P, Joly C, Humblot P, Dalbiès-Tran R (2007). Differential regulation of abundance and deadenylation of maternal transcripts during bovine oocyte maturation in vitro and in vivo. *BMC Developmental Biology*, 7(1): 125.

Thomas S, Li X-Y, Sabo PJ, Sandstrom R, Thurman RE, Canfield TK, Giste E, Fisher W, Hammonds A, Celniker SE, Biggin MD, Stamatoyannopoulos JA (2011). Dynamic reprogramming of chromatin accessibility during Drosophila embryo development. *Genome Biology*, 12(5): R43.

Tiemann U (2008) The Role of Platelet-activating Factor in the Mammalian Female Reproductive Tract. *Reproduction in Domestic Animals*, 43(6): 647–655.

Töhönen V, Katayama S, Vesterlund L, Jouhilahti E-M, Sheikhi M, Madissoon E, Filippini-Cattaneo

G, Jaconi M, Johnsson A, Bürglin TR, Linnarsson S, Hovatta O, Kere J (2015). Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development. *Nature Communications*, 6: 8207.

Turan N, Katari S, Gerson LF, Chalian R, Foster MW, Gaughan JP, Coutifaris C, Sapienza C (2010). Inter- and Intra-Individual Variation in Allele-Specific DNA Methylation and Gene Expression in Children Conceived using Assisted Reproductive Technology. *PLoS Genetics,* 6(7): e1001033.

Van Beneden É (1870). Recherches sur la composition et la signification de l'oeuf. Hayez. l'Académie Royale de Belgique 4.

Van Beneden É (1875). La maturation de l'œuf, la fécondation, et les premières phases du développement embryonnaire des mammifères d'après des recherches faites chez le lapin. *Bulletins de l'Académie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique*. 40(7): 686–736

Van Beneden É (1883). Recherches sur la maturation de l'oeuf, la fécondation, et la division cellulaire. Librairie Clemm.

Vassena R, Boue S, Gonzalez-Roca E, Aran B, Auer H, Veiga A, Belmonte JCI (2011). Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. *Development,* 138(7): 3699–3709.

Verona RI, Mann MR, Bartolomei MS (2003). Genomic imprinting: intricacies of epigenetic regulation in clusters. *Annual review of cell and developmental biology*, 19(1): 237–259.

Von Baer KE (1827). De ovi mammalium et hominis genesi. *Epistolam ad Academiam Imperialem Scientiarum Petropolitanam*. L Vossi, Leipzig.

Walker E (1968). Mammals of the World. Baltimore, MD: Johns Hopkins.

Warner CM, Versteegh LR (1974). *In vivo* and *in vitro* effect of α-amanitin on preimplantation mouse embryo RNA polymerase. *Nature* 248, 678–680.

Watanabe D, Suetake I, Tada T, Tajima S (2002). Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. *Mechanisms of Development,* 118(1): 187–190.

Weksberg R, Shen DR, Fei YL, Song QL, Squire J (1993). Disruption of insulin–like growth factor 2 imprinting in Beckwith–Wiedemann syndrome. *Nature Genetics*, 5(2): 143–150.

Wells XE, O'Neill C (1994). Detection and preliminary characterization of two enzymes involved in biosynthesis of platelet-activating factor in mouse oocytes, zygotes and preimplantation embryos: dithiothreitol-insensitive cytidinediphosphocholine: 1-O-alkyl-2-acetyl-sn-glycerol cholinephosphotransferase and acetyl-coenzyme A:1-O-alkyl-2-lyso-sn-glycero-3-phosphocholine acetyltransferase. *Reproduction*, 101(2): 385–391.

Wessel AD, Gumalla M, Grosshans J, Schmidt CF (2015). The Mechanical Properties of Early Drosophila Embryos Measured by High-Speed Video Microrheology. *Biophysical Journal,* 108(8): 1899–1907.

Whitelaw E, Garrick D (2005) The epigenome: epigenetic regulation of gene expression in mammalian species, in *Mammalian genomics* (pp. 179–200). CABI Publishing, Wallingford, UK.

Wiekowski M, Miranda M, Depamphilis ML (1991). Regulation of gene expression in preimplantation mouse embryos: Effects of the zygotic clock and the first mitosis on promoter and enhancer activities. *Developmental Biology*, 147(2): 403–414.

Wiekowski M, Miranda M, Depamphilis ML (1993). Requirements for Promoter Activity in Mouse Oocytes and Embryos Distinguish Paternal Pronuclei from Maternal and Zygotic Nuclei. *Developmental Biology*, 159(1): 366–378.

Wilson DE, Reeder DM, eds. (2005). Mammal species of the world: a taxonomic and geographic reference. JHU Press.

Wong CC, Loewke KE, Bossert NL, Behr B, Jonge CJD, Baer TM, Pera RAR (2010). Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst stage. *Nature Biotechnology,* 28(10): 1115–1121.

Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, Boiani M, Arand J, Nakano T, Reik W, Walter J (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nature Communications,* 2: 241.

Wray GA (2001). Dating branches on the tree of life using DNA. *Genome biology*, 3(1): 0001-1.

Wu SC, Zhang Y (2010) Active DNA demethylation: many roads lead to Rome. *Nature Reviews Molecular Cell Biology*, 11(9): 750–750.

Wu X, Viveiros MM, Eppig JJ, Bai Y, Fitzpatrick SL, Matzuk MM (2003). Zygote arrest 1 (Zar1) is a novel maternal-effect gene critical for the oocyte-to-embryo transition. *Nature genetics*. 33(2): 187-91.

Xu Z, Williams CJ, Kopf GS, Schultz RM (2003). Maturation-associated increase in IP3 receptor type 1: role in conferring increased IP3 sensitivity and Ca2  oscillatory behavior in mouse eggs. *Developmental Biology*, 254(2): 163–171.

Xue Z, Huang K, Cai C, Cai L, Jiang C-Y, Feng Y, Liu Z, Zeng Q, Cheng L, Sun YE, Liu J-Y, Horvath S, Fan G (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature,* 500(7464): 593–597.

Yandell M, Ence D (2012). A beginner's guide to eukaryotic genome annotation. *Nature Reviews Genetics*, 13(5): 329–342.

Yoder JA, Walsh CP, Bestor TH (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends in genetics*, 13(8): 335–340.

Young DA (1976). Breeding and fertility of the Egyptian spiny mouse, Acomys cahirinus: effect of different environments. *Laboratory animals*, 10(1): 15-24.

Yu J (2001). Expression of MSY2 in Mouse Oocytes and Preimplantation Embryos. *Biology of Reproduction*, 65(4): 1260–1270.

Yu J, Hecht NB, Schultz RM (2003). Requirement for RNA-binding activity of MSY2 for cytoplasmic localization and retention in mouse oocytes. *Developmental Biology*, 255(2): 249–262.

Zalokar M (1976). Autoradiographic study of protein and RNA formation during early development of Drosophila eggs. *Developmental biology*, 49(2): 425–437.

Zampieri M, Ciccarone F, Calabrese R, Franceschi C, Bürkle A, Caiafa P (2015). Reconfiguration of DNA methylation in aging. *Mechanisms of Ageing and Development*. In press.

Zeng F, Baldwin DA, Schultz RM (2004). Transcript profiling during preimplantation mouse development. *Developmental Biology*, 272(2): 483–496.

Zeng F, Schultz RM (2005). RNA transcript profiling during zygotic gene activation in the preimplantation mouse embryo. *Developmental biology*. 283(1): 40-57.

Zhang Y, Vastenhouw NL, Feng J, Fu K, Wang C, Ge Y, Pauli A, Hummelen PV, Schier AF, Liu XS (2013). Canonical nucleosome organization at promoters forms during genome activation. *Genome Research*, 24(2): 260–266.

Zheng W, Gorre N, Shen Y, Noda T, Ogawa W, Lundin E, Liu K (2010). Maternal phosphatidylinositol 3-kinase signalling is crucial for embryonic genome activation and

preimplantation embryogenesis. *EMBO reports*. 11(11):890-5.

Zhao BS, Wang X, Beadell AV, Lu Z, Shi H, Kuuspalu A, Ho RK, He C (2017). m6A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition. *Nature*, 542(7642): pp.475-478.

# Chapter 1: Methods development

## 1.1 Overview

The original aim of my PhD was to use the spiny mouse (*Acomys cahirinus*) to investigate early mammalian development, specifically to investigate placental physiology, and the role fetal sex plays in pregnancy outcome. The spiny mouse provides several biological advantages over other species for this purpose, however translation of scientific techniques and methods developed in other species often requires extensive optimisation and validation before use. A series of technical and time limitations prevented me from completing these studies, however I was successful in developing, optimising and validating several original methodologies in pursuit of my aims. This chapter provides the rationale and background for these original ideas and describes the methods developed. Subsequent chapters in this thesis investigate and characterise key components of spiny mouse biology, including *de novo* assembly and validation of a transcriptome, and characterising the initiation of gene transcription in the spiny mouse embryo to establish the timing and progression of embryonic genome activation.

## 1.2 Rationale

In humans, physiological differences between males and females become evident early in development (Hughes, 2002; Yuron et al., 2002). Compared to female fetuses, male fetuses are more vulnerable to maternal stress and disease states (Engel et al., 2008; Gabory et al., 2013; Wells, 2000) including preeclampsia (Muralimanoharan et al., 2013), maternal asthma (Saif et al., 2014), and premature preterm rupture of membranes (Di Renzo et al., 2007). Male fetuses are at greater risk of reduced fetal size (Vaughan et al., 2011), reduced placental weight, energy reserve and function (Mandò et al., 2016;

Roseboom et al., 2011; Wallace et al., 2012), and are at higher risk of morbidity and mortality (Eriksson et al., 2010; Verburg et al., 2016). Males born prematurely are significantly more likely to die than premature females (Logan, 1950; Stevenson et al., 2000), and maternal food restriction or severe trauma early in pregnancy has been shown to increase the risk of male mortality to a greater degree (Rosenfeld, 2011; Torche & Kleinhaus, 2012). The underlying mechanisms responsible for these sex-specific differences are not fully characterised, and as a result we often fail to account for the sex of the fetus in current clinical care, which may exacerbate the 'male disadvantage' in neonatal outcome.

Evidence of increased risk of death in male neonates can be identified in large epidemiological studies (Del Principe et al., 2013). Census data collated by the Human Mortality Database (University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org; accessed 03/04/2017) illustrate greater male mortality during the perinatal period, compared to the expected 1:1 ratio of males to females (Figure 1). Despite a significant decrease in total neonatal mortality over the past few hundred years, there are still approximately 40% more male neonates dying during early life than females. The true degree of male disadvantage is likely more pronounced than these data suggest, as these data do not account for factors such as historically high rates of female infanticide and preference for male offspring (Hesketh & Xing, 2006). Regardless, a clear correlation between fetal sex and neonatal survival can be seen.

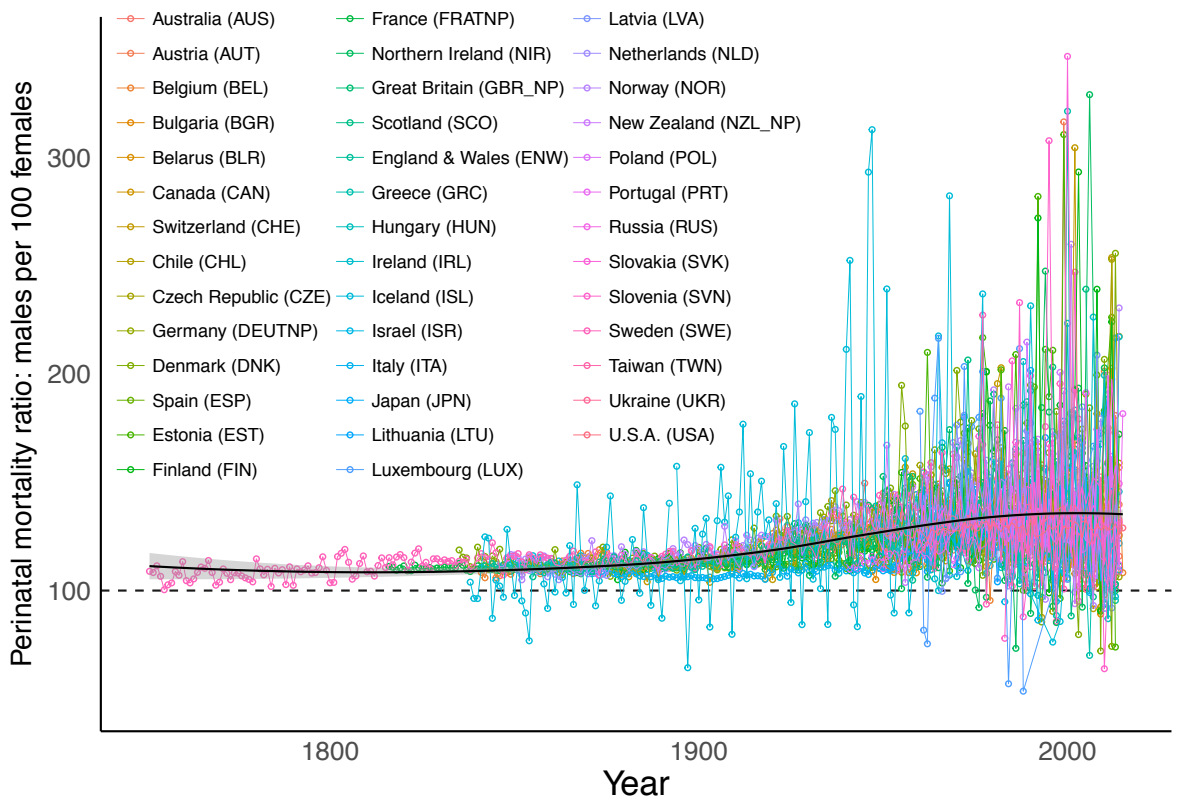



Figure 1: Ratio of male/female mortality during the perinatal period (0-2 years) for each country in the Human Mortality Database (n=41) from 1751-2015. Dashed line represents gender parity. Trend line (black) represents locally weighted regression (LOESS) with 95% CI (shaded). Code to generate figure: https://github.com/jpmam1/humanmortalitydatabase

In addition to higher neonatal mortality, males have an increased risk of morbidity in early life (Stevenson et al., 2000; Townsel et al., 2017), which corresponds to an increased incidence of chronic illness later in life (Barker, 1995; Barker, 1998; Fall & Osmond, 2013). This effect is termed the 'Barker Hypothesis', or the 'Developmental Origins of Health and Disease' (DOHaD), and states that fetal 'programming' through environmental conditions *in utero*, such as exposure to maternal stress hormones or malnutrition, predisposes individuals to diseases in adulthood, such as cardiovascular disease and obesity (Hales &

Barker, 2001; Navarro et al., 2017; Seckl & Holmes, 2007; Stanner et al., 1997). As with neonatal mortality rates, evidence for this theory can be found in large scale epidemiological data: males lead on average 5-10% shorter lives than females (Figure 2), and the prevalence of males in the population decreases as age increases (Figure 3). The DOHaD hypothesis emphasises the importance of identifying and characterising sex-specific differences in neonatal development; according to this theory, if the incidence and severity of specific diseases are correlated with being male or female, failure to adapt clinical treatment to account for fetal sex will result in a poorer start to life and a life-long predisposition toward poorer health in males.

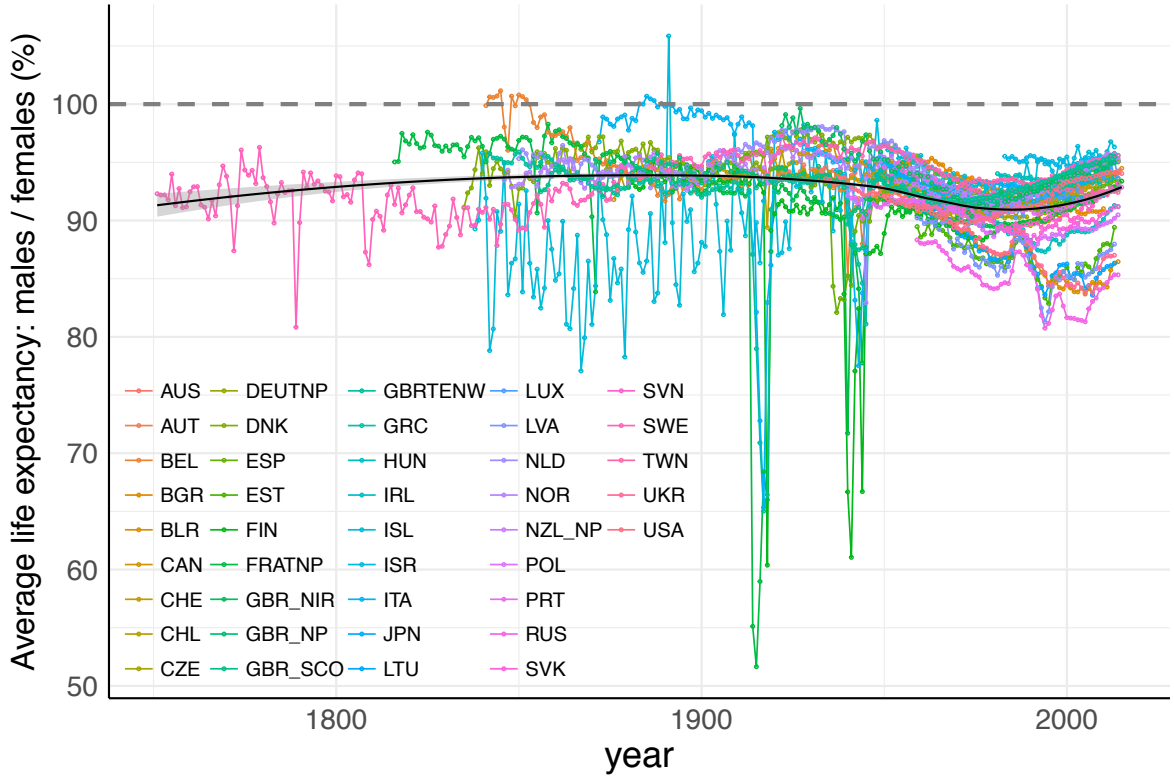



Figure 2: Ratio of male/female life expectancy for each country in the Human Mortality Database (n=41) from 1751-2012. Dashed line represents gender parity. Trend line (black) represents LOESS regression with 95% CI (shaded).

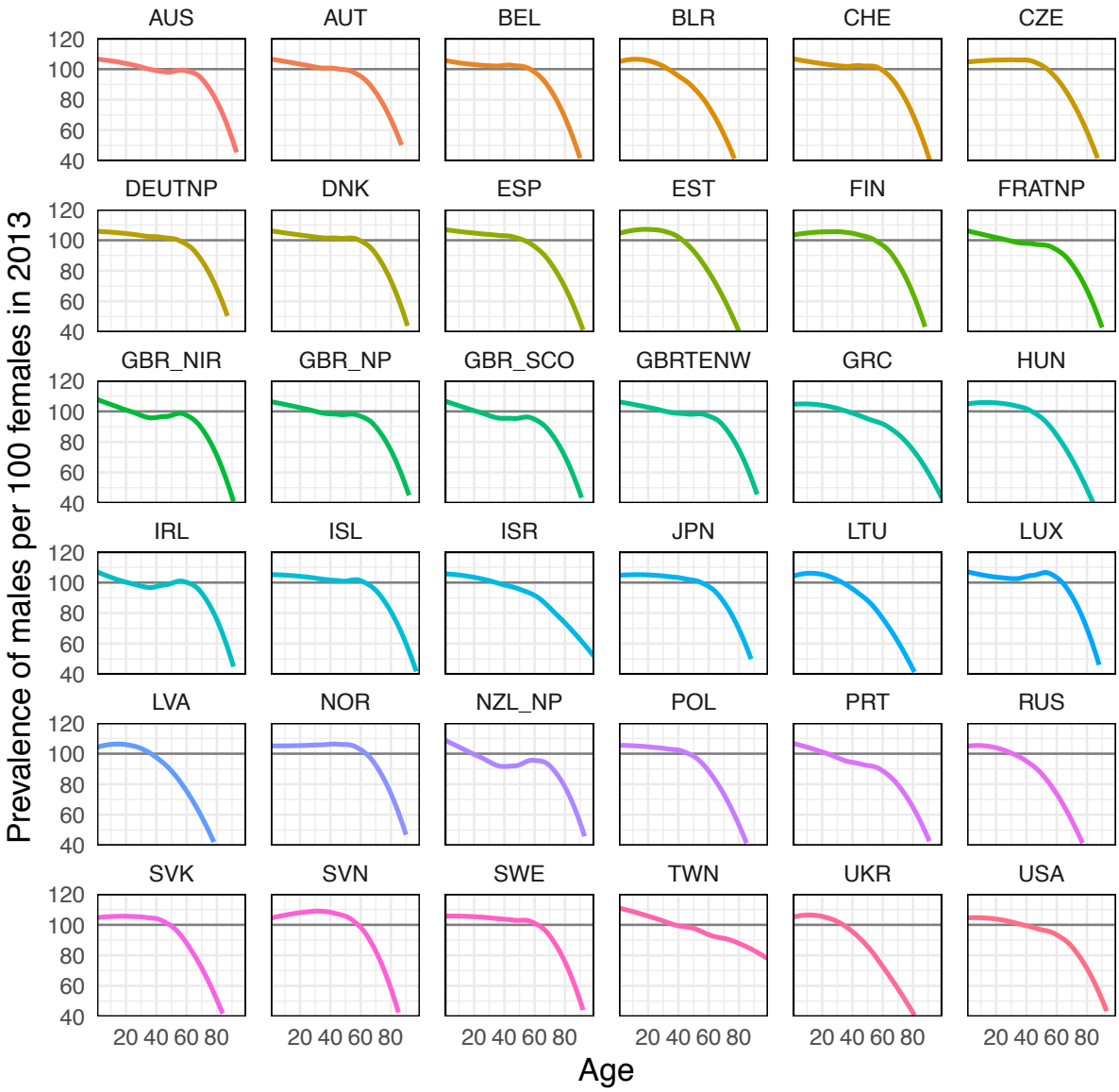



Figure 3: Number of males per 100 females in 2013. In all countries sampled (n=36) more male births were recorded than females, however higher male mortality throughout life reduces the male/female ratio in the population as age increases.

Despite our limited understanding of physiological mechanisms responsible for sex-specific susceptibility to morbidity and mortality, several targets for investigation have been

identified (Aiken & Ozanne, 2013; Tarrade et al., 2015). One of the most promising targets for unravelling the causes of male disadvantage is the placenta, which plays an integral role in fetal and neonatal health (see 'Burton et al., 2016' for comprehensive review). The placenta is a highly specialized organ that supports fetal growth *in utero* and forms the interface between mother and fetus. It is a barrier-type structure, made of cells from both maternal and fetal origin, and it is essential for normal growth and development in eutherian mammals. The placenta acts to regulate nutrient delivery, gas exchange and waste removal, it insulates against the maternal immune system, prevents transmission of potentially harmful factors, protects against adverse maternal conditions, and synthesises specific hormones, cytokines and growth factors (Jansson & Powell, 2007; Vaughan et al., 2011). Significant structural and functional adaptations can be required to perform these tasks, and the placenta has the capacity to respond to maternal insults such as hypoxia or malnutrition to mitigate adverse effects on the fetus (Brett et al., 2014). Many of the developmental processes involved in placentation are also directly involved in fetal programming, however the correlation between placental physiology and neonatal outcome is typically measured in humans only when the organ is accessible (i.e. at birth), as prenatal intervention can jeopardise the health of the fetus.

When pregnancy complications occur, such as maternal asthma (Clifton & Murphy, 2004; Murphy et al., 2003; Murphy et al., 2005), preeclampsia (Muralimanoharan et al., 2013; Stark et al., 2006; Stark et al., 2009), or malnutrition (Imdad et al., 2016; Stanner et al., 1997), male and female fetuses are seen to institute different 'coping' strategies in response to insult long before they are born (Barker et al., 2012; Eriksson et al., 2010). The placenta is typically implicated in these strategies: for example, male fetuses typically have a smaller placenta than female fetuses when exposed to suboptimal maternal environments (Clifton, 2010). We, and others, hypothesise that male fetuses maintain fetal

growth during poor *in utero* conditions utilising energy that would otherwise have been incorporated into placental growth, whereas female fetuses reduce fetal growth and subsequently place less demand on the placenta maintaining placental size to a greater degree (Kalisch-Smith et al., 2016; Sandman et al., 2013). It is not yet known whether sex-specific mechanisms regulating placental size in response to adverse conditions are 'controlled' by the fetus or by the placenta, however the answer to this question has important clinical implications for developing, optimizing, and administering treatments that account for fetal sex. Using an animal model, the identification of fundamental differences between male and female placentas *in utero* at various stages of gestation is possible. Our research group has employed a unique animal model of pregnancy for this purpose, the spiny mouse (*Acomys Cahirinus*).

The Common or Cairo spiny mouse (*Acomys cahirinus*) is well suited to modelling human disease. As detailed in the literature review, they exhibit a precocial pattern of development (Brunjes, 1990; Dickinson et al., 2005), non-rodent-like synthesis of hormones such as cortisol and dehydroepiandosterone (Lamers, 1986; Quinn et al., 2013; 2015), and a menstrual cycle instead of an estrous cycle (Bellofiore et al., 2017). They have a relatively long gestation (~39 days: almost double that of *Mus musculus*), small litters of 1 - 5 pups (typically 2-3 per litter), and a discoid haemotrichorial placenta with a developmental trajectory and umbilical blood flow approximating the human placenta (King & Hastings, 1977; O'Connell et al., 2012; Dickinson, unpublished). These characteristics support the use of the spiny mouse to model human reproductive physiology as a preferential model compared to rodents for hormonal and developmental timing reasons, and compared to primates for practical and logistical reasons such as cost of maintaining a colony and ethical status.

Placental and fetal physiology of *Acomys cahirinus* were first reported in 1905, with significant differences to small rodents such as mice and rats described (Assheton, 1905). A substantial amount of research in this area has followed, characterising and investigating many aspects of placental, fetal and maternal physiology during pregnancy. Spiny mice have been successfully used to model maternal stressors: placental response to excess maternal glucocorticoid exposure has been examined, revealing significant differences in placental physiology between males and females (O'Connell et al., 2011; O'Connell et al., 2013a). This sexually dimorphic response has not been reported in other rodent models, further supporting the spiny mouse as a preferential model of human pregnancy. Recently, we examined placental gene expression at different stages of gestation, identifying distinct sexual dimorphism in the placenta of the spiny mouse. Fetal growth promoters *SLC2A1* and *MAP2K1* were expressed earlier in male placentas and *IGF1R* was more highly expresses in male placentas (O'Connell et al., 2013b). This finding provides novel insight into the mechanism driving the male 'coping' response to suboptimal conditions and suggests sexually dimorphic fetal and placental phenotypes are established extremely early in development. To further explore this mechanism we look to the earliest stages of fetal development: when the embryo begins to differentiate.

Formation and subsequent development of mammalian embryos has been described extensively in the literature (Moore et al., 2015). Differentiation of embryonic cells first occurs at the blastocyst stage: pluripotent embryonic cells lose their pluripotent phenotype and begin to acquire characteristics required for implantation into the uterus. These differentiated cells - trophectoderm (TE) cells - further differentiate to form the cytotrophoblast and syncytiotrophoblast, which mediate implantation, and ultimately form the placenta. Contrary to the trophoblast lineage, the fetus is comprised from cells of the epiblast lineage, derived from the Inner Cell Mass (ICM) (Surani et al., 2007). Cells

comprising the ICM retain their pluripotent potential: they can differentiate into any cell in the body and they are capable of continually proliferating when cultured in a suitable *in vitro* environment. All tissues in the fetus are derived from cells of this origin, just as tissues of the placenta are derived from cells of trophoblastic origin. Interactions between the ICM and the TE in the early embryo are the most likely origin of sexually dimorphic physiological characteristics identified in the fetus and placenta.

Molecular interactions between the ICM and TE, and subsequent effects on embryonic, prenatal, and placental development, are not fully understood (Vaughan et al., 2011). This is in part due to the temporal component of many of the processes under investigation: the myriad of biological processes required to establish life are dynamic in nature, being activated and inactivated at specific time-points in development (Rice et al., 2016). Striving to understand these processes and their contribution to fetal health will likely require more than a *prima facie* understanding of anatomy and physiology. Modern scientific techniques will be required to provide novel insight into underlying mechanisms driving development and to elucidate physiological characteristics that have not yet been identified. One such technique for gaining unique insight into developmental biology is the creation of chimeric organisms.

In ancient Greek mythology the chimera was a fire-breathing monster with a lion's head, a goat's body, and a serpent's tail (Oxford Dictionary, 2017). In biology the modern usage denotes an organism comprised of two or more genotypes. In mammals, chimeras are found to occur naturally in a variety of species, however most are rare occurrences caused by dysfunctional processes in development. Research utilising chimeras has been successful in investigating sex determination, germ-cell differentiation, immunology, tumour clonality, size regulation, melanocyte migration and pigment patterns, and cell lineages, among other developmental processes (McLaren, 1976). Chimeras have also

been used extensively to validate the pluripotency of human embryonic stem cell lines, where stem cells are introduced into an immune-compromised mouse, and the resulting tumours comprised of various cell lineages are used as evidence of pluripotency (Lensch et al., 2007). The creation of chimeras for research, and utilisation of many of the techniques and methodologies used in the creation of gene-edited organisms, has provided researchers with novel insight into the underlying biological processes involved in development. This is evident in many experimental models, and in a highly pertinent example: embryo reconstruction.

In 2005, Zheng et al. reported their findings on the developmental ability of chimeric 'reconstructed' mouse blastocysts comprised of ICM and TE from different origins. 'ICM transfer' was conducted using the ICM from a blastocyst derived from a C57BL/6 mouse and the ICM-less TE from a blastocyst derived from a Kunming mouse. The reconstructed blastocysts were injected into the uterus of pseudopregnant recipient mice to implant and continue to develop *in utero*. The primary finding of this study was the derivation of a single live young (C57BL/6) attached to the placenta of a different mouse strain (Kunming). This result was obtained after transferring 124 reconstructed blastocysts into recipient female mice, providing evidence that reconstructed blastocysts can form viable fetal-placental chimeras. This result has not been reproduced since this paper was published, likely due to the substantial amount of work required for a single successful outcome, however several technological advances in embryo manipulation have occurred during the past 12 years. Addressing technical shortcomings in the method published by Zheng et al. (2005) by optimising each stage of the procedure using modern techniques may increase the success rate and facilitate the development and application of 'ICM transfer' as an investigative tool for translation to other species.

Once ICM transfer was successfully performed in mice, the methods could then be translated to the spiny mouse. This translation will require development and optimisation of techniques used in the spiny mouse for the first time, such the injection of *in vitro* embryos into the uterus of a receptive female (embryo transfer). This procedure is used clinically in IVF and extensively in medical research to produce genetically modified organisms; it is one of the most commonly performed surgical procedures worldwide (Department of Primary Industries, 2014; European Commission, 2014; U.K. Government: Home Office, 2013). Embryo transfer requires preparation of the recipient's uterine lining for implantation of donor embryos. In mice and other non-menstruating species the uterine lining can be manipulated to increase receptivity via administration of exogenous progesterone or pairing with a vasectomised male: a technique known as pseudopregnancy induction (Behringer et al., 2014).

My efforts to establish methods for pseudopregnancy induction and translation of ICM transfer to the spiny mouse predate our discovery that the spiny mouse has a fundamentally different reproductive physiology to other rodents: the spiny mouse has a menstrual cycle like humans/primates rather than an estrous cycle like >98% of all mammals (Bellofiore et al., 2017). This discovery provides new context for interpretation of the results described.

## 1.3 Aims and hypotheses

The aims of this chapter were to optimise the ICM transfer protocol using F1(C57BL/6 X CBA) mice, and to apply the optimised ICM transfer technique to the spiny mouse to investigate the effects of placental genotype on fetal growth in normal and pathophysiological pregnancies. We hypothesized reconstructed embryos created from ICM and TE cells from independent mouse blastocysts will interact and the ICM will

adhere to the TE, exhibiting minimal cell death and re-expanding in culture. Furthermore, we hypothesized that reconstructed mouse embryos transferred to the uterus of recipient female mice would implant and the technique would be successfully translated to the spiny mouse to investigate sex-specific placental physiology using ICM transfer derived fetal-placental chimeras.

## 1.4 Methods

### 1.4.1 Animal housing and ethics

All experiments were conducted in accordance with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes and were approved by the Monash University/Monash Medical Centre Animal Ethics Committee. Mice were housed under constant temperature of 22 ± 0.5˚C, constant humidity of 30 ± 5% and a 12 hour light-dark cycle (lights on at 08:00am). Spiny mice were sourced from our own research colony and housed as previously described (Dickinson & Walker, 2007), with temperature 25–27°C, humidity 30–40%, and a 12-hour light-dark cycle (lights on 7:00 am).

### 1.4.2 Embryo collection and culture

Female F1(C57BL/6 X CBA) mice were superovulated via interperitoneal administration of 5IU pregnant mare serum gonadotrophin (PMSG) followed 48h later with 5IU of human chorionic gonadotrophin (hCG). They were housed overnight with a male F1(C57BL6 X CBA) mouse of proven fertility and separated the following morning to identify vaginal plugs, indicating successful copulation and staging the mice at Day 0.5 post-coitum (0.5dpc). Successfully mated female mice were culled by cervical dislocation 40-42h post-hCG administration and reproductive organs were collected.

Dissection of the reproductive organs to obtain embryos was conducted per Gardner & Lane (2014). In summary, all instruments used in dissection were cleaned thoroughly with 70% ethanol, the uterus, oviduct and ovary were excised, and organs were immersed at 37°C in 3(N-Morpholino)propanesulfonic acid (MOPS) buffered 'embryo handling' medium (GMOPSplus™, Vitrolife). This process was conducted as quickly as possible to minimise exposure to non-physiological temperatures. The oviduct was separated from the ovary and uterus using forceps and fine spring scissors. Care was taken to avoid damaging the oviduct. Fine spring scissors were used to sever the bursar and to separate the oviduct from the uterus.

Warmed GMOPSplus™ (.75ml) was drawn into a 1ml syringe fitted with a 33-gauge flat-ended needle. The needle was inserted into the infundibulum and the outer oviduct was held in place using fine forceps. The plunger was depressed, releasing embryos into the dish in a 'cloud' of oviducal cells and debris. In the rare cases where flushing was not possible due to blockage, damaged oviduct or trouble accessing the infundibulum, the oviduct was torn with fine forceps to release embryos into the dish. The embryos were then quickly submerged twice in clean GMOPSplus™ to remove oviducal cells and debris, then transferred into culture medium using a pulled borosilicate glass pipette.

Prior to embryo collection, G1plus™ embryo culture medium was pipetted into 7 X 10µl droplets using a flushed tip in a 35mm X 10mm embryo culture dish. Droplets were overlaid with 4ml of embryo culture grade paraffin oil as quickly as possible to prevent evaporation from the droplets. A further 10µl of culture medium was then added to each 10 drop (10µl on top of 10µl to provide 20µl total). These dishes were then incubated at 37°C, 5%$O_2$, 6%$CO_2$, 89%$N_2$ to equilibrate for 12h.

Embryo culture medium was changed in the morning of day 2.5 post-coitum and day 4.5 post-coitum into G2plus™ blastocyst medium. New droplets were equilibrated for 12-18h in an incubator (37°C, 5%$O_2$, 6%$CO_2$, 89%$N_2$) before use. Embryos were transferred to the new dishes of fresh embryo culture medium using flushed, pulled borosilicate glass pipettes.

### 1.4.3 ICM transfer

Embryos were cultured for 110-112h post-hCG (4.5 dpc) to the early blastocyst stage of development. They were then transferred individually to 10 droplets of GMOPSplus™ embryo handling medium in Falcon 50mm petri dishes overlaid with paraffin oil. A non-contact infra-red laser (ZILOS tk®: Hamilton Thorne) was used to ablate a small (<5um) section of the zona pellucida above the ICM without outwardly impacting the outer TE cells (Figure 4). Embryos already hatching at this stage were allocated to immunosurgical ICM isolations. Laser ablation was conducted late on the evening of D4 (~102h post-hCG: 3AA grade embryos) or very early on D5 (~110h post-hCG: 4AA grade embryos) on 10 embryos each time.

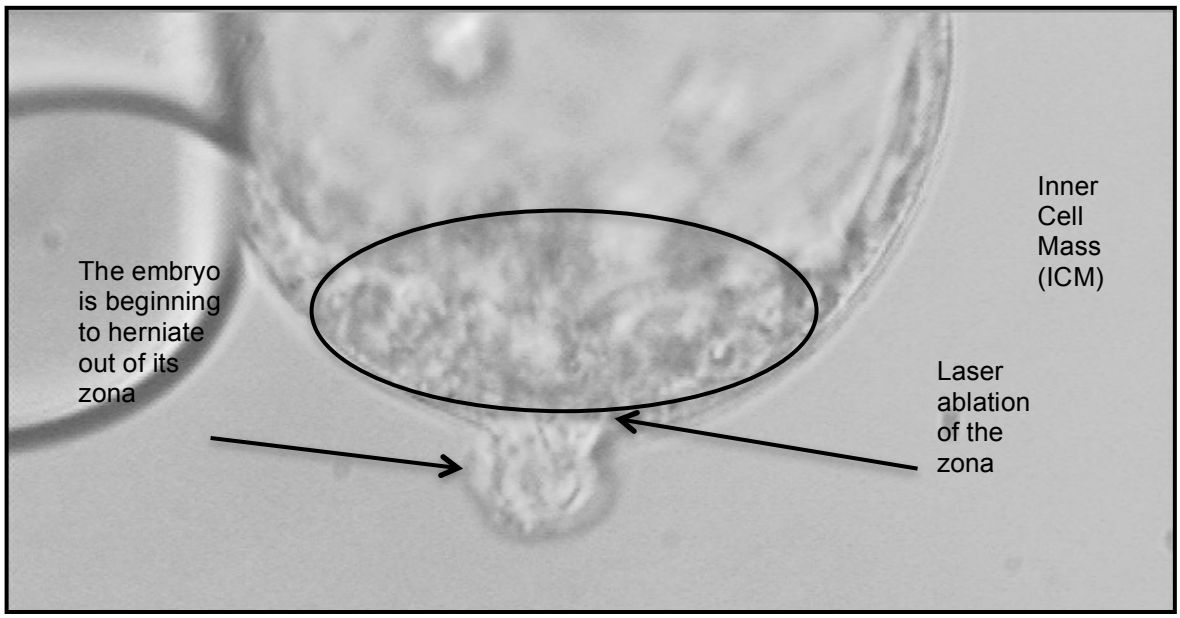



Figure 4: A blastocyst is held with suction to a holding pipette and a small ablation (<5µm) of the zona pellucida is conducted using a laser. After 2-6h, the ICM of the blastocyst herniates out of the zona.

Embryo immunosurgery was conducted per Solter & Knowles (1975). In brief, the zona was dissolved with pronase (5.9U/mg at a final concentration of 0.5% (w/v) as described by Tanaka et al., 2006), trophectoderm cells (TE) were labelled with rabbit anti-mouse whole serum for 20 min, and exposed to Guinea Pig Serum compliment (GPS) for 30 min. Only high morphological grade embryos were used for this procedure (4AA: Gardner & Schoolcraft, 1999). Immunosurgical ICM isolations were conducted immediately after embryos underwent laser ablation. Media was equilibrated at 37°C, 5%$O_2$, 6%$CO_2$ and 89%$N_2$ for 12-18h before use. The addition of pronase, rabbit anti-mouse antibody and GPS solutions was done using a 10µl pipette 2-3 hours before the procedure. Trophectoderm lysis (also known as 'blebbing') is shown in Figure 5.

After incubation in GPS, embryos were agitated mechanically to separate the live ICM cells from the lysed trophectoderm cells using a very thin pulled glass pipette (internal diameter <40µm). Sheared cells were washed in Mg2+/Ca2+-free PBS and snap frozen in liquid nitrogen for genotyping to determine the sex of the ICM in fetuses developing from implanted reconstructed embryos.

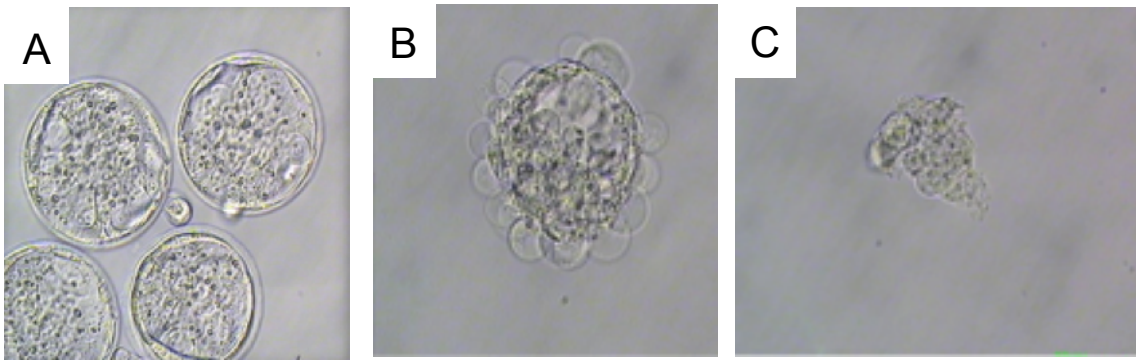



Figure 5: (A) Normal blastocysts, (B) after removal of the zona and exposure to GPS, showing ‘blebbing’ of the trophectoderm (TE) cells, and (C) isolated ICM cells after mechanically shearing off lysed TE cells.

Herniated ICMs were severed from the blastocyst using a laser (Figure 6). The left micromanipulator was equipped with a standard holding pipette and, the right was equipped with an angled flat-ended blastomere biopsy pipette (28-32um tip, 30-35 degree bend). Isolated donor ICM cells were injected into the recipient blastocyst (Figure 7). The herniated ICM from the recipient blastocyst was removed, sheared cells were washed in Mg2+/Ca2+-free PBS then snap frozen in liquid nitrogen for genotyping to determine the

sex of the TE in placentas developing from implanted reconstructed embryos. Reconstructed embryos were cultured individually in 10µl drops of G2plus medium under oil at 37°C, 5%$O_2$, 6%$CO_2$ and 89%$N_2$. Reconstructed blastocysts were cultured until re-expansion (1-6h), and immunofluorochemically-stained or surgically transferred to the uterus of a pseudopregnant female mouse recipient. Each stage of this process was optimised, as shown in Table 1.

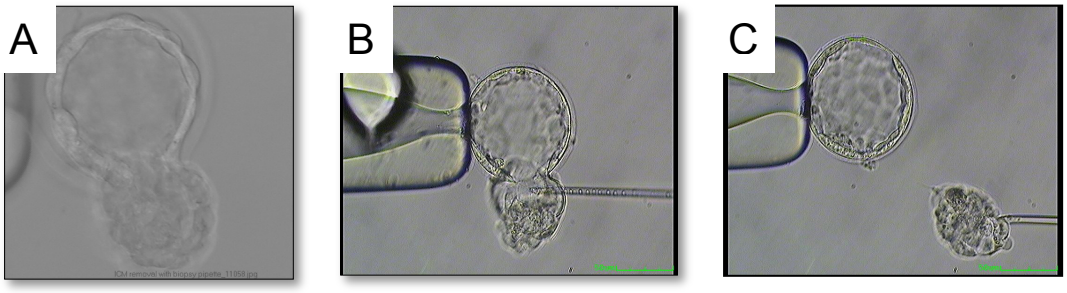



Figure 6: (A) herniated blastocyst held in position using a holding pipette, (B) positioning and removal of the ICM using the laser and ICSI pipette, and (C) the ICM cells separated from the TE cells.

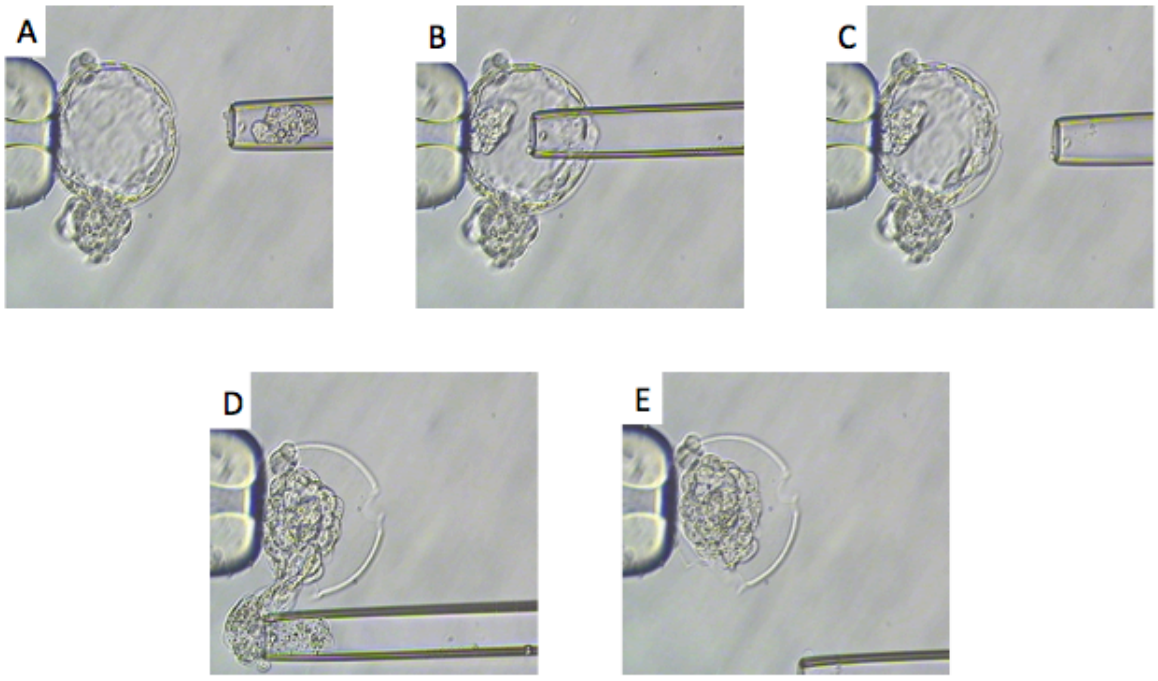



Figure 7: The reconstruction process: (A) inserting the biopsy pipette, (B) expelling the donor ICM, (C) retracting the biopsy pipette, (D) removing the herniated ICM (E) the resulting reconstructed embryo.

Table 1: Summary of optimisation for each stage of the ICM transfer procedure.

| Task | Variable | Tested | Optimal |
|---|---|---|---|
| Embryo collection (i.e. mice culled, dissected, embryos obtained and transferred into *in vitro* culture in an incubator) | Number of mice culled together and the time taken from dissection to *in vitro* embryo culture | Embryos collected from 4 mice per collection; embryos collected from 2 mice per collection | Embryos collected from 2 mice per collection |
| Timing of zona ablation | When to ablate the zona pellucida to induce hatching of the ICM | 96h – 110h in increments of 2h | 102h |
| ICM isolation | When to isolate the ICM from donor embryos | 96h – 110h in increments of 2h | 108h |
| Time in Guinea Pig Serum compliment | How long to expose embryos to Guinea Pig Serum compliment to lyse TE cells | 15 min – 45 min in increments of 5 minutes | 30 mins |
| Micropipette used for ICM transfer | Is an angled or flat injection pipette more effective | Angled (45°) and flat ended (90°) | Flat ended (90°) |
| Time in culture after ICM transfer | How long is required to assess ICM-TE adhesion | 1h – 6h in 30 min increments | 2h |
| Embryo transfer | Which method of skin incision is fastest for embryo transfer to both uterine horns | Single dorsal skin incision dual laparotomy, or two skin incisions (one of each flank) dual laparotomy | Single dorsal skin incision, dual laparotomy (per Gardner et al., 2016) |

### 1.4.4 Immunofluorochemistry

Immunofluorochemical cell labelling was conducted at room temperature. Blastocysts were fixed with 4% PFA (20 mins), washed in PBS (5 min X 3), permeabilized with 0.1% Triton X-100 (5 mins), blocked with 10% DAKO protein block (1 hr; Agilent Pathology Solutions), then incubated with primary antibodies at 4°C overnight: monoclonal mouse anti-human OCT-3/4 (1:250, Santa Cruz Biotechnology sc-5279) and monoclonal rat anti-mouse TROMA-1 (1:500; Developmental Studies Hybridoma Bank, University of Iowa). After incubation, blastocysts were washed with PBS (3 X 5 min) and incubated with secondary antibodies for 1 hour in the dark: Alexa Fluor 488 goat anti-mouse (Green, 1:1000; Molecular Probes) and Alexa Flour 566 goat anti rat (Red, 1:1000; Molecular Probes). Blastocysts were then washed and counterstained with DAPI (30 secs), rinsed with PBS, and mounting on coverslips with DAKO mounting medium for fluorescence (Agilent Pathology Solutions).

### 1.4.5 Embryo transfer

Blastocysts were surgically transferred to the uterus of female recipient mice on Day 4.5 of pseudopregnancy. Pseudopregnant recipient females were generated via natural mating with vasectomised males (Behringer et al., 2014). Pairing took place in the afternoon (approximately 4:00pm), with pseudopregnancy established via the detection of a vaginal plug the morning after pairing (0.5 dpc). The embryo transfer procedure was conducted on a warmed stage at 37°C. Anaesthesia was induced in recipient mice with 4.9% isofluorane in air, and was maintained throughout the experiment with 2.5% isofluorane in air. Carprofen was administered subcutaneously (5mg/kg in 0.1ml PBS) whilst mice were anaesthetised, and up to ten blastocyst-stage embryos (Day 5) were transferred into the lumen of each uterine horn through a small dorsal incision using a pulled and flame-

polished glass Pasteur pipette. Pipettes had a lumen slightly larger than the blastocysts being transferred, and a relatively thick wall to prevent breakage when inserted into the reproductive tract. Pipettes were flushed after transfer to determine whether all blastocysts had been expelled into the uterine lumen. Following confirmation that all blastocysts had been expelled, the peritoneum was sutured (Maxon™, Covidien, US) and the skin wound closed with sterile surgical clips. Lignocaine (Xylocaine; 0.1ml, 2% w/v) or Bupivacaine (0.1ml, 0.25% w/v) was applied topically, and mice recovered for 30 min in a dark cage on a 37°C warming plate before being transferred to their original cage. After transfer, recipients were administered 3 mg/ml paracetamol in their drinking water and were monitored for signs of distress.

### 1.4.6 Vasectomy in the spiny mouse

To translate the ICM transfer technique to the spiny mouse, protocols for vasectomy and pseudopregnancy induction were developed. Male spiny mice with proven mating performance (>3 months old) were anesthetised with 4.9% isofluorane in air. Anaesthesia was maintained throughout the experiment with 2.5% isofluorane in air. The male was placed on a warm stage in supine position (tail towards operator) and 3 cm of fur was removed from the ventral area above the penis. The shaved area was sprayed with 70% ethanol and covered with a sterile sheet of plastic containing a 3X3 cm hole. A 10-15 mm longitudinal incision was made along the medial line of the abdomen, approximately 1.5 cm above the penis. A smaller incision (5-10 mm) was made in the linea alba to expose the testicular adipose pad. This pad was gently pulled through the incision to expose the vas deferens. A disposable cautery pen was used to remove a portion of the vas deferens (~10 mm) sealing both ends, and organs/tissues were gently eased back into the cavity. The same procedure was then conducted on the other side. The muscle wall was sutured (Maxon™, Covidien, US), and skin wound closed with sterile surgical clips. Spiny mice

then recovered for 30 min in a dark cage on a 37°C warming plate before being transferred to their original cage. Wound clips were removed 10-14 days after vasectomy with a wound clip removal tool. Male infertility was confirmed with successive pairings of virgin post-pubertal females (>3 days per pairing). If neither female fell pregnant after pairing the procedure was considered successful.

### 1.4.7 Pseudopregnancy induction in the spiny mouse

Female spiny mice (n=25) between 90 and 150 days old were used. Prior to the experiment commencing, vaginal cytology was examined to determine the stage of reproductive cycle. Each animal underwent vaginal lavage using sterile saline, with cellular debris smeared on histological slides, air dried, and fixed. Slides were stained with haematoxylin and eosin and cell counts were conducted. The relative proportion of each exfoliated vaginal cell type was used to establish the stage of the estrus cycle (McLean et al., 2012). Regular cycles were established based on this vaginal cytology and animals were randomly allocated to three groups: Progesterone treatment (n=10), Vasectomised male pairing (n=5), and vehicle only (n=10). Progesterone (0.3ml) was administered as a single dose in sesame oil (30mg/ml) via subcutaneous injection. Sterile mating was conducted by pairing with a vasectomised male spiny mouse for ~18h (mice were paired at ~4:00pm), and females were examined for the presence of a vaginal mucous plug the morning after pairing to confirm successful mating (~10:00am). Vaginal lavage was conducted with sterile saline on all animals (progesterone treated with sterile male pairing, sterile male pairing only, and vehicle only). Cellular debris was smeared across glass slides and air-dried. Slides were stained with haematoxylin and eosin and cell counts were conducted. The relative proportion of each exfoliated vaginal cell type was used to establish the stage of the estrus cycle (McLean et al., 2012).

## 1.5 Results

### 1.5.1 Embryo collection and culture

Female F1(C57BL/6 X CBA) mice (n=202) between 37 and 70 days of age (average 6.7 weeks) were superovulated, their embryos were collected, and *in vitro* embryo culture was performed in 49 individual trials (~4 mice per trial). Embryos were obtained from superovulated mice in 78% of cases. Embryo collection was conducted on average 40.5h post-hCG administration. Time required for embryo collection from four mice decreased from 3 hours to 1 hour (from *in* vivo to the incubator) (Figure 8).

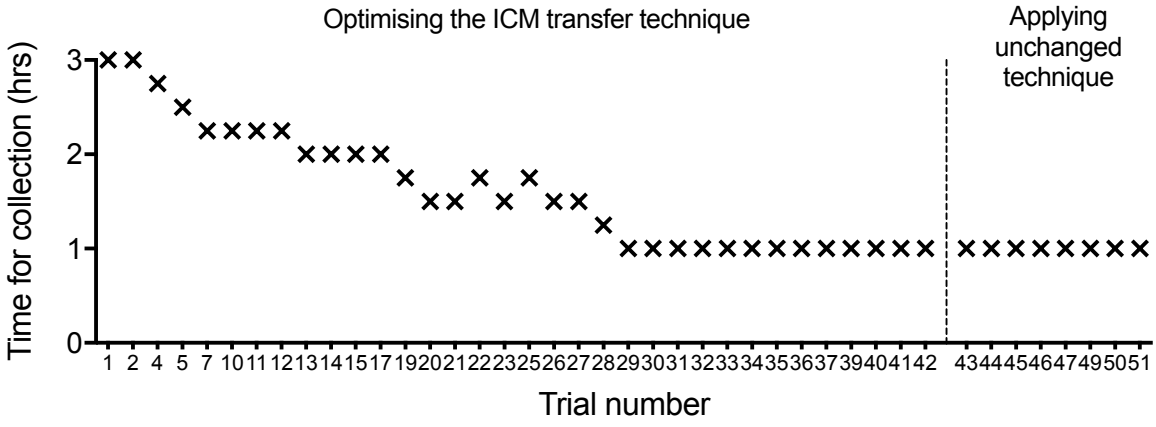



Figure 8: Time taken to cull female mice (n=4 per trial), collect embryos, and transfer embryos into *in vitro* culture (two-trial rolling average). Trials in which embryos were not obtained due to failed superovulation are excluded.

In total, 4076 embryos were collected from 157 successfully superovulated females (25.9±1.6 embryos per female). Younger mice did not generate more embryos than older mice in response to superovulation (Figure 9). The proportion of embryos that reached the blastocyst stage of development in each trial is shown in Figure 10.

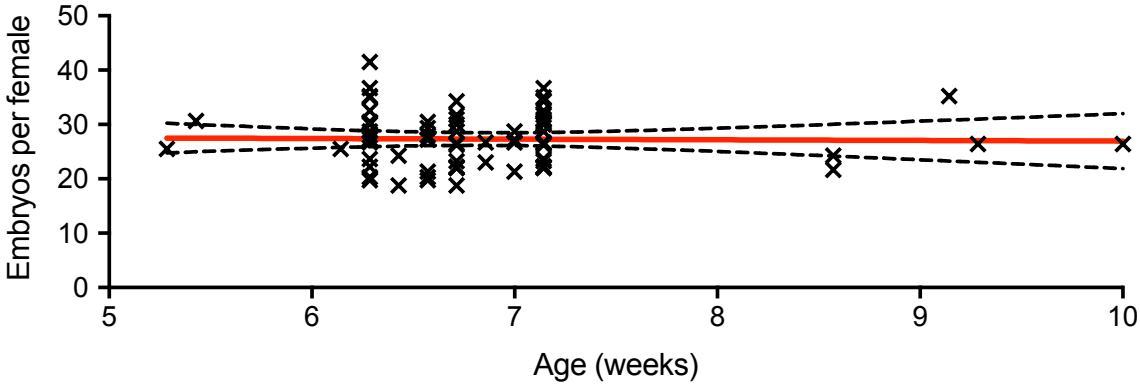



Figure 9: Embryos collected per superovulated female. Regression analysis (red line; black dashed lines 95% CI) shows no correlation between age and number of embryos collected for mice between 5.5 and 10 weeks of age.

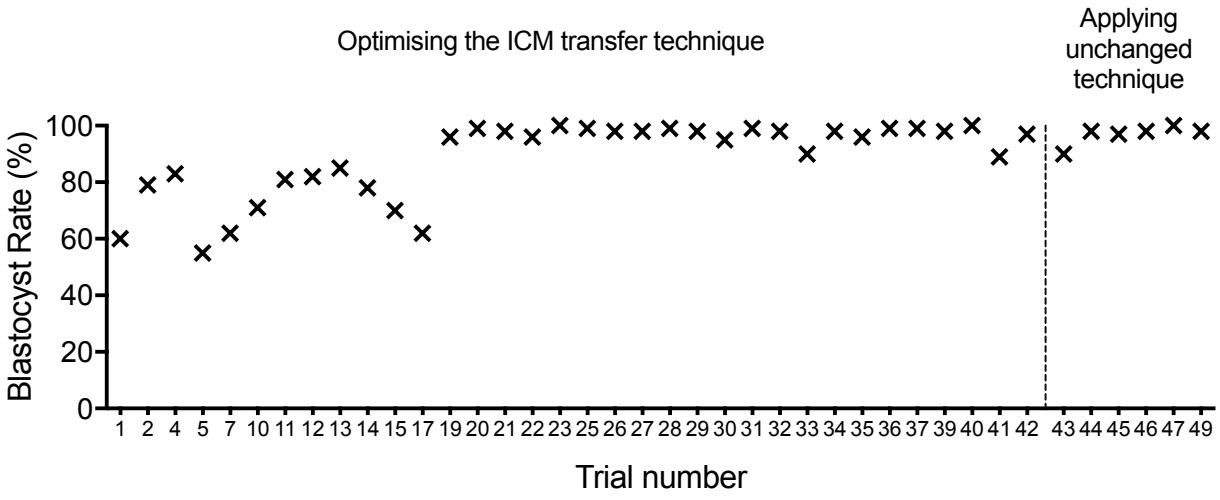



Figure 10: Proportion of *in vitro* cultured mouse embryos that reached the blastocyst stage of development in each trial.

### 1.5.2 ICM transfer

The protocol for ICM transfer was developed and optimised over 42 trials, then applied unchanged for a further 9 trials. During the optimisation trials, 535 embryos were used to develop the ICM transfer procedure: 202 embryos were allocated to immunosurgical ICM isolations, and 333 were allocated to ICM herniation. This resulted in 90 successful ICM transfers from 140 attempts. Of these, 65 reconstructed blastocysts were used for embryo transfer, 9 reconstructed blastocysts were immunofluorolabelled for cell counts, and 16 reconstructed blastocysts were returned to *in vitro* culture to investigate ICM attachment and re-expansion rates. After optimisation the developed technique (as described in the Methods) was applied unchanged, using an additional 509 embryos to successfully create 54 reconstructed blastocysts from 148 attempts (36 of these were transferred to pseudopregnant recipients). The number of ICM transfers conducted per trial is shown in figure 11.

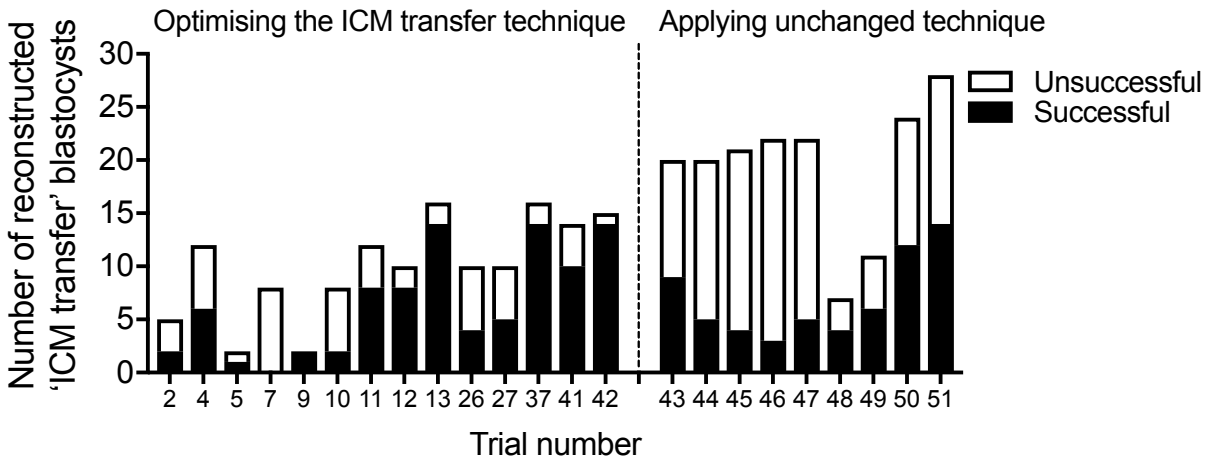



Figure 11: Number of ICM transfer procedures conducted. Closed bars represent successful reconstructions and open bars represent failed attempts.

Post-ICM transfer reconstructed embryos maintained the ability to re-expand in culture. Donor ICM cells adhered to the TE wall (Figure 12A) and reconstructed blastocysts maintained the ability to hatch from the zona (Figure 12B).

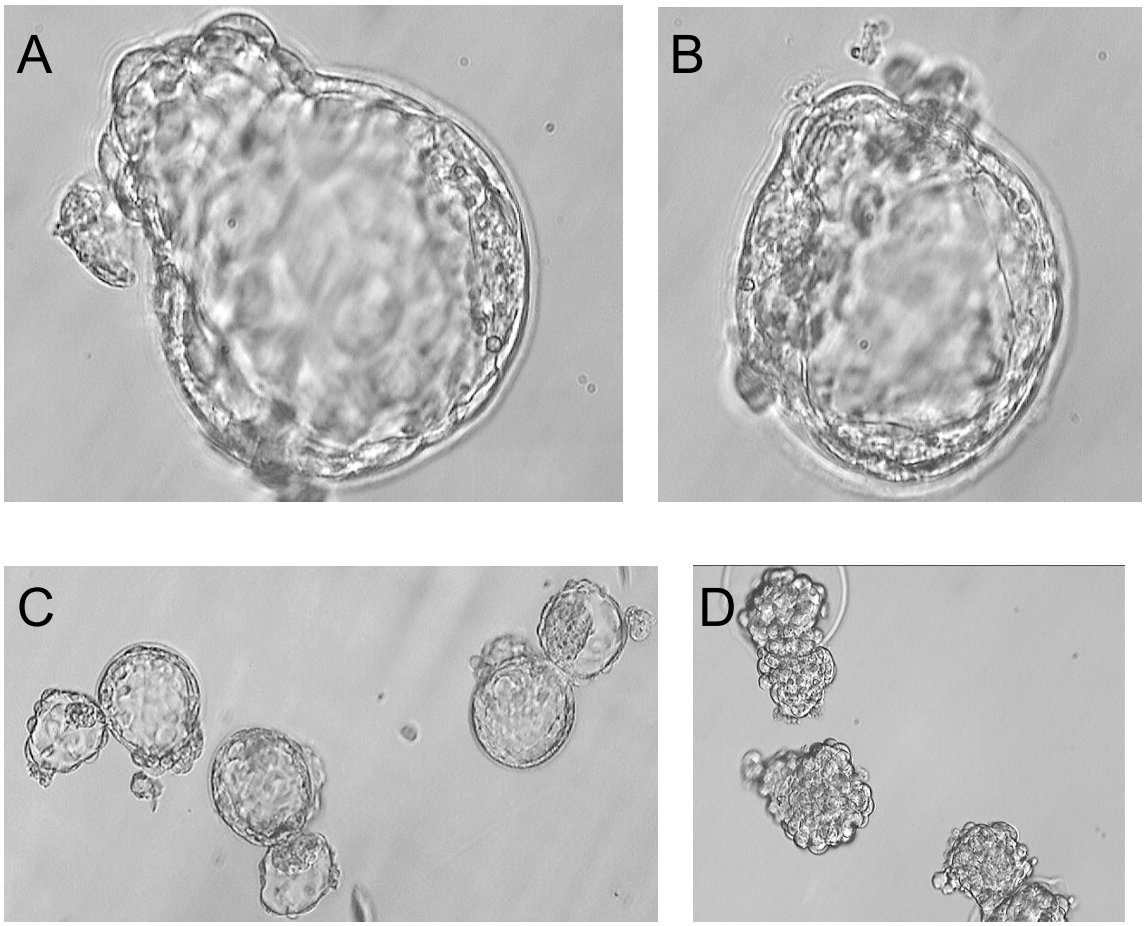



Figure 12: Re-expansion and development of reconstructed blastocysts at (A) 1 hour (B) 6 hours, (C) 24 hours, and (D) 48 hours post-ICM transfer.

### 1.5.3 Embryo Immunofluorochemistry

Immunofluorochemistry was implemented on a subset of embryos not transferred to pseudopregnant recipients to identify cells as belonging to either the inner cell mass (ICM) or the trophectoderm (TE). Cell counts were performed on unmanipulated blastocysts (n=6), and reconstructed blastocysts post-ICM transfer (n=6). On average, 31.5 ICM cells were identified per unmanipulated embryo cultured *in vitro* (31.5 ± 1.18), and 25 ICM cells were identified per embryo after undergoing the ICM transfer procedure (25 ± 1.0), with

108.17 cells of TE origin identified per stained embryo (108.17 ± 9.65), and 57.8 TE cells were identified in embryos after undergoing ICM transfer (57.80 ± 3.60).

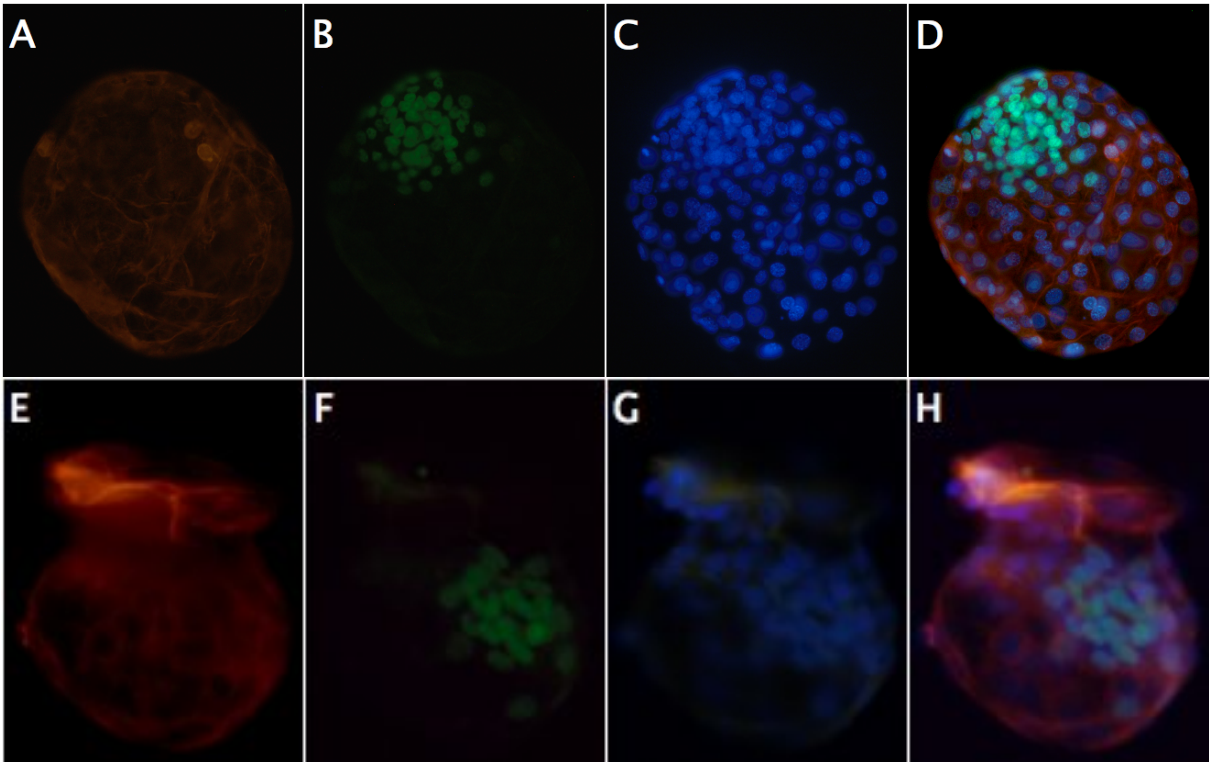



Figure 13: Immunofluorolabelled blastocysts. Unmanipulated embryo: (A) TROMA-1 labelled TE, (B) OCT3/4 labelled ICM, (C) DAPI counterstain (all cell nuclei), (D) Combined. Manipulated/reconstructed embryo: (E) TROMA-1 labelled TE, (F) OCT3/4 labelled ICM, (G) DAPI counterstain (all cell nuclei), (H) Combined.

Cell numbers in unmanipulated and reconstructed stained blastocysts are presented in Figure 14. The ratio of ICM/TE cells in unmanipulated embryos (3.40 ± 0.197) and manipulated embryos (2.30 ± 0.113) are also shown (Figure 14B).

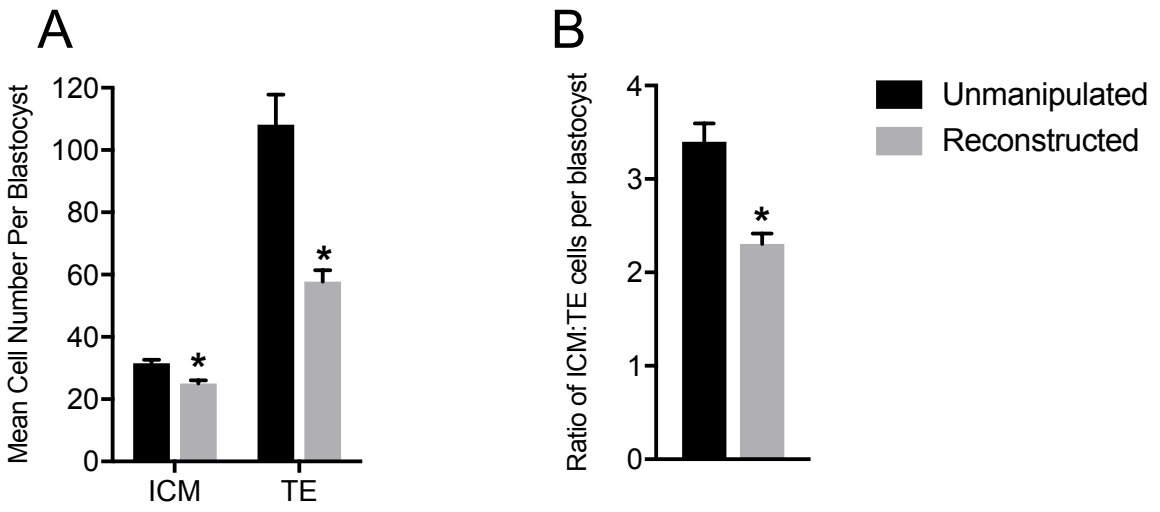



Figure 14: (A) Average number of ICM and TE cells per blastocyst in unmanipulated (n=6) and reconstructed embryos post-ICM transfer (n=6). (B) Ratio of ICM:TE cells in manipulated and reconstructed blastocysts. *p<0.05, two-tailed student's t-test.

### 1.5.4 Embryo transfer

Reconstructed embryos (n=101) were surgically transferred into uterine horns of recipient females (n=12), and implantation was assessed 10 days post-transfer. No signs of implantation or fetal resorption were detected in any recipient female mice after transfer of reconstructed blastocysts.

Control embryo transfers were also conducted. Unmanipulated blastocysts were transferred to separate pseudopregnant females during transfer of reconstructed blastocysts. The number of fetuses/pups derived from embryo transfer of unmanipulated blastocysts is depicted in Figure 15: 20 recipient mice received unmanipulated blastocysts during ICM transfer optimisation (recipients 1-20), and 12 mice received blastocysts during application of the optimised technique (recipients 21-32). The average number of viable

fetuses was 16.66% during the optimisation period, and 37% during application of the optimised technique.

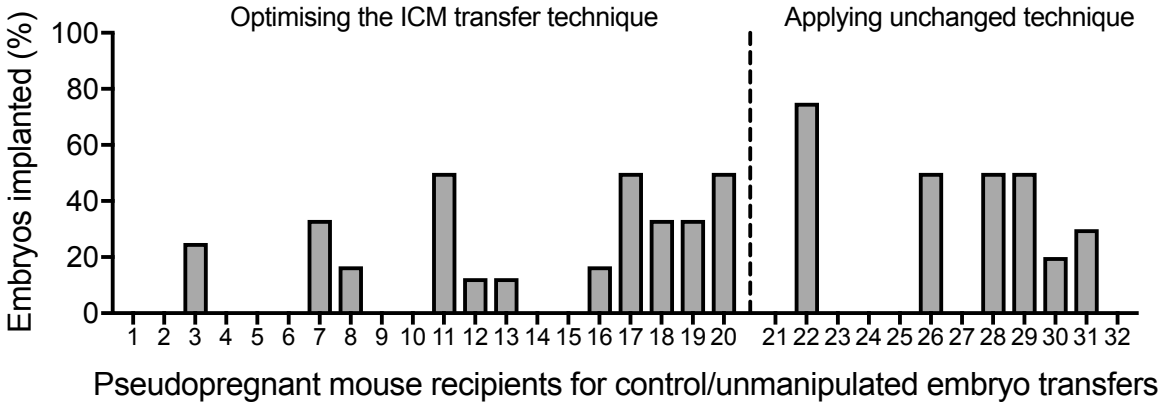



Figure 15: Proportion of viable fetuses per unmanipulated embryo transferred (n=~4 per uterine horn) to each recipient female (n=32) during optimisation and after the technique was optimised.

### 1.5.5 Vasectomy in the spiny mouse

Vasectomised males showed minimal signs of discomfort post-surgery. Wounds healed quickly with no scarring evident, and movement was not restricted: they were able to jump, flip and climb as normal. Vasectomised males failed to impregnate post-pubertal virgin females after successive pairings, and were used in attempts to induce pseudopregnancy without impregnating any of the females used. Vasectomy was successful in all male spiny mice in this experiment (n=6).

### 1.5.6 Pseudopregnancy induction in the spiny mouse

Pseudopregnancy induction in female spiny mice was unsuccessful. Untreated spiny mice (vehicle only: no progesterone or sterile-male pairing) spent 65±5h (mean±SEM) in diestrus. Sterile-male pairing and progesterone treatment increased the average time spent in diestrus (Figure 16), however the effect was highly variable (Figure 17A) and not all females responded to treatment (Figure 17B).

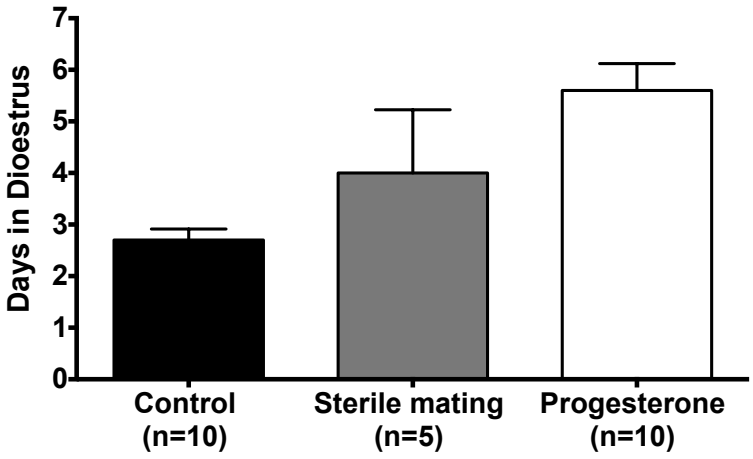



Figure 16: Average time spent in diestrus for each treatment group.

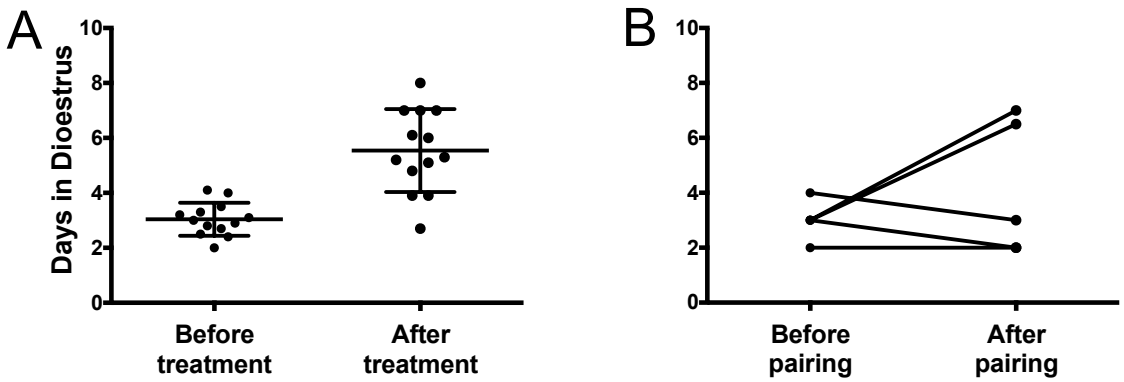



Figure 17: The effect of (A) progesterone treatment and sterile-male pairing, and (B) sterile-male pairing only, on the time spent in diestrus.

The increased length of time spent in diestrus in response to treatment was not sufficient to be considered pseudopregnant. Sterile-pairing alone was also insufficient to induce pseudopregnancy in spiny mice.

## 1.6 Discussion

Development and refinement of the ICM transfer protocol produced 144 reconstructed chimeric mouse blastocysts. Reconstructed blastocysts maintained the capacity to re-expand in culture and fully hatch from the zona, suggesting viability was not irreversibly compromised by the procedure. Immuno-labelling of ICM-specific and TE-specific markers revealed attachment of injected ICM cells to the trophoblast wall after re-expansion, suggesting cells from the ICM and TE lineages survive the procedure, although cell counts revealed a decrease in total cell number and a decreased ratio of ICM to TE cells. This suggests the capacity for development is reduced, however the degree to which viability is impeded is difficult to ascertain from this measure alone. A conclusive measure of viability is successful implantation and development to term, however in this experiment 0/101 reconstructed embryos transferred to the uterus of pseudopregnant recipients implanted.

Substantial variation in embryo transfer success rate between replicates using unmanipulated 'control' embryos was also identified in this experiment. This may indicate the presence of an uncontrolled variable that may have affected the potential for reconstructed blastocysts to implant successfully. The implantation rate for unmanipulated *in* vitro cultured blastocyst transfer is typically reported to be 50-70% in the literature (e.g. Truong et al., 2016; Walker et al., 2015), however my success rate was 37% when applying the optimised technique. Whilst this may be an indication of poor technical proficiency, an alternative explanation is the presence of an unidentified confounding

variable in the embryo culture or embryo transfer procedure (described further in Chapter 2). Despite inconclusive validation of the embryo reconstruction technique, several aspects of the procedure were optimised and produced positive outcomes.

Methods developed to translate ICM transfer to the spiny mouse were described for the first time. Vasectomy was successfully developed and validated in males. Pseudopregnancy was not successfully induced in female spiny mice. The increased time spent in diestrus exhibited by high-responders suggests the dose of progesterone used was appropriate, however the adaptation of established rodent-specific methods was unsuccessful in prolonging the time spent in the embryo-receptive phase for >5 days. This inability to successfully induce pseudopregnancy was a definitive obstacle for translating ICM transfer to this species, and prevented further development of methods for translating ICM transfer, such as species-specific protocols for embryo collection, culture, manipulation/reconstruction, and embryo transfer.

It was not immediately clear why spiny mice failed to respond to traditional methods of pseudopregnancy induction. Another student in our lab pursued this aspect of the project (Nadia Bellofiore). Nadia conducted a complete characterisation the spiny mouse reproductive cycle and discovered that spiny mice do not exhibit a typical estrus cycle; instead, the spiny mouse has a menstrual cycle, like humans and higher-order primates (Bellofiore et al., 2017). This extraordinary finding provides context to interpret the results described here: female spiny mice failed to respond to traditional 'rodent-based' methods due to fundamental differences in their reproductive physiology. There are several future directions for this work: I plan to apply primate-specific methods to prepare the uterine lining for embryo implantation (e.g. Tachibana et al., 2012) to facilitate translation of the ICM transfer procedure to this species.

An important limitation in this project was time. This was an ambitious project with a substantial amount of work required to develop and apply highly technical skills. Without a clear plan for translating ICM transfer to the spiny mouse to investigate male/female and inter-species differences in placentation, development and fetal mortality the original plan for my PhD project could not be pursued further. My new goal was to further investigate the embryo transfer procedure. Specifically, I was interested in the variability in implantation rates found after transfer of unmanipulated mouse blastocysts, and I planned to investigate potential confounding factors in the induction and maintenance of pseudopregnancy in recipient female mice to determine whether the procedure was taking place within the 'window of opportunity'. We believe optimising the embryo transfer technique in the mouse will aid translation of the ICM transfer technique to the spiny mouse once the prerequisite techniques (pseudopregnancy induction and *in vitro* embryo culture past the 4-cell block) have been developed.

## 1.7 References


Agulnik, SI and Silver, LM (1996), The Cairo spiny mouse Acomys cahirinus shows a strong affinity to the Mongolian gerbil Meriones unguiculatus, *Molecular Biology and Evolution,* 13 (1), 3-6.

Aiken, CE and Ozanne, SE (2013). Sex differences in developmental programming models, *Reproduction,* 145 (1), R1-R13.

Alwasel, SH, Abotalib, Z, Aljarallah, JS, Osmond, C, Alkharaz, SM, Alhazza, IM, Badr, G, and Barker, DJP (2010). Changes in placental size during Ramadan, *Placenta,* 31 (7), 607-10.

Assheton, R (1905). On the Foetus and Placenta of the Spiny Mouse (Acomys cahirinus), *Proceedings of the Zoological Society of London* (75: Wiley Online Library). 280-88.

Barker, DJP (1995). Fetal origins of coronary heart disease, *BMJ: British Medical Journal,* 311 (6998), 171.

Barker, DJP (1998). In utero programming of chronic disease, *Clinical science,* 95(2), 115-28.

Barker, DJP and Thornburg, KL (2013). Placental programming of chronic diseases, cancer and lifespan: a review, *Placenta,* 34 (10), 841-45.

Barker DJ, Lampl M, Roseboom T, Winder N. (2012) Resource allocation in utero and health in later life. *Placenta*, 33, e30-4.

Behringer R, Gertenstein M, Nagy KV, Nagy A (2014) Manipulating the Mouse Embryo: A Laboratory Manual, fourth ed. *Cold Spring Harbour Laboratory Press*, New York.

Bellofiore N, Ellery SJ, Mamrot J, Walker DW, Temple-Smith P, Dickinson H (2017). First evidence of a menstruating rodent: the spiny mouse (Acomys cahirinus). *American Journal of Obstetrics & Gynecology*, 216(1): 40-e1.

Brett KE, Ferraro ZM, Yockell-Lelievre J, Gruslin A, Adamo KB. (2014) Maternal–fetal nutrient transport in pregnancy pathologies: the role of the placenta. *International journal of molecular sciences*, 15(9), 16153-85.

Brunjes PC (1990). The precocial mouse, Acomys cahirinus. *Psychobiology*, 18(3):339-50.

Burton GJ, Fowden AL, Thornburg KL (2016) Placental Origins of Chronic Disease. *Physiological Reviews*, 96 (4), 1509-1565.

Challis, J, Newnham, J, Petraglia, F, Yeganegi, M, and Bocking, A (2013). Fetal sex and preterm birth, *Placenta,* 34 (2), 95-99.

Chambers, GM., Wang, YA., Chapman, MG., Sullivan, EA., Abdalla, HI., and Ledger, W. (2013). What can we learn from a decade of promoting safe embryo transfer practices? A comparative analysis of policies and outcomes in the UK and Australia, 2001–2010, *Human Reproduction,* 28 (6), 1679-86.

Chambers, PL and Hearn, JP (1985). Embryonic, foetal and placental development in the common marmoset monkey (Callithrix jacchus), *Journal of Zoology,* 207 (4), 545-61.

Clifton, VL (2010). Review: sex and the human placenta: mediating differential strategies of fetal growth and survival, *Placenta,* 31, S33-S39.

Clifton, VL and Murphy, VE (2004). Maternal asthma as a model for examining fetal sex-specific effects on maternal physiology and placental mechanisms that regulate human fetal growth, *Placenta,* 25, S45-S52.

De Geyter, Christian, Sterthaus, Oliver, Miny, Peter, Wenzel, Friedel, Lapaire, Olav, De Geyter, Maria, and Sartorius, Gideon (2013). First successful pregnancy in Switzerland after prospective

sex determination of the embryo through the separation of X-chromosome bearing spermatozoa, *Swiss Med Wkly,* 143, w13718.

Del Principe, D, Marconi, M, Matarrese, P, Villani, A, and Malorni, W (2013). Gender disparity in pediatric diseases, *Current molecular medicine,* 13 (4), 499-513.

Department of Primary Industries, (2014) Statistics of Animal Use in Research and Teaching, Report Number 30, 1 January 2012 to 31 December 2012.

Di Renzo, Gian Carlo, Rosati, Alessia, Sarti, Roberta Donati, Cruciani, Laura, and Cutuli, Antonio Massimo (2007). Does fetal sex affect pregnancy outcome?, *Gender medicine,* 4 (1), 19-30.

Dickinson H and Walker DW (2007). Managing a colony of spiny mice (Acomys cahirinus) for perinatal research, *Australian and New Zealand Council for the Care of Animals in Research and Training (ANZCCART) News,* 20 (1), 4-11.

Dickinson H, Walker DW, Cullen-McEwen L, Wintour E, and Moritz K (2005). The spiny mouse (Acomys cahirinus) completes nephrogenesis before birth, *American Journal of Physiology-Renal Physiology,* 289 (2), F273-F79.

Emera D, Romero R, Wagner G, (2012). The evolution of menstruation: A new model for genetic assimilation. *Bioessays*, 34 (1), 26-35.

Engel, Patricia J, Smith, Roger, Brinsmead, Maxwell W, Bowe, Stephen J, and Clifton, Vicki L (2008). Male sex and pre-existing diabetes are independent risk factors for stillbirth, *Australian and New Zealand Journal of Obstetrics and Gynaecology,* 48 (4), 375-83.

European Commission, (2014). Seventh Report from the Commission to the Council and the European Parliament on the Statistics on the number of animals used for experimental and other scientific purposes in the member states of the European Union. http://ec.europa.eu/environment/chemicals/lab_animals/reports_en.htm

Faisal, M, Kim, H, and Kim, J (2014). Sexual differences of imprinted genes expression levels, *Gene,* 533 (1), 434-38.

Fall, C and Osmond, C (2013). Commentary: The developmental origins of health and disease: an appreciation of the life and work of Professor David JP Barker, 1938–2013, *International journal of epidemiology,* 42 (5), 1231-32.

Gabory, A, Roseboom, TJ, Moore, T, Moore, LG, and Junien, C (2013). Placental contribution to the origins of sexual dimorphism in health and diseases: sex chromosomes and epigenetics, *Biol Sex Differ,* 4 (5), 10.1186.

Gardner DK, Lane M (2014). Mammalian preimplantation embryo culture. In *Mouse Molecular Embryology* (pp. 167-182). Humana Press, Boston, MA.

Gardner DK, Schoolcraft WB (1999). Culture and transfer of human blastocysts. *Current Opinion in Obstetrics and Gynecology*, 11(3): 307-11.

Hales, C Nicholas and Barker, David JP (2001). The thrifty phenotype hypothesis, *British medical bulletin,* 60 (1).

Hesketh T, Xing ZW (2006). Abnormal sex ratios in human populations: causes and consequences. Proceedings of the National Academy of Sciences, 103(36), 13271-5.

Hughes IA. (2001) Minireview: sex differentiation. *Endocrinology*, 142(8), 3281-7.

Imdad A, Lassi Z, Salaam R, Bhutta ZA (2016). Prenatal Nutrition and Nutrition in Pregnancy: Effects on Long-Term Growth and Development, In: *Early Nutrition and Long-Term Health: Mechanisms, Consequences, and Opportunities*. http://doi.org/10.1016/B978-0-08-100168-4.00001-X

Jansson T, Powell TL. (2007) Role of the placenta in fetal programming: underlying mechanisms and potential interventional approaches. *Clinical science*, 113(1), 1-3.

King, Barry F and Hastings, Richard A (1977). The comparative fine structure of the interhemal membrane of chorioallantoic placentas from six genera of myomorph rodents, *American Journal of Anatomy,* 149 (2), 165-79.

Kalisch-Smith JI, Simmons DG, Dickinson H, Moritz KM (2016) Review: Sexual dimorphism in the formation, function and adaptation of the placenta, *Placenta*, in press.

Lamers WH, Mooren PG, Griep HA, Endert ER, Degenhart HJ, Charles RO (1986). Hormones in perinatal rat and spiny mouse: relation to altricial and precocial timing of birth. *American Journal of Physiology-Endocrinology and Metabolism*, 251(1): E78–85.

Lensch, M William, Schlaeger, Thorsten M, Zon, Leonard I, and Daley, George Q (2007). Teratoma formation assays with human embryonic stem cells: a rationale for one type of human-animal chimera, *Cell Stem Cell,* 1 (3), 253-58.

Logan, WPD (1950). Mortality in England and Wales from 1848 to 1947: A survey of the changing causes of death during the past hundred years, *Population Studies,* 4 (2), 132-78.

Mamrot J, Pangestu M, Walker D, Gardner DK, Dickinson H. (2015). Confirmed dioestrus in pseudopregnant mice using vaginal exfoliative cytology improves embryo transfer implantation rate. *Reproductive biomedicine online*, 31(4), 538-43.

Mandò, C., Calabrese, S., Mazzocco, M.I., Novielli, C., Anelli, G.M., Antonazzo, P. and Cetin I (2016). Sex specific adaptations in placental biometry of overweight and obese women. *Placenta*, *38*, pp.1-7.

Martín J, Cervero A, Mir P, Conejero M, Jose A, Pellicer A, Simón C (2013). The impact of next-generation sequencing technology on preimplantation genetic diagnosis and screening, *Fertility and sterility,* 99 (4), 1054-61. e3.

Martin, Yvonne, Gerlach, Gabriele, Schlötterer, Christian, and Meyer, Axel (2000). Molecular Phylogeny of European Muroid Rodents Based on Complete Cytochrome *b* Sequences, *Molecular Phylogenetics and Evolution,* 16 (1), 37-47.

McLaren A (1976). Genetics of the early mouse embryo. *Annual review of genetics*, 10(1): 361-88.

McLean AC, Valenzuela N, Fai S, Bennett SA (2012) Performing vaginal lavage, crystal violet staining, and vaginal cytological evaluation for mouse estrous cycle staging identification. *JoVE* (Journal of Visualized Experiments). 15 (67), e4389.

Moore KL, Persaud TV, Torchia MG. (2015) Before We Are Born: Essentials of Embryology and Birth Defects. *Elsevier Health Sciences*.

Muralimanoharan, S, Maloyan, A, and Myatt, L (2013). Evidence of sexual dimorphism in the placental function with severe preeclampsia, *Placenta,* 34 (12), 1183-89.

Murphy, Vanessa E, Gibson, Peter, Talbot, Philippa I, and Clifton, Vicki L (2005). Severe asthma exacerbations during pregnancy, *Obstetrics & Gynecology,* 106 (5, Part 1), 1046-54.

Murphy, Vanessa E, Gibson, Peter G, Giles, Warwick B, Zakar, Tamas, Smith, Roger, Bisits, Andrew M, Kessell, Carolyn G, and Clifton, Vicki L (2003). Maternal asthma is associated with reduced female fetal growth, *American journal of respiratory and critical care medicine,* 168 (11), 1317-23.

Navarro E, Funtikova AN, Fíto M, Schröder H (2017) Prenatal nutrition and the risk of adult obesity: Long-term effects of nutrition on epigenetic mechanisms regulating gene expression. The Journal of nutritional biochemistry, 39, 1-4.

O'Connell, B. A., Moritz, K. M., Roberts, C. T., Walker, D. W., and Dickinson, H. (2011). The placental response to excess maternal glucocorticoid exposure differs between the male and female conceptus in spiny mice', *Biology of reproduction,* 85 (5), 1040-7.

O'Connell, BA, Moritz, KM, Walker, DW, and Dickinson, H (2013a). Treatment of pregnant spiny mice at mid gestation with a synthetic glucocorticoid has sex-dependent effects on placental glycogen stores, *Placenta,* 34 (10), 932-40.

O'Connell, BA, Moritz, KM, Walker, DW, and Dickinson, H (2013b). Sexually dimorphic placental development throughout gestation in the spiny mouse (Acomys cahirinus), *Placenta,* 34 (2), 119-26.

O'Connell, Bree A, Moritz, Karen M, Walker, David W, and Dickinson, Hayley (2012). Synthetic Glucocorticoid Dexamethasone Inhibits Branching Morphogenesis in the Spiny Mouse Placenta 1, *Biology of reproduction,* 88 (1).

Oxford Dictionary (2017). *"chimera". Oxford Dictionaries* (accessed April 11, 2017). Oxford University. https://en.oxforddictionaries.com/definition/chimera

Quinn TA, Ratnayake U, Castillo-Melendez M, Moritz KM, Dickinson H, Walker DW (2014). Adrenal steroidogenesis following prenatal dexamethasone exposure in the spiny mouse. *Journal of endocrinology*, 221(2): 347-62.

Quinn, Tracey A, Ratnayake, Udani, Dickinson, Hayley, Nguyen, Tri-Hung, McIntosh, Michelle, Castillo-Melendez, Margie, Conley, Alan J, and Walker, David W (2013). Ontogeny of the adrenal gland in the spiny mouse, with particular reference to production of the steroids cortisol and dehydroepiandrosterone, *Endocrinology,* 154 (3), 1190-201.

Rice WR, Friberg U, Gavrilets S (2016) Sexually antagonistic epigenetic marks that canalize sexually dimorphic development, *Molecular ecology*, 25 (8), 1812-1822.

Roseboom, TJ, Painter, RC, De Rooij, SR, Van Abeelen, AFM, Veenendaal, MVE, Osmond, C, and Barker, DJP (2011). Effects of famine on placental size and efficiency, *Placenta,* 32 (5), 395-99.

Rosenfeld, Cheryl S (2011). Periconceptional influences on offspring sex ratio and placental responses, *Reproduction, Fertility and Development,* 24 (1), 45-58.

Saif, Zarqa, Hodyl, Nicolette A, Hobbs, Eleanor, Tuck, Astrud R, Butler, Miriam S, Osei-Kumah, Annette, and Clifton, Vicki L (2014). The human placenta expresses multiple glucocorticoid receptor isoforms that are altered by fetal sex, growth restriction and maternal asthma, *Placenta*, *35*(4), 260-268.

Sandman, Curt A, Glynn, Laura M, and Davis, Elysia Poggi (2013). Is there a viability–vulnerability tradeoff? Sex differences in fetal programming, *Journal of psychosomatic research,* 75 (4), 327-35.

Seckl, Jonathan R and Holmes, Megan C (2007). Mechanisms of disease: glucocorticoids, their placental metabolism and fetal 'programming' of adult pathophysiology, *Nature clinical practice Endocrinology & metabolism,* 3 (6), 479-88.

Solter, Davor and Knowles, Barbara B (1975). Immunosurgery of mouse blastocyst, *Proceedings of the National Academy of Sciences,* 72 (12), 5099-102.

Stanner, Sara A, Bulmer, K, Andres, C, Lantseva, Olga E, Borodina, V, Poteen, VV, and Yudkin, John S (1997). Does malnutrition in utero determine diabetes and coronary heart disease in adulthood? Results from the Leningrad siege study, a cross sectional study, *BMJ: British Medical Journal*, 1342-48.

Stark, Michael J, Clifton, Vicki L, and Wright, Ian (2009). Neonates born to mothers with preeclampsia exhibit sex-specific alterations in microvascular function, *Pediatric research,* 65 (3), 291-95.

Stark, Michael J, Dierkx, L, Clifton, Vicki Lee, and Wright, Ian MR (2006). Alterations in the maternal peripheral microvascular response in pregnancies complicated by preeclampsia and the impact of fetal sex, *Journal of the Society for Gynecologic Investigation,* 13 (8), 573-78.

Stevenson, DK, Verter, J, Fanaroff, AA, Oh, W, Ehrenkranz, RA, Shankaran, S, Donovan, EF, Wright, LL, Lemons, JA, and Tyson, JE (2000). Sex differences in outcomes of very low birthweight infants: the newborn male disadvantage, *Archives of Disease in Childhood-Fetal and Neonatal Edition,* 83 (3), F182-F85.

Surani MA, Hayashi K, Hajkova P (2007). Genetic and epigenetic regulators of pluripotency, *Cell*, 128 (4), 747-62.

Tachibana M, Sparman M, Ramsey C, Ma H, Lee HS, Penedo MC, Mitalipov S. (2012) Generation of chimeric rhesus monkeys. *Cell*. 148(1), 285-95.

Tanaka N, Takeuchi T, Neri QV, Sills ES, Palermo GD (2006). Laser-assisted blastocyst dissection and subsequent cultivation of embryonic stem cells in a serum/cell free culture system: applications and preliminary results in a murine model. *Journal of translational medicine*, 4(1):20.

Tarrade A, Panchenko P, Junien C, Gabory A (2015). Placental contribution to nutritional programming of health and diseases: epigenetics and sexual dimorphism. *Journal of Experimental Biology*, 218 (1), 50-8.

Torche, Florencia and Kleinhaus, Karine (2012). Prenatal stress, gestational age and secondary sex ratio: the sex-specific effects of exposure to a natural disaster in early pregnancy, *Human reproduction,* 27 (2), 558-67.

Townsel CD, Emmer SF, Campbell WA, Hussain N. (2017). Gender differences in respiratory morbidity and mortality of preterm neonates, *Frontiers in Pediatrics*, 5.

Truong TT, Soh YM, Gardner DK (2016). Antioxidants improve mouse preimplantation embryo development and viability. *Human Reproduction*. 31(7), 445-54.

Vaughan, Owen R, Sferruzzi-Perri, AN, Coan, PM, and Fowden, AL (2011). Environmental regulation of placental phenotype: implications for fetal growth, *Reproduction, Fertility and Development,* 24 (1), 80-96.

Verburg, P.E., Tucker, G., Scheil, W., Erwich, J.J.H., Dekker, G.A. and Roberts, C.T., (2016). Sexual Dimorphism in Adverse Pregnancy Outcomes-A Retrospective Australian Population Study 1981-2011. *PLoS One*, *11*(7), p.e0158807.

Walker KJ, Green MP, Gardner DK (2015). Spatial asynchronous transfer of cleavage-stage mouse embryos to the uterus compromises fetal development. *Molecular reproduction and development*. 82(2), 80.

Wallace, J. M., Horgan, G. W., and Bhattacharya, S. (2012). Placental weight and efficiency in relation to maternal body mass index and the risk of pregnancy complications in women delivering singleton babies, *Placenta,* 33 (8), 611-18.

Wells JC. (2000). Natural selection and sex differences in morbidity and mortality in early life. *Journal of theoretical Biology*, 202(1), 65-76.

Yuron Y, Lehavi O, Orr-Urtreger A, Gull I, Lessing JB, Amit A, Ben-Yosef D. (2002). Maternal serum HCG is higher in the presence of a female fetus as early as week 3 post-fertilization. *Human reproduction*, 17(2), 485-9.

Zheng, Y.L., Jiang, M.X., Ouyang, Y.C., Sun, Q.Y., and Chen, D.Y. (2005). Production of mouse by inter-strain inner cell mass replacement, *Zygote,* 13 (1), 73-78.

# Chapter 2: Embryo transfer

## 2.1 Overview

This publication describes the development of methods to improve the implantation rate after embryo transfer in the common mouse (*Mus musculus*). The original source of this work is:

Mamrot, J., Pangestu, M., Walker, D., Gardner, D. K., & Dickinson, H. (2015). Confirmed dioestrus in pseudopregnant mice using vaginal exfoliative cytology improves embryo transfer implantation rate. *Reproductive biomedicine online*, *31*(4), 538-543. DOI: https://doi.org/10.1016/j.rbmo.2015.06.020

Reproductive BioMedicine Online (2015) **31**, 538–543

www.sciencedirect.com
www.rbmonline.com

ARTICLE

# Confirmed dioestrus in pseudopregnant mice using vaginal exfoliative cytology improves embryo transfer implantation rate

**Jared Mamrot [a,b,*], Mulyoto Pangestu [b], David Walker [a,b], David K Gardner [c], Hayley Dickinson [a,b]**

[a] *The Ritchie Centre, Hudson Institute of Medical Research, Clayton, Australia;* [b] *Department of Obstetrics and Gynaecology, Monash University, Clayton, Australia;* [c] *School of BioSciences, The University of Melbourne, Parkville, Australia*

* Corresponding author. *E-mail address:* jared.mamrot@hudson.org.au (J Mamrot).

Jared Mamrot obtained a Masters in Clinical Embryology from Monash University in 2012, and began his PhD in The Ritchie Centre for Perinatal Research in 2013. His research interests include embryo micromanipulation and embryo transfer, with a strong focus on assisted reproduction and developmental physiology.

**Abstract** Embryo transfer is a commonly performed surgical technique. In mice, protocols typically specify pairing recipient females with vasectomized males to induce a receptive uterine environment for embryo implantation. However, this induced receptive state is not always maintained until implantation occurs. The use of a well-characterized correlation between oestrous state and exfoliative vaginal cytology was therefore evaluated to assess uterine receptivity immediately before embryo transfer. Eight- to 12-week-old virgin female CD1 mice ($n = 22$) were paired overnight with vasectomized males and successfully mated, indicated by the presence of a vaginal plug. These dams underwent embryo transfer 3 days later with embryos obtained from superovulated 4-week-old $F_1$ (C57BL/6 × CBA) females. Non-invasive vaginal lavage was conducted immediately before transfer. Dams were killed 6 days after transfer and the uterus collected for histological analysis. Embryo implantation rate in mice was 96% when cytological analysis of the lavage samples signified dioestrus ($n = 6$), whereas the implantation rate was <15% ($n = 16$) when cytology signified other stages of oestrous. This simple, quick, non-invasive measure of receptivity was accurate and easily adopted and, when applied prospectively, will avoid unnecessary surgery and subsequent culling of non-suitable recipients, while maximizing the implantation potential of each recipient female.



**KEYWORDS:** blastocyst, mouse, optimization, pregnancy, recipient, smear

http://dx.doi.org/10.1016/j.rbmo.2015.06.020
1472-6483/

## Introduction

Successful embryo transfer requires a receptive uterine environment to support embryo implantation. In female laboratory animals inducing a receptive state is typically accomplished by mating with a sterile male (Behringer et al., 2014; Tarkowski, 1959). This process initiates a cascade of physiological changes termed 'pseudopregnancy', with the uterine decidua undergoing substantial remodelling in preparation for implantation of blastocysts (Paria et al., 2002; Wang and Dey, 2006). The successful induction of pseudopregnancy is typically assessed by the presence of a vaginal mucous plug the day after pairing (Bronson and McLaren, 1970; Hogan et al., 1986; Yang et al., 2009); however, this does not guarantee that the female will remain in the pseudopregnant state until implantation of embryos occurs, which is approximately 4 days after mating in mice, as pseudopregnant females have been found to re-enter the oestrous cycle in response to various stimuli (Whitten, 1956). Verifying that recipient mice have remained in the receptive 'dioestrus' state until the time of embryo transfer may provide a key measure of the suitability of the recipient to undergo the procedure and consequently increase the overall success of the transfer process.

The mouse oestrous cycle can be observed indirectly through changes in the reproductive tract. Cyclic changes in epithelial cell structure within the vaginal and uterine lumen have been characterized in many species, with cell type and relative prevalence correlating strongly with oestrous state (Byers et al., 2012; Gal et al., 2014; Nelson et al., 1982; Papanicolaou, 1933). This relationship was first documented in the guinea pig by Stockard and Papanicolaou (1917), and shortly after in the rat (Long and Evans, 1922) and mouse (Allen, 1922). The criteria used to determine oestrous state in these landmark papers has remained relatively unchanged (Bertolin and Murphy, 2013; Cora et al., 2015; Thung et al., 1956); however, its application in embryo transfer experiments has not been routine because vaginal cell sampling typically requires instrumental penetration of the vagina (Caligioni, 2009; Nelson et al., 1982), which can induce inflammatory processes, resulting in loss of the dioestrus state (Bertolin and Murphy, 2013; McLean et al., 2012). McLean and colleagues have recently addressed the 'negative' aspects of the sampling technique, proposing a protocol for non-invasive sampling of exfoliative epithelial cells in mice (McLean et al., 2012). Therefore, we hypothesized that embryo transfer done at the time of cytologically-proven dioestrus would result in an increased implantation rate, and application of this non-invasive protocol would allow accurate staging of the oestrous cycle while avoiding the unwanted effects of penetrative sampling via the vagina.

## Materials and methods

### Animals and treatments

All experiments were approved by the Monash Medical Centre Animal Ethics Committee on 6 June 2013 (reference number MMC A- 2011/84) and conducted in accordance with the 8th edition of the Australian Code of Practice for the care and use of animals for scientific purposes (2013). Mice were obtained from Monash Animal Services (Clayton, Victoria, Australia), and housed in high barrier specific pathogen free (SPF) housing ≤4 per cage, 12 h light cycle (lights on at 08:00 h) at 22–23°C, 40–55% relative humidity, and food and water were available *ad libitum*.

### Inducing pseudopregnancy

Pseudopregnancy was induced by pairing 8- to 12-week-old CD1 females with vasectomized CD1 males. Oestrus was first induced in the females with the introduction of soiled bedding from a sexually mature male cage 3 days prior to mating (Whitten, 1956). CD1 males were housed one per cage, with a female introduced at 14:00 h and separated at 09:30 h the following morning. Females were independently examined for the presence of a vaginal plug and plug-positive mice were moved into conventional housing for 3 days prior to embryo transfer. The morning of plug detection was designated day 1 of pseudopregnancy, with embryo transfer conducted on the morning of day 4 (~72 h after plug detection). Non-plugged mice were returned to stock for re-pairing >2 weeks later.

### Vaginal lavage

Avoiding contact with the vagina, 20 μl of sterile saline was repeatedly expelled (three to five times) onto the vaginal opening and re-drawn up into the pipette tip (McLean et al., 2012). The fluid was then expelled onto a glass slide and smeared using the pipette tip to facilitate evaporation at room temperature.

For the purposes of this study, vaginal lavage samples were not stained or examined until after embryo transfers had been performed to reduce potential experimenter bias.

### Exfoliative vaginal cell staining

After lavage fluid had evaporated, cells were fixed with 10% neutral buffered formalin for 5 min. Smeared lavage fluid was stained with haematoxylin and eosin. Cells were rehydrated in deionized water for 15 s, stained with Harris haematoxylin (Amber Scientific, Victoria, Australia) for 6 min, differentiated with acid ethanol and submersed in basic ammonium for up to 10 s. Slides were counterstained with 1% aqueous eosin (Amber Scientific, Victoria, Australia) for 3 min, then dehydrated, cleared and coverslipped using DPX mounting medium. Slides were scanned using Aperio ScanScope (Leica Biosystems, NSW, Australia) and viewed using Aperio ImageScope software (Leica Biosystems; ver 11).

### Cytological analysis

Oestrous state was determined from vaginal exfoliative cytology smears (as shown in **Figure 1**) following specific criteria based on Byers et al. (2012). Proestrus was defined as the presence of nucleated epithelial cells and leukocytes and the absence of cornified epithelial cells. Dioestrus was defined as the presence of leukocytes only, or with leukocytes

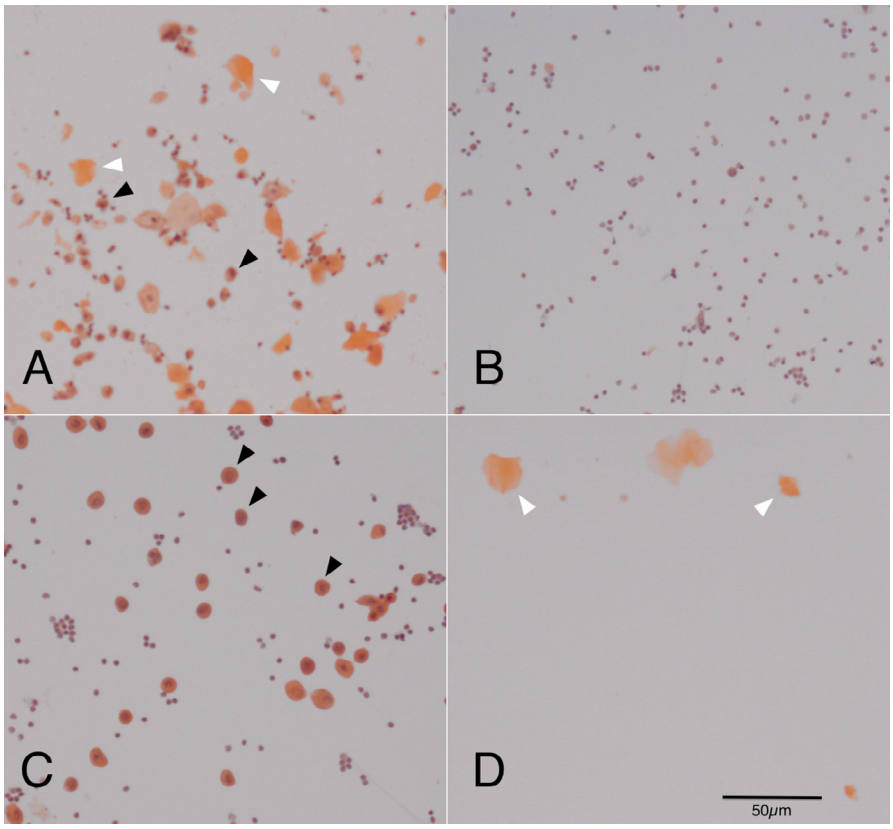



**Figure 1** Representative images of haematoxylin/eosin-stained exfoliative vaginal cytology corresponding to: (A) metoestrus, with anucleated cornified epithelial cells (white arrows), nucleated epithelial cells (black arrows) and leukocytes present; (B) dioestrus, with leukocytes the most prevalent cell type, with minimal presence of nucleated epithelial cells; (C) proestrus, with nucleated cells (black arrows) and leukocytes present and anucleated cornified epithelial cells absent; (D) oestrus, with anucleated cornified epithelial cells only (white arrows).

comprising 91% or more of the total cells, and these other <9% of cells being both cornified epithelial cells and nucleated epithelial cells. Metoestrus was defined as the presence of all three cell types, with leukocytes comprising 90% or less of the total cells. Oestrus was defined as the presence of cornified epithelial cells alone.

The number of anucleated cornified epithelial cells, nucleated epithelial cells and leukocytes were counted on the centre 0.25 $mm^2$ area of each slide (corresponding to ×200 magnification), with one slide processed per animal. Image J version 1.48 (Rasband, 1997–2014) was used to quantify cell type and number with each image undergoing threshold adjustment, binary conversion and watershed adjustment to clearly identify individual cells. Cells were analysed by volume: leukocytes between 10–80 $\mu m^2$ in size, nucleated epithelial cells between 90–500 $\mu m^2$ and cornified squamous epithelial cells were counted manually. Detailed instructions for cell counting using Image J are available online (e.g. imagej.net/Particle_Analysis).

## Generating embryos for transfer

Donor embryos were generated as previously described (Gardner and Lane, 2014). Four-week-old $F_1$ (C57BL/6 × CBA) female mice were given an intraperitoneal (i.p.) injection of 5 IU pregnant mare serum gonadotrophin (PMSG; Intervet, Australia) and, 48 h later, an intraperitoneal (i.p.) injection of 5 IU human chorionic gonadotrophin (HCG; Intervet, Australia). They were then paired with $F_1$ (C57BL/6 × CBA) hybrid males; pairing was conducted 1 day prior to pairing of recipient females to induce pseudopregnancy. At 1.5 days postcoitus (dpc), donor females were killed by cervical dislocation, oviducts were excised and embryos were flushed into MOPS-buffered medium G1 (Gardner and Lane, 2014) supplemented with human serum albumin (HSA, 5 mg/ml; Vitrolife). Embryos were cultured at 37°C, 6% $CO_2$, 5% $O_2$, 89% $N_2$ in groups of 10 embryos per 20 μl G1 microdrop supplemented with HSA (5 mg/ml) and covered with Ovoil™ liquid paraffin (Vitrolife). After 24 h, when embryos were at the 8–16 cell stage, they were transferred into 20 μl G2 embryo culture medium droplets supplemented with HSA (5 mg/ml), with 10 embryos per 20 μl microdrop, as outlined in Gardner and Lane (2014). Expanded blastocysts of excellent morphology were selected for transfer (graded as per Gardner and Lane, 2014); morphologically abnormal blastocysts were excluded.

## Embryo transfer

Uterine transfer, of day 5 embryos, was performed immediately after vaginal lavage of the pseudopregnant CD1 recipient

females on day 4 after pairing, i.e., day 5 embryos into day 4 recipient female. Anaesthesia was induced in the recipient mice with 4.9% isofluorane in air, and was maintained throughout the experiment with 2.5% isofluorane in air. Four to eight embryos at the blastocyst stage were transferred into the lumen of each uterine horn through a small dorsal incision using a pulled and flame-polished glass Pasteur pipette. Following embryo transfer, the peritoneum was sutured (Maxon, Covidien, US), and the skin wound closed with sterile surgical clips. The procedure was conducted on a warmed stage and typically took <20 min. Lignocaine (Xylocaine) was applied topically (0.1 ml), and mice recovered for 30 min in a dark cage at 37°C before being transferred to their original cage. After transfer, recipients had 3 mg/ml paracetamol added to their drinking water and were monitored for signs of distress. All females were killed 6 days post-surgery (day 10 post-pairing), with the uterus excised and fixed in 10% neutral buffered formalin post-mortem.

## Uterine morphology

After fixation in formalin for 48 h, each uterus was processed through graded ethanols, xylene and paraffin wax and sectioned at 4 μm. Representative sections were stained to verify the number of implantation sites. Slides were dewaxed through xylene and graded ethanol in preparation for Masson's trichrome staining. Samples were post-fixed in Bouin's fluid for 1 h at room temperature then stained with Weighert's iron haematoxylin (Amber Scientific, Victoria, Australia) for 10 min and differentiated in 1% acetic acid. Slides were then stained with beibrich scarlet-acid fuchsin (POCD, NSW, Australia) for 5 min and differentiated in phosphotungstic-phosphomolybdic acid (POCD, NSW, Australia), before being stained with aniline blue (POCD, NSW, Australia) for 2 min, dehydrated, cleared and coverslipped in DPX mounting medium. Slides were scanned using Aperio ScanScope slide scanner and viewed using ImageScope (version 11). Embryo implantation sites were counted for each uterine horn in the stained sections by three independent scorers. Scorers were blinded to the cytological data obtained.

## Statistics

Statistical analysis was conducted using GraphPad Prism (version 6.0, GraphPad Software, San Diego, CA). The Kruskal-Wallis non-parametric statistical comparison was used, as the data did not meet normality and homogeneity of variance needed for parametric analysis. Data are presented as means ± SEM with significance established at $P < 0.05$.

# Results

Representative exfoliative vaginal cytology for each stage of the oestrous cycle is presented in **Figure 1**. The relative prevalence of each cell type was used to ascertain the oestrous state for each recipient mouse (**Table 1**).

When vaginal cytology was retrospectively analysed to determine oestrous state (**Table 1**), the implantation rate in

**Table 1** Relative epithelial cell type and prevalence for each embryo transfer recipient sorted by oestrus state, with vaginal lavage smears, conducted immediately prior to embryo transfer ($n = 22$).

| *Mouse number* | *Cornified epithelial cells* | *Nucleated epithelial cells* | *Leukocytes* | *Leukocytes/total cells (%)* | *Oestrus state* | *Implantation/embryo* |
|---|---|---|---|---|---|---|
| 10 | 0 | 2 | 15 | 88.2 | Proestrus | 0/8 |
| 19 | 0 | 25 | 290 | 92.1 | Proestrus | 2/10 |
| 20 | 0 | 24 | 1060 | 97.8 | Proestrus | 1/10 |
| 1 | 4 | 0 | 0 | 0.00 | Oestrus | 0/10 |
| 2 | 2 | 0 | 0 | 0.00 | Oestrus | 2/10 |
| 9 | 1 | 0 | 0 | 0.00 | Oestrus | 0/8 |
| 17 | 2 | 0 | 0 | 0.00 | Oestrus | 0/10 |
| 18 | 5 | 0 | 0 | 0.00 | Oestrus | 0/10 |
| 22 | 1 | 0 | 0 | 0.00 | Oestrus | 1/10 |
| 3 | 196 | 6 | 1450 | 87.8 | Metoestrus | 2/10 |
| 5 | 5 | 1 | 22 | 78.6 | Metoestrus | 0/10 |
| 6 | 5 | 2 | 69 | 90.8 | Metoestrus | 6/10 |
| 7 | 1 | 5 | 15 | 71.4 | Metoestrus | 0/5 |
| 12 | 36 | 1 | 4 | 9.8 | Metoestrus | 0/8 |
| 13 | 30 | 1 | 7 | 18.4 | Metoestrus | 0/8 |
| 16 | 3 | 1 | 20 | 83.3 | Metoestrus | 3/10 |
| 4 | 27 | 12 | 1919 | 98.0 | Dioestrus | 10/10 |
| 8 | 0 | 0 | 200 | 100.0 | Dioestrus | 13/13 |
| 11 | 0 | 0 | 72 | 100.0 | Dioestrus | 7/8 |
| 14 | 60 | 8 | 687 | 91.0 | Dioestrus | 8/8 |
| 15 | 0 | 0 | 80 | 100.0 | Dioestrus | 9/10 |
| 21 | 4 | 14 | 1409 | 98.7 | Dioestrus | 10/10 |

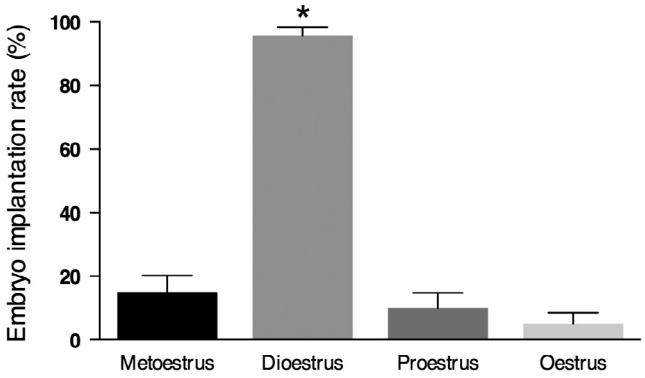



**Figure 2** Implantation rate for female recipients with oestrous state determined by exfoliative vaginal cytology. *Significantly different from all other stages of oestrus; $P < 0.05$.

dioestrus-stage females ($n = 6$) was 96 ± 2% (Figure 2). The average implantation rate for females not in dioestrus at the time of embryo transfer ($n = 16$) was 10 ± 4%, with no difference in implantation rate detected between uterine horns (left versus right, data not shown) or between non-dioestrus states (Figure 2).

In all dioestrus recipients ($n = 6$) histological evaluation of pregnancy confirmed implantation sites for 57 of 59 embryos transferred. Representative stained histological sections displaying individual implantation sites within the uterus are presented in Figure 3.

## Discussion

This study describes a simple non-invasive technique for determining the oestrous state in female recipient mice prior to embryo transfer. It was shown that this established technique allows dioestrus to be determined quickly and accurately, and that the decision to conduct embryo transfers on the basis of these findings will significantly increase the efficacy of the embryo transfer process. The results show that simple lavage of the vaginal opening provides a sufficient number of cells for accurate determination of oestrous state, with the result that embryo transfer done at the time of histologically verified dioestrus is highly efficient. The implantation rate in females was 96% when retrospective cytological analysis signified dioestrus, whereas, if the presence of the copulatory plug alone is used to assess suitability for use, the average implantation rate for all recipient mice used in this study was 36%. A large number of common stressors have been reported to affect the maintenance of pseudopregnancy in plug-positive pseudopregnant females, such as described in detail by Whitten (1956), yet the results of this study show that objective confirmation of dioestrus prior to surgery can reduce variation in embryo transfer experiment outcomes, and increase the overall implantation rate independent of the method of embryo transfer employed.

Based on the original protocol for embryo transfer (McLaren and Michie, 1956), a large number of publications have reported improved success rates from technical improvements and optimizations to the embryo transfer procedure, such as non-surgical transcervical embryo transfer (Green et al., 2009; Moreno-Moya et al., 2014), and embryo transfer via the utero-tubal junction (Bermejo-Alvarez et al., 2014). Innovations such as these report near-perfect implantation rates, however these success rates remain predicated on the continued presence of dioestrus in the pseudopregnant recipients. Conducting an objective assessment of this aspect of uterine receptivity prior to embryo transfer surgery in mice merits inclusion in current and future embryo transfer protocols and optimizations.

This simple, quick and non-invasive exfoliative cytological analysis requires minimal training to implement, with an instructional video of the technique available (McLean et al., 2012). Quantifying cell type and number in cytological smears is quick and simple to conduct, minimizing bias and subjectivity, with the definitions proposed here shown to accurately reflect defined oestrous states. The key benefits of implementing this technique, regardless of the specific protocol used to conduct embryo transfer, is minimizing variation in embryo transfer outcome and minimizing the number of mice required. As embryo transfer in mice is one of the most frequently conducted surgical procedures

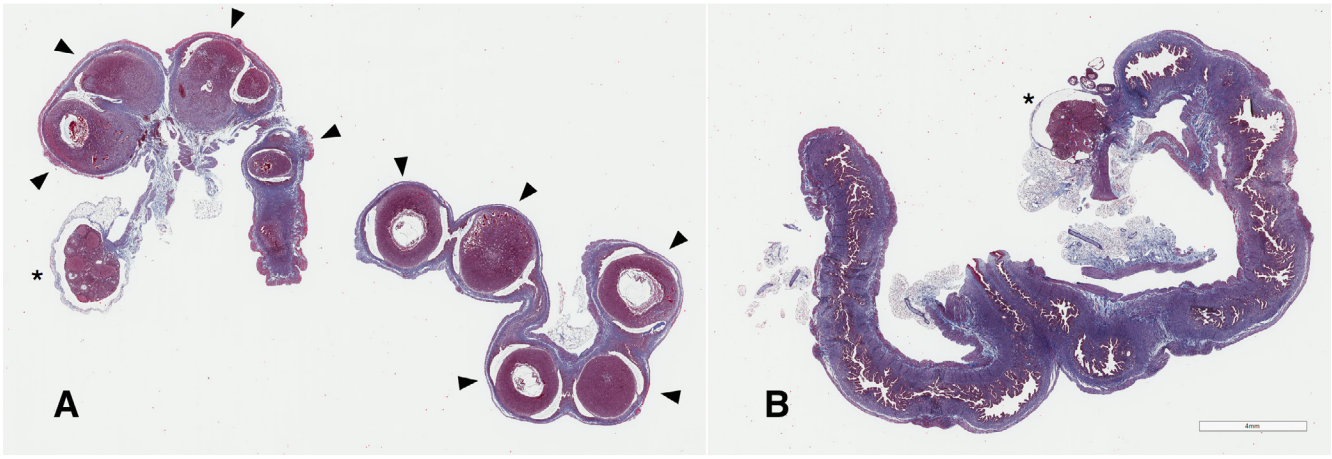



**Figure 3** Representative stained uterine sections from female mice in dioestrus (A) and oestrus (B) at day 10 of gestation. Implantation sites are indicated with black arrows, and ovaries indicated with asterisks.

worldwide (Department of Primary Industries, 2014; European Commission, 2014; U.K. Government: Home Office, 2013), this technique will substantially reduce the number of animals needed in such experiments; not only because of the increased implantation success, but also because early identification of non-suitable recipients will allow their use elsewhere.

## Acknowledgements

The authors acknowledge Hudson Institute of Medical Research, Monash University and the Operational Infrastructure Support Scheme (OIS), as well as the scientific and technical assistance of Mai Truong, and the assistance and support of Lesley Wiadrowski and the Histology Facility, Hudson Institute of Medical Research.

## References

Allen, E., 1922. The oestrous cycle in the mouse. Am. J. Anat. 30, 297–371.

Behringer, R., Gertenstein, M., Nagy, K.V., Nagy, A., 2014. Manipulating the Mouse Embryo: A Laboratory Manual, fourth ed. Cold Spring Harbour Laboratory Press, New York.

Bermejo-Alvarez, P., Park, K.E., Telugu, B.P., 2014. Utero-tubal embryo transfer and vasectomy in the mouse model. J. Vis. Exp. 84, e51214.

Bertolin, K., Murphy, B.D., 2013. Monitoring mouse estrous cycles. In: Croy, A., Yamada, A.T., Demayo, F.J., Adamson, S.L. (Eds.), The Guide to Investigation of Mouse Pregnancy. Academic Press, San Diego, p. 375.

Bronson, R.A., McLaren, A., 1970. Transfer to the mouse oviduct of eggs with and without the zona pellucida. J. Reprod. Fertil. 22, 129–137.

Byers, S.L., Wiles, M.V., Dunn, S.L., Taft, R.A., 2012. Mouse estrous cycle identification tool and images. PLoS ONE 7, e35538.

Caligioni, C.S., 2009. Assessing reproductive status/stages in mice. Curr. Protoc. Neurosci. Appendix 4: Appendix 4I.

Cora, M.C., Kooistra, L., Travlos, G., 2015. Vaginal cytology of the laboratory rat and mouse review and criteria for the staging of the estrous cycle using stained vaginal smears. Toxicol. Pathol. pii:0192623315570339.

Department of Primary Industries, 2014. Statistics of Animal Use in Research and Teaching, Report Number 30, 1 January 2012 to 31 December 2012.

European Commission, 2014. Seventh Report from the Commission to the Council and the European Parliament on the Statistics on the number of animals used for experimental and other scientific purposes in the member states of the European Union. <http://ec.europa.eu/environment/chemicals/lab_animals/reports_en.htm> (accessed 02.12.14).

Gal, A., Lin, P.C., Barger, A.M., MacNeill, A.L., Ko, C., 2014. Vaginal fold histology reduces the variability introduced by vaginal exfoliative cytology in the classification of mouse estrous cycle stages. Toxicol. Pathol. 42, 1212–1220.

Gardner, D.K., Lane, M., 2014. Culture of the mammalian embryo. Methods Mol. Biol. 1092, 167–182.

Green, M.A., Bass, S., Spear, B.T., 2009. Transcervical transfer of mouse embryos using the non-surgical embryo transfer (NSET™) device. Biotechniques 52.

Hogan, B., Costantini, F., Lacy, E., 1986. Manipulating the Mouse Embryo: A Laboratory Manual. Cold Spring Harbor Laboratory Press, New York.

Long, J.A., Evans, H.M., 1922. The Oestrous Cycle in the Rat and Its Associated Phenomena, vol. 6. University of California Press.

McLaren, A., Michie, D., 1956. Studies on the transfer of fertilized mouse eggs to uterine foster-mothers. 1. Factors affecting the implantation and survival of native and transferred eggs. J. Exp. Biol. 33, 394–416.

McLean, A.C., Valenzuela, N., Fai, S., Bennett, S.A.L., 2012. Performing vaginal lavage, crystal violet staining, and vaginal cytological evaluation for mouse estrous cycle staging identification. J. Vis. Exp. 67, e4389. doi:10.3791/4389.

Moreno-Moya, J.M., Ramírez, L., Vilella, F., Martínez, S., Quiñonero, A., Noguera, I., Pellicer, A., Simón, C., 2014. Complete method to obtain, culture, and transfer mouse blastocysts nonsurgically to study implantation and development. Fertil. Steril. 101, e13.

Nelson, J.F., Felicio, L.S., Randall, P.K., Sims, C., Finch, C.E., 1982. A longitudinal study of estrous cyclicity in aging C57BL/6J mice: I. Cycle frequency, length and vaginal cytology. Biol. Reprod. 27, 327–339.

Papanicolaou, G.N., 1933. The sexual cycle in the human female as revealed by vaginal smears. Am. J. Anat. 52 (S1), 519–637.

Paria, B.C., Reese, J., Das, S.K., Dey, S.K., 2002. Deciphering the cross-talk of implantation: advances and challenges. Science 296, 2185–2188.

Rasband, W.S., 1997–2014. ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA. <http://imagej.nih.gov/ij/> (accessed 04.17.14).

Stockard, C.R., Papanicolaou, G.N., 1917. The existence of a typical oestrous cycle in the guinea-pig – with a study of its histological and physiological changes. Am. J. Anat. 22, 225–283.

Tarkowski, A.K., 1959. Expirements on the transplantation of ova in mice. Acta Theriol. 2, 251–267.

Thung, P.J., Boot, L.M., Muhlbock, O., 1956. Senile changes in the oestrous cycle and in ovarian structure in some inbred strains of mice. Acta Endocrinol. 23, 8–32.

U.K. Government: Home Office, 2013. Annual statistics of scientific procedures on living animals Great Britain 2012. <http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/212610/spanimals12.pdf> (accessed 02.12.14).

Wang, H., Dey, S.K., 2006. Roadmap to embryo implantation: clues from mouse models. Nat. Rev. Genet. 7, 185–199.

Whitten, W.K., 1956. Modification of the oestrous cycle of the mouse by external stimuli associated with the male. J. Endocrinol. 13, 399–404.

Yang, J.J., Larsen, C.M., Grattan, D.R., Erskine, M.S., 2009. Mating-induced neuroendocrine responses during pseudopregnancy in the female mouse. J. Neuroendocrinol. 21, 30–39.

*Declaration: The authors report no financial or commercial conflicts of interest.*

Received 14 April 2015; refereed 17 June 2015; accepted 17 June 2015.

# Chapter 3: Transcriptome assembly

## 3.1 Overview

This manuscript describes the assembly and annotation of a transcriptome for the spiny mouse. The original source of this work is:

Mamrot, J., Legaie, R., Ellery, S.J., Wilson, T., Seemann, T., Powell, D.R., Gardner, D.K., Walker, D.W., Temple-Smith, P., Papenfuss, A.T. and Dickinson, H. (2017). De novo transcriptome assembly for the spiny mouse (Acomys cahirinus). *Scientific Reports*, *7*(1), p.8996. DOI: https://doi.org/10.1038/s41598-017-09334-7

SCIENTIFIC REPORTS

OPEN

eceived: 28 April 2017
ccepted: 17 July 2017
ıblished online: 21 August 2017

# *De novo* transcriptome assembly for the spiny mouse (*Acomys cahirinus*)

Jared Mamrot[1,2], Roxane Legaie[3], Stacey J. Ellery[1,2], Trevor Wilson[4], Torsten Seemann[5], David R. Powell[6], David K. Gardner[7], David W. Walker[1,2,13], Peter Temple-Smith[2,8], Anthony T. Papenfuss[9,10,11,12] & Hayley Dickinson[1,2]

**Spiny mice of the genus *Acomys* display several unique physiological traits, including menstruation and scar-free wound healing; characteristics that are exceedingly rare in mammals, and of considerable interest to the scientific community. These unique attributes, and the potential for spiny mice to accurately model human diseases, are driving increased use of this genus in biomedical research, however little genetic information is accessible for this species. This project aimed to generate a draft transcriptome for the Common spiny mouse (*Acomys cahirinus*). Illumina sequencing of RNA from 15 organ types (male and female) produced 451 million, 150 bp paired-end reads (92.4Gbp). An extensive survey of *de novo* transcriptome assembly approaches using Trinity, SOAPdenovo-Trans, and Oases at multiple kmer lengths was conducted, producing 50 single-kmer assemblies from this dataset. Non-redundant transcripts from all assemblies were merged into a meta-assembly using the EvidentialGene *tr2aacds* pipeline, producing the largest gene catalogue to date for *Acomys cahirinus*. This study provides the first detailed characterization of the spiny mouse transcriptome. It validates use of the EvidentialGene *tr2aacds* pipeline in mammals to augment conventional *de novo* assembly approaches, and provides a valuable scientific resource for further investigation into the unique physiological characteristics inherent in the genus *Acomys*.**

The Common or Cairo spiny mouse (*Acomys cahirinus*) is a small rodent species endemic to the semi-arid deserts of Africa and the Middle East[1]. Used in research to model human disease, spiny mice exhibit physiological characteristics not typically found in rodents: they exhibit a precocial pattern of development[2,3], atypical synthesis of hormones such as cortisol and dehydroepiandosterone[4–6], and a menstrual cycle[7]. These traits are common to humans and other higher order primates, but rare in other mammals. For example, menstruation has been identified in only six non-primate species (from >5,000 extant mammals), none of which are rodents[8]. The discovery of human-like physiological characteristics in a rodent is highly valuable for those in the scientific community looking to model human conditions, however fundamental aspects of their biology remain unexplored; for instance, there is little genetic information accessible for this species.

Publically available genetic information for the spiny mouse consists of the mitochondrial genome[9], and two RNA sequencing (RNA-Seq) datasets: PRJNA184055[10], and PRJNA292021[11]. These next-generation sequencing (NGS) datasets were created with specific aims: to establish incipient sympatric speciation as a mode of natural selection in mammals inhabiting divergent microclimates[9], to examine the molecular basis for natural variation in mammalian lifespan[10], and to characterize and investigate another characteristic unique to *Acomys*: scar-free wound healing and skin regeneration[11]. *De novo* assembly of NGS reads was conducted for each specific organ/tissue sequenced in these projects in order to investigate differential gene expression, however the accuracy

[1]The Ritchie Centre, Hudson Institute of Medical Research, Melbourne, Australia. [2]Department of Obstetrics and Gynaecology, Monash University, Melbourne, Australia. [3]MHTP node - Monash Bioinformatics Platform, Monash University, Melbourne, Australia. [4]MHTP Medical Genomics Facility, Melbourne, Australia. [5]Melbourne Bioinformatics, The University of Melbourne, Melbourne, Australia. [6]Monash Bioinformatics Platform, Monash University, Melbourne, Australia. [7]School of BioSciences, University of Melbourne, Melbourne, Australia. [8]Education Program in Reproduction and Development, Monash University, Melbourne, Australia. [9]Bioinformatics Division, Walter and Eliza Hall Institute, Parkville, Australia. [10]Computational Cancer Biology Program, Peter MacCallum Cancer Centre, Melbourne, Australia. [11]Department of Medical Biology, University of Melbourne, Melbourne, Australia. [12]Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, Australia. [13]Present address: RMIT University, Bundoora Campus, Bundoora, Australia. Anthony T. Papenfuss and Hayley Dickinson jointly supervised this work. Correspondence and requests for materials should be addressed to H.D. (email: hayley.dickinson@hudson.org.au)

| | Forward | Reverse | Total |
|---|---|---|---|
| Yield total (Gb) | 45.9 | 45.9 | 91.77617 |
| Aligned (%) | 0.45 | 0.25 | 0.3509741 |
| Error rate (%) | 0.62 | 1.6 | 1.003602 |
| Intensity cycle 1 | 3255 | 3054 | 3154.379 |
| %>=Q30 | 80.5 | 46.5 | 63.47869 |
| Total raw read pairs | | | 451,182,406 |
| Total read pairs (Q>=30) | | | 305,920,540 |
| GC content | | | 45% |

**Table 1.** Spiny mouse RNA-Seq summary statistics.

and completeness of resulting assemblies was not explicitly described. Accurate identification of differentially expressed genes is dependent on accurate read mapping[12], and an accurate reference assembly requires transcripts from multiple organ types.

Here, we describe a survey of *de novo* transcriptome assembly methods, utilizing both single-kmer and multi-kmer approaches, with the aim to generate a comprehensive *de novo* transcriptome assembly for the Common spiny mouse (*Acomys cahirinus*).

## Results

**Sample preparation and sequencing.** Tissues were collected from male (n = 1), non-pregnant female (n = 1) and placenta from 2 pregnant female (1 male fetus, 1 female fetus) adult spiny mice in accordance with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes with approval from the Monash Medical Centre Animal Ethics Committee. Total RNA was extracted from skin, lung, liver, small intestine, kidney, adrenal gland, brain, thymus, spleen, diaphragm, heart, skeletal muscle (male only), testis (male only), ovary (female only), and placenta. All samples returned RNA integrity numbers (RIN scores) >7.0 and were pooled for sequencing. Samples were not multiplexed at the time of sequencing due to cost. Each tissues RNA sample is stored individually and able to be resequenced. The Illumina HiSeq. 1500 was used to produce 150 bp paired-end reads. In total, 451 million read pairs were generated, with yield, proportion aligned, error rate, intensity, and GC content provided in Table 1. RNA-Seq reads are available from the NCBI as Bioproject PRJNA342864, run accessions SRR4279903 and SRR4279904. Summary statistics for data yield, percent pass-filter (%PF), raw cluster percentage per lane, and quality score summary are provided in Supplementary Table S1. Filtering of poor quality reads (Q < 30) removed 32% of the original 451 million read pairs, with 305 million high-quality paired reads used for assembly. FastQC reports for raw, filtered and *in silico* normalized data are provided in Supplementary Figure S1.

***De novo* transcriptome assembly.** A detailed protocol describing the assembly and validation of the spiny mouse RNA-Seq dataset is available at https://dx.doi.org/10.17504/protocols.io.ghebt3e. This protocol provides a brief description, documentation, citations, dependencies/requirements, parameters and commands used for all software employed in this manuscript.

In total, 50 unique single-kmer transcriptome assemblies were produced from 305 million paired reads, with and without digital normalization and read error correction, as described in Fig. 1. Detailed metrics for all transcriptome assemblies are provided in Supplementary Table S2. Digital normalization using Trinity reduced the size of the dataset by >80%, however assemblies constructed using normalized data contained fewer 'Benchmarking Universal Single-Copy Orthologs' (BUSCOs)[13] (Figs 2 and 3), had decreased backmapping rates ('backmapping' is aligning the reads used for *de novo* assembly to the assembled transcripts) (Fig. 4), decreased mapping of independent spiny mouse read data (Fig. 5), and worse TransRate scores[14] (Fig. 6), compared to assemblies generated from unnormalized data. Size distribution of assembled transcripts comprising each assembly is described in Fig. 7.

Clustering highly similar Trinity[15] contiguous sequences (contigs) using CD-HIT-EST[16,17] resulted in a modest reduction in the total number of Trinity contigs, with the majority of clustered contigs corresponding to transcript isoforms (Supplementary Figure S2). Clustering highly similar contigs increased the proportion of single copy BUSCO orthologs detected, however it also increased the number of fragmented BUSCO orthologs (Supplementary Figure S3).

Read sequencing errors identified using probabilistic error correction program SEECER[18] were comprised of 14,821,705 substitutions (4.84%), 1,760,162 deletions (0.57%), and 1,614,908 insertions (0.53%), affecting 6% of reads in total. Error correction provided a modest improvement to BUSCO score (Figs 2 and 3) and mapping of independent reads when assembled using Trinity (Fig. 4), however it also resulted in slightly poorer backmapping rate and TransRate score, compared to assemblies generated from non-corrected reads (Figs 5 and 6).

Trinity produced the largest and most complete single-kmer assemblies. The Trinity_v2.3.2 assembly contained the greatest number of BUSCOs (Figs 2 and 3), the highest proportion of back-mapped reads (Fig. 4), and the highest proportion of aligned RNA-Seq reads from the National Center for Biotechnology Information (NCBI) projects PRJNA184055 and PRJNA292021 (Fig. 5). The distribution of contigs is more negatively skewed, with more contigs of larger size (Figs 7 and 8). The Trinity_v2.3.2 assembly also has the highest number of 'Basic Local Alignment Search Tool' (BLAST) hits, and unique 'single best' BLAST hits (the single highest scoring assembled

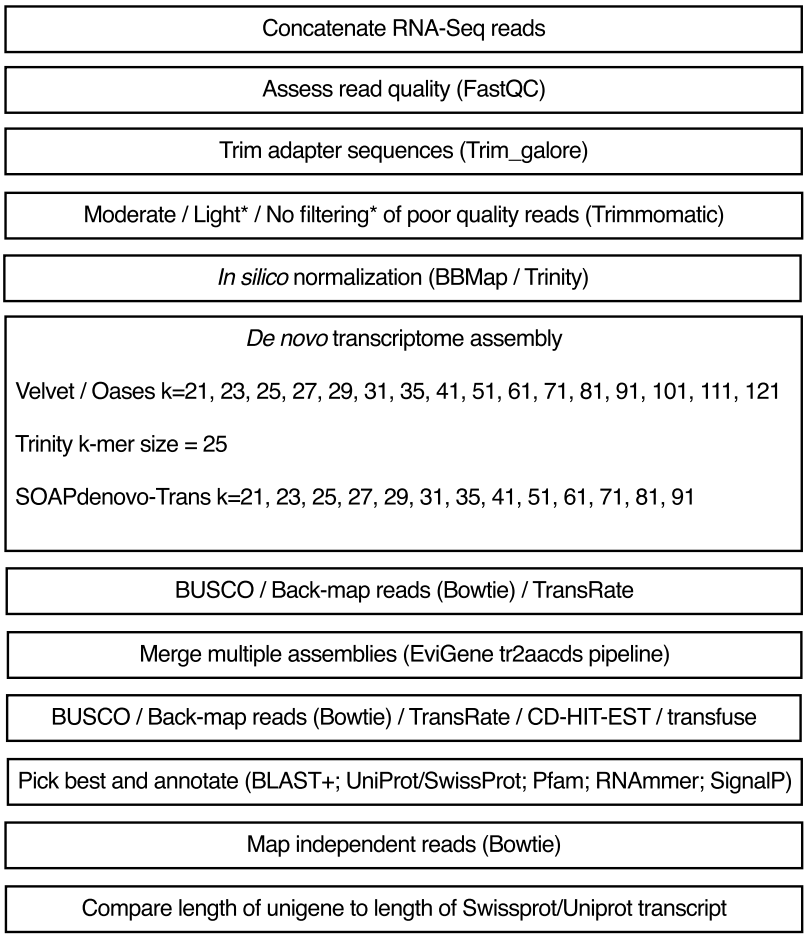



**Figure 1.** Flow chart of transcriptome assembly pipeline. *SEECER probabilistic error correction conducted on these datasets.

transcript alignment for each database entry above a specified significance threshold), compared to *Mus musculus* RefSeq transcripts, UniProtKB/SwissProt database[19,20] and UniRef90 database[21,22] (Figs 9, 10 and 11). Gene Ontology (GO) terms corresponding to annotated transcripts were well distributed between the categories of biological process, cellular component and molecular function (Supplementary Figure S4).

**Collating non-redundant transcripts from multiple assemblies.** Merging non-redundant transcripts from all assemblies using the EvidentialGene 'transcript to amino acid coding sequence' (*tr2aacds*) pipeline[23,24] (Supplementary Table S2) increased the proportion of complete BUSCOs found, and reduced the number of fragmented and missing BUSCOs (Fig. 3). The BUSCO values obtained are consistent with the most complete reference transcriptomes from other vertebrate and eukaryote taxa (BUSCO[13]: Supplementary Online Material).

**Annotation and identification of non-coding RNAs.** The most accurate and complete single-kmer assembly was produced by Trinity_v2.3.2 from the 'non-normalized' dataset. It contains 2,219,978 contigs (2,026,183 'genes' as defined by Trinity), representing a 1.29 Gb transcriptome. Of these, 546,398 transcripts were identified as non-coding by the 'Coding-Non-Coding Index' (CNCI[25]) corresponding to 44,572 unique NONCODE ncRNAs[26], 277,565 transcripts aligned to UniProtKB/SwissProt entries (BLASTx, 'Expect value' (e-value) $\leq 1 \times 10^{-20}$), and 145,658 transcripts contained an open reading frame (ORF). Many UniProtKB/SwissProt entries have more than one aligning transcript from the assembly: the highest scoring transcript for each UniProtKB/SwissProt entry (the 'single best' BLAST hit) was identified and collated. In total, 28,847 unique 'single best' alignments to the UniProtKB/SwissProt database were identified, 21,762 of which contain an ORF. In comparison, the *tr2aacds*-generated assembly ("tr2aacds_v2") contains fewer than half the number of transcripts compared to the Trinity_v2.3.2 assembly, with 1,034,437 transcripts in total, representing a 491 Mb transcriptome. Of these, 258,400 transcripts were identified as non-coding by CNCI corresponding to 31,642 unique NONCODE ncRNAs, 135,139 transcripts show significant sequence similarity to the UniProtKB/SwissProt database (e-value $\leq 1 \times 10^{-20}$), and 117,250 transcripts contain an ORF. Alignment to the UniProtKB/SwissProt database revealed 22,852 unique 'single best' transcripts, 21,135 of which contain an ORF.

Despite containing fewer 'single best' UniProtKB/SwissProt-aligning transcripts, analysis of single best transcripts revealed the tr2aacds_v2 assembly contains the greatest number of full-length transcripts (transcript

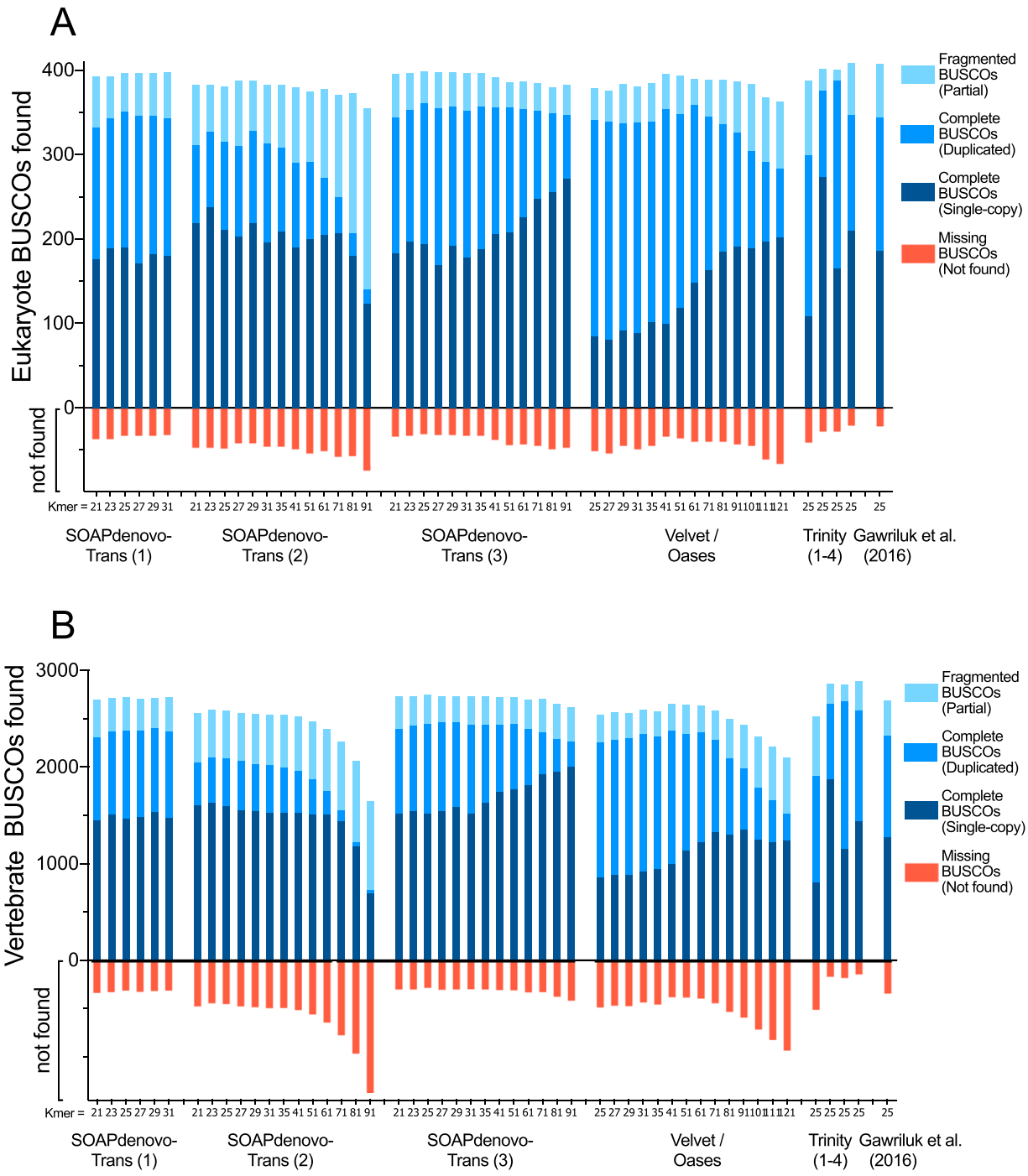



**Figure 2.** BUSCOs identified in each single-kmer transcriptome assembly. SOAPdenovo-Trans (1): reads prior to quality filtering; SOAPdenovo-Trans (2): light filtering of poor quality read pairs, combined with *in silico* normalization; SOAPdenovo-Trans (3): moderate filtering of poor quality read pairs; Velvet/Oases: moderate filtering of poor quality read pairs; Trinity (1): light filtering of poor quality read pairs, combined with *in silico* normalization; Trinity (2): light filtering of poor quality read pairs, combined with SEECER error correction; Trinity (3): moderate filtering of poor quality read pairs; Trinity (4): light filtering of poor quality read pairs, assembled with ver2.3.2.

length >90% of the reference UniProtKB/SwissProt entry) (Fig. 10). Combined with a greater number of complete BUSCOs, and fewer fragmented BUSCOs, this result suggests the tr2aacds_v2 assembly is the most accurate and complete catalogue of protein-coding transcripts for the spiny mouse.

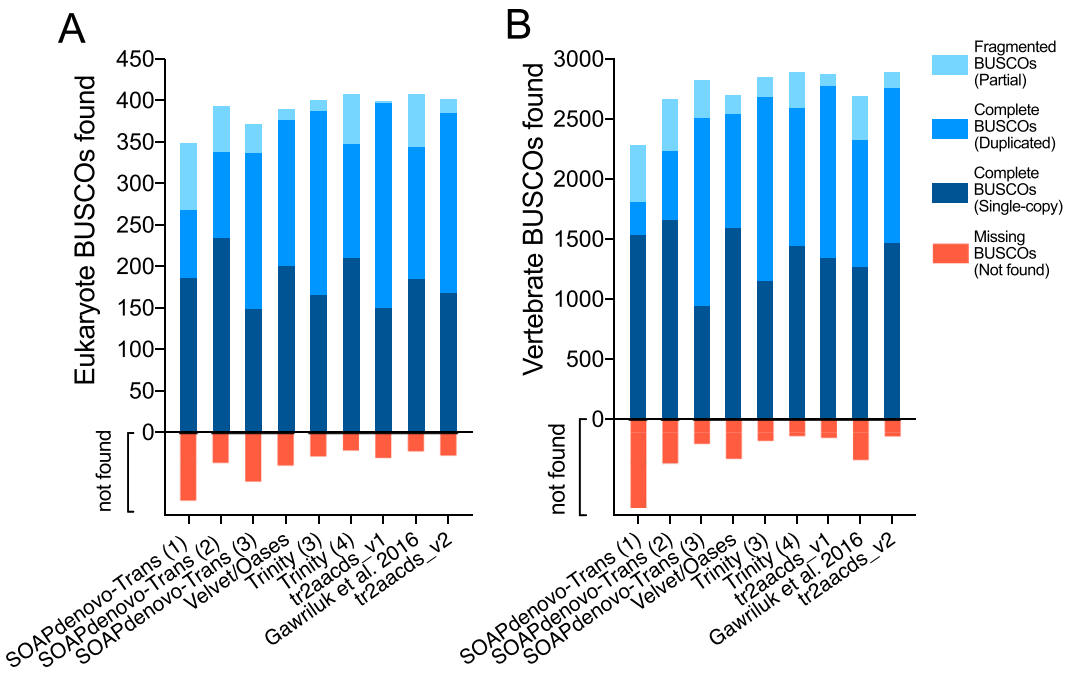



**Figure 3.** BUSCOs identified in each clustered/multi-kmer transcriptome assembly. SOAPdenovo-Trans (1): no filtering; (2): light filtering of poor quality read pairs, combined with *in silico* normalization; (3): moderate filtering of poor quality read pairs; Velvet/Oases: moderate filtering of poor quality read pairs; Trinity (1): light filtering of poor quality read pairs, combined with *in silico* normalization; (2): light filtering of poor quality read pairs, combined with SEECER error correction; (3): moderate filtering of poor quality read pairs; (4) moderate filtering of poor quality reads and SEECER error correction; tr2aacds merged assemblies (1): pre-v2.3.2 Trinity; (2) including v2.3.2 Trinity.

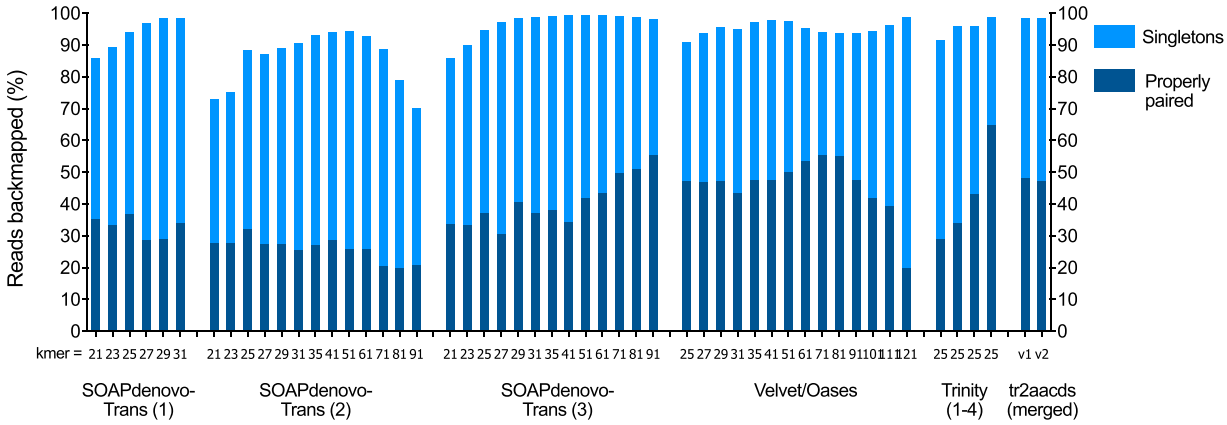



**Figure 4.** Proportion of reads backmapping to each transcriptome assembly. SOAPdenovo-Trans (1): no filtering; (2): light filtering of poor quality read pairs, combined with *in silico* normalization; (3): moderate filtering of poor quality read pairs; Velvet/Oases: moderate filtering of poor quality read pairs; Trinity (1): light filtering of poor quality read pairs, combined with *in silico* normalization; (2): light filtering of poor quality read pairs, combined with SEECER error correction; (3): moderate filtering of poor quality read pairs; (4): moderate filtering of poor quality reads and SEECER error correction; tr2aacds merged assemblies (1): pre-v2.3.2 Trinity; (2): including v2.3.2 Trinity.

## Discussion

The transcriptome assemblies produced and validated here comprise an important new resource for spiny mouse research, increasing the value and accessibility of this species as an animal model in biomedical science. In total, 50 assemblies were produced using three *de novo* assemblers: Trinity[15], SOAPdenovo-Trans[27], and Oases[28]. Combining unique assembled transcripts from all single-kmer assemblies using EvidentialGene *tr2aacds* produced the largest collection of full-length protein-coding transcripts. Each transcriptome performed well in

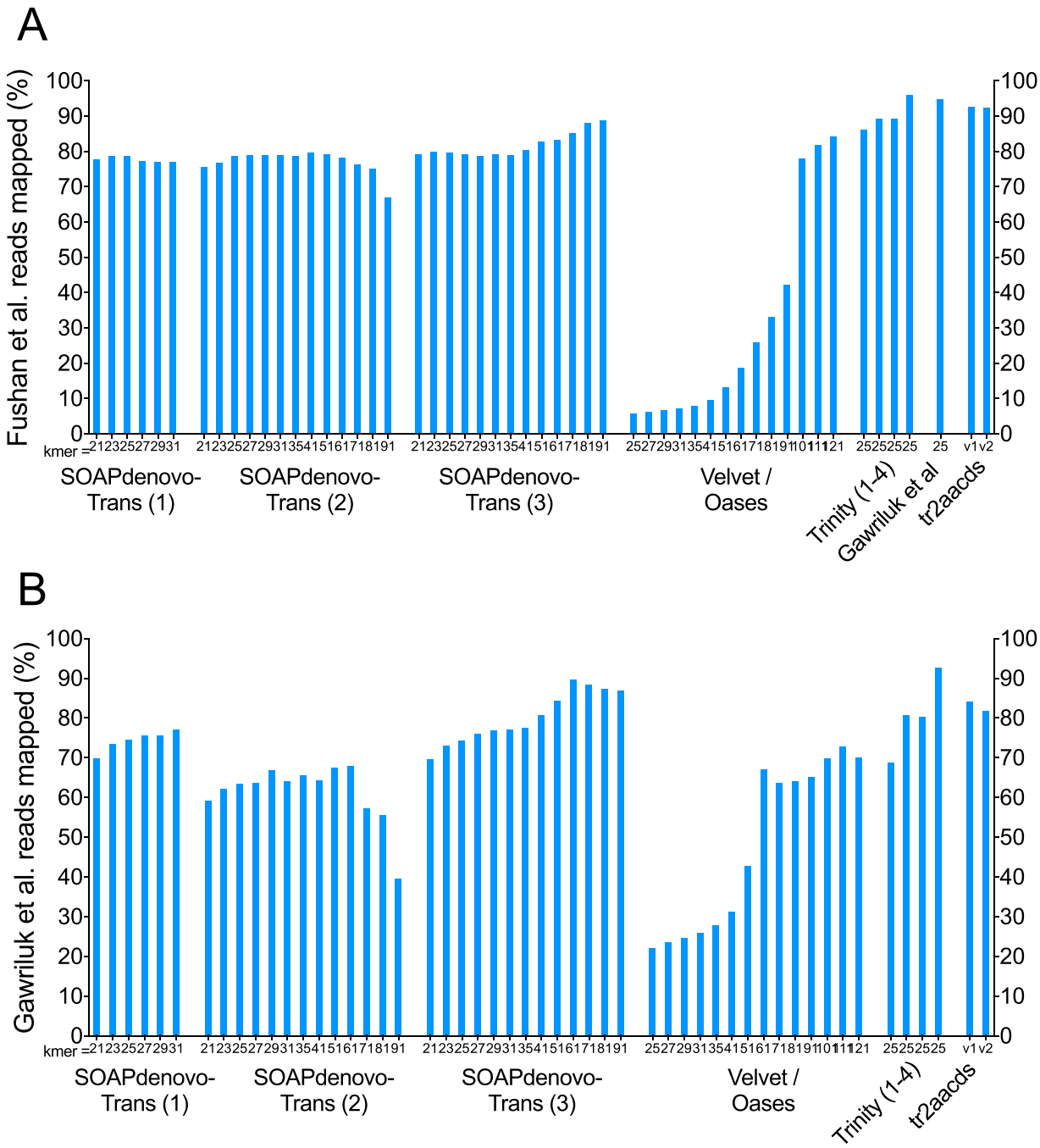



**Figure 5.** Proportion of independent reads mapping to each transcriptome assembly for (**A**) PRJNA184055 (Fushan *et al.*, 2015), and (**B**) PRJNA300275 (Gawriluk *et al.*, 2016). SOAPdenovo-Trans (1): no filtering; (2): light filtering of poor quality read pairs, combined with *in silico* normalization; (3): moderate filtering of poor quality read pairs; Velvet/Oases: moderate filtering of poor quality read pairs; Trinity (1): light filtering of poor quality read pairs, combined with in silico normalization; (2): light filtering of poor quality read pairs, combined with SEECER error correction; (3): moderate filtering of poor quality read pairs; (4): moderate filtering of poor quality reads and SEECER error correction; (5): Gawriluk *et al.* transcriptome assembly; tr2aacds merged assemblies (1): pre-v2.3.2 Trinity; (2): including v2.3.2 Trinity.

measures of assembly integrity and completeness (e.g. BUSCO, TransRate, backmapping and BLAST[29,30]), however TransRate scores were lower than expected for high-quality *de novo* transcriptome assemblies.

The approximate median TransRate score for assemblies uploaded to the NCBI Transcriptome Shotgun Assembly database is 0.2, however the best scoring assemblies were between 0.15 and 0.2. Potential explanations for the lower-than-expected TransRate scores are the proportion of read errors identified by SEECER, and quality of the RNA-Seq reads. TransRate scoring is contingent on accurate alignment of reads to transcripts, as it evaluates assemblies based on whether each base has been called correctly, whether bases are truly part of transcripts,

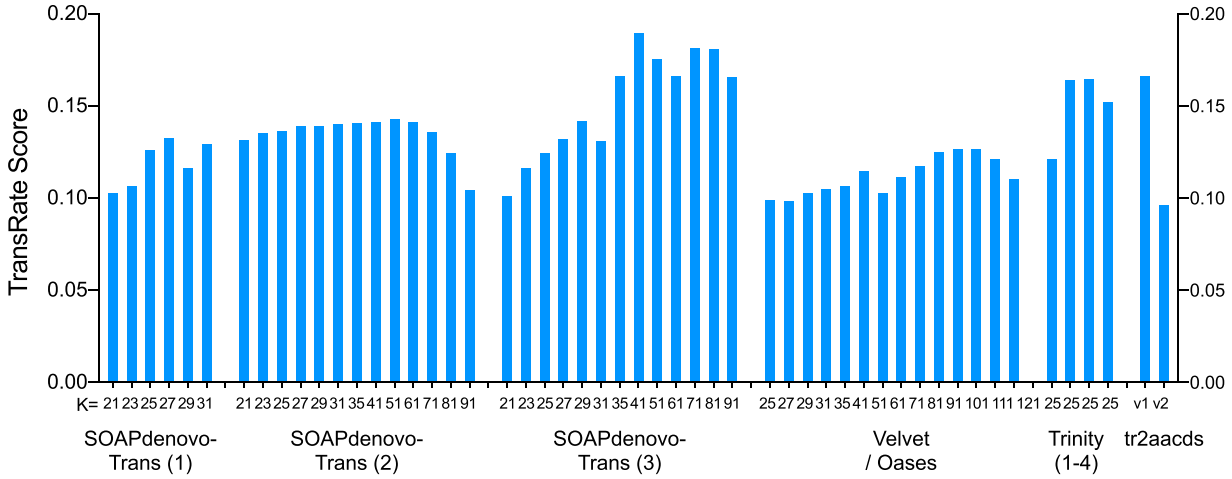



**Figure 6.** TransRate scores for each transcriptome assembly. SOAPdenovo-Trans (1): no filtering; (2): light filtering of poor quality read pairs, combined with *in silico* normalization; (3): moderate filtering of poor quality read pairs; Velvet/Oases: moderate filtering of poor quality read pairs; Trinity (1): light filtering of poor quality read pairs, combined with *in silico* normalization; (2): moderate filtering of poor quality read pairs; (3): light filtering of poor quality read pairs, combined with SEECER error correction; (4) Trinity v2.3.2 after moderate filtering of poor quality read pairs; (5) Gawriluk *et al*. transcriptome assembly; tr2aacds merged assemblies (1): pre-v2.3.2 Trinity; (2) including v2.3.2 Trinity.

whether contigs are derived from a single transcript, and whether contigs are structurally complete and correct[14]. Alignment rates calculated by TransRate were below the alignment rates expected (based on read alignment using Bowtie), and this may have negatively impacted the TransRate score. Using error-corrected reads for TransRate alignment may increase TransRate scores compared to uncorrected reads[31], however this was not examined in the present study.

The largest catalogue of 'full length' (>90%) transcripts aligning to the UniProtKB/SwissProt database was produced using the EvidentialGene *tr2aacds* pipeline. This finding correlates with similar projects incorporating *tr2aacds*-based meta-assembly, with more accurate and complete gene sets produced compared to transcriptomes assembled with a single software package (for instance, the mosquito *Aedes aegypti*: http://arthropods.eugenes.org/EvidentialGene/arthropods/mosquito/aedes_aegypti/). Generating accurate and complete transcripts is fundamental for gene annotation, and for subsequent identification of gene function, however a transcriptome assembly is comprised of more than protein-coding transcripts alone. Non-coding transcripts such as micro-RNAs and long non-coding RNAs perform essential roles in cellular function, with novel investigative methodologies driving increased interest in this area. The Trinity_v2.3.2 assembly contained the largest number of non-coding transcripts of all assemblies produced, and this resource will be made available in addition to the tr2aacds_v2 assembly.

A fundamental goal in generating this dataset is to facilitate access to spiny mouse transcript sequence information for external collaborators and researchers. The sequence reads and metadata are available from the NCBI (PRJNA342864) and assembled transcriptomes (Trinity_v2.3.2 and tr2aacds_v2) are available from the Zenodo repository (https://doi.org/10.5281/zenodo.808870), however accessing and utilizing this data can be challenging for researchers lacking bioinformatics expertise. To address this problem we are hosting a SequenceServer[32] BLAST-search website (http://spinymouse.erc.monash.edu/sequenceserver/). This resource provides a user-friendly interface to access sequence information from the tr2aacds_v2 assembly (to explore annotated protein-coding transcripts) and/or the Trinity_v2.3.2 assembly (to explore non-coding transcripts).

The public spiny mouse BLAST database has already been used by the spiny mouse research community. One member of our research group has used this resource to successfully design quantitative real-time polymerase chain reaction (qPCR) primer sets for lung-specific genes, markers of hypoxia, inflammation and apoptosis, and nuclear coding genes associated with mitogenesis (n = 41; personal communication). Sanger sequencing of PCR products confirmed >85% success under standard qPCR conditions, which is a significant improvement compared to ~40% success rate reported for primer design based on homologous regions from human, mouse (*Mus musculus*) and rat (*Rattus norvegicus*)[33–35]. Another collaborator requested the sequence of the spiny mouse protein-coding Beta Amyloid transcript, the product of which is implicated in the etiology of Alzheimer's disease. Prior to release of our dataset, the Gawriluk *et al*. transcriptome assembly was the only potential source for this information, however a complete Beta Amyloid transcript was not found in the assembly. The complete transcript from our database shows biological variation between the spiny mouse, *Mus musculus* (NM_001198823.1), and human orthologs (M15532.1).

RNA-Seq provides an unprecedented opportunity for cost-effective, large-scale genetic analysis in non-model organisms for which a genome sequence is unavailable. *De novo* assembly of millions/billions of RNA-Seq reads into a reference transcriptome can provide a valuable scientific resource, with applications in phylogenetics[36], novel gene identification[37], RNA editing[38] and alternative splicing investigation[39], qPCR primer design[40],

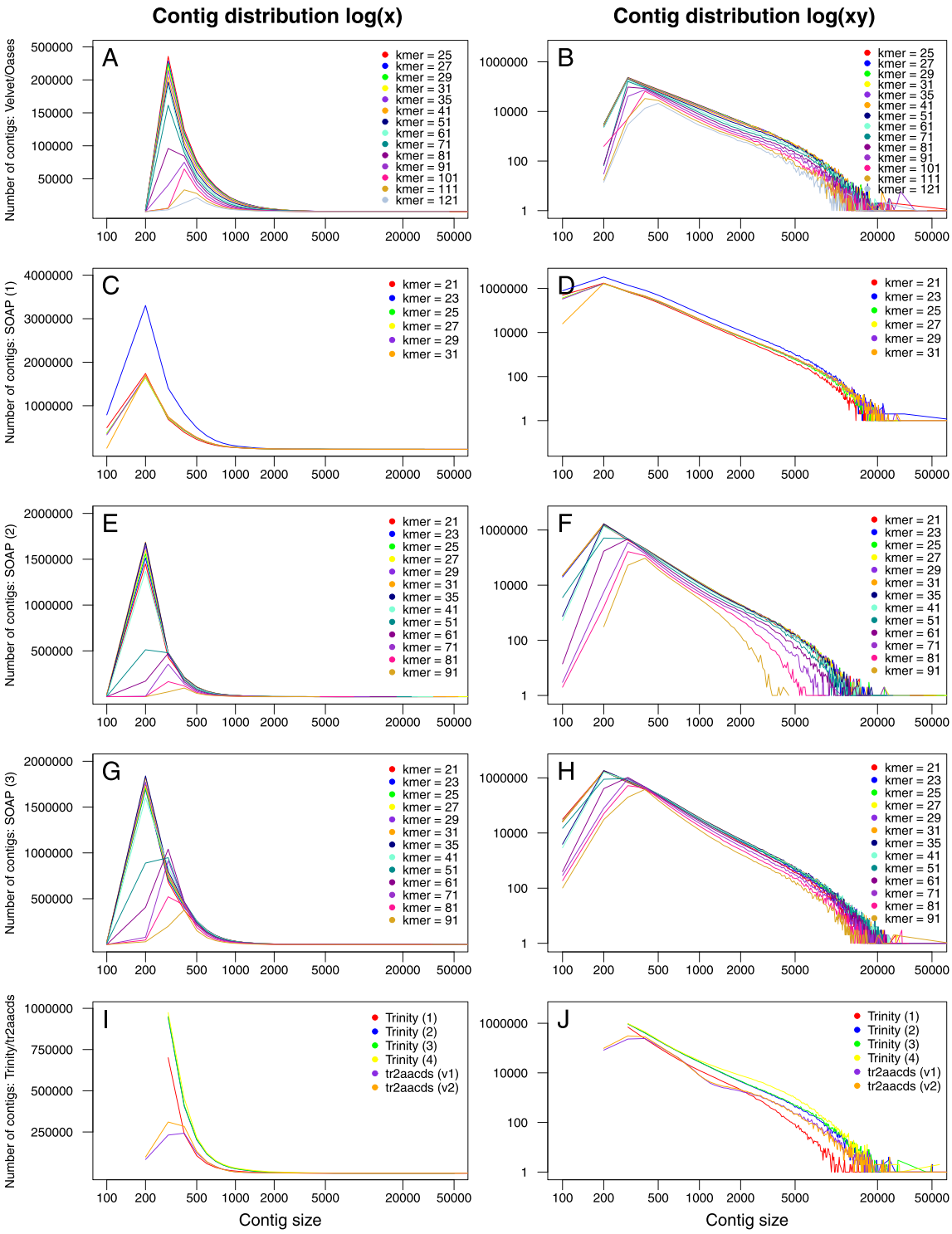



**Figure 7.** Distribution of contig number and size for all assemblies. Velvet/Oases: moderate filtering of poor quality read pairs; SOAPdenovo-Trans (1): no filtering; (2): light filtering of poor quality read pairs, combined with *in silico* normalization; (3): moderate filtering of poor quality read pairs; Trinity (1): light filtering of poor quality read pairs, combined with *in silico* normalization; (2): moderate filtering of poor quality read pairs; (3): light filtering of poor quality read pairs, combined with SEECER error correction; (4) Trinity v2.3.2 after moderate filtering of poor quality read pairs; tr2aacds merged assemblies (v1): pre-v2.3.2 Trinity; (v2) including v2.3.2 Trinity output.

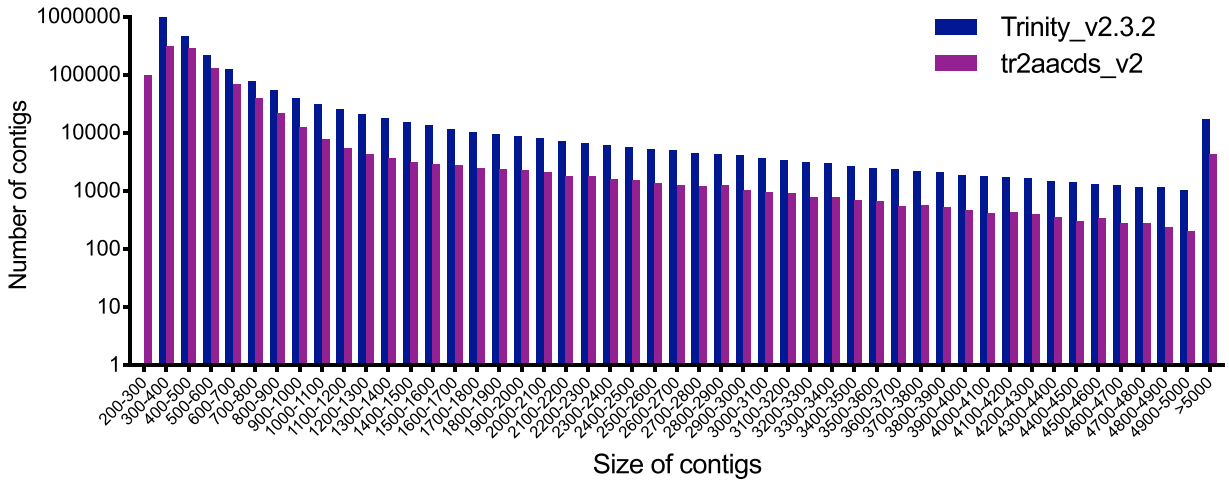



**Figure 8.** Comparison of contig size and number for Trinity_v2.3.2 (moderate filtering of poor quality read pairs) and tr2aacds_v2 (including v2.3.2 Trinity output) assemblies.

development and refinement of bioinformatics software[41], augmenting proteomic research[42], and investigation of gene expression profiles underlying complex physiological traits[43,44]. The utility of a transcriptome assembly is highly dependent on its completeness and accuracy, however there is not yet consensus within the transcriptomics community on a universal 'gold-standard' protocol or quality metric for transcriptome assembly. Many popular *de novo* assembly software packages provide detailed protocols for their use (eg. Trinity[45]), and practical guidelines such as the comprehensive Oyster River protocol[31] provide clear advice and guidance for optimising *de novo* assembly (http://oyster-river-protocol.readthedocs.io/en/latest/), however substantial customization and optimization of the sequencing and assembly pipeline is often necessary to produce high-quality, meaningful results[46,47]. The transcriptome assembly pipeline optimised for this project is easily accessible (https://dx.doi.org/10.17504/protocols.io.ghebt3e), facilitating reproduction and replication in other species.

## Conclusion

In conclusion, we have generated the most accurate and complete *de novo* transcriptome for the spiny mouse (*Acomys cahirinus*) to date, using the combined output of three *de novo* transcriptome assemblers: Trinity, SOAPdenovo-Trans, and Oases. All assemblies produced (n = 50) were analysed for accuracy and completeness, and validated using multiple quality metrics. The highest quality single-kmer transcriptome was generated using Trinity (v2.3.2). It is comprised of 2,219,978 transcripts, representing a 1.29 Gb transcriptome. The EvidentialGene *tr2aacds* pipeline was effective in identifying and collating unique transcripts from all 50 assemblies, producing a 491 Mb transcriptome comprised of 1,034,437 transcripts. This meta-assembly contained a greater number of full-length protein-coding transcripts than all individual single-kmer assemblies. This is the first study to implement the EvidentialGene *tr2aacds* pipeline to augment transcriptome assembly in a mammal. This study has produced the largest gene catalogue to date for the spiny mouse, providing an important resource for medical research. This dataset is now being used to further investigate physiological traits unique to the spiny mouse.

## Methods

**Data processing.** The protocol used to assemble and validate the spiny mouse RNA-Seq dataset is available at https://dx.doi.org/10.17504/protocols.io.ghebt3e.

Sequence reads were quality checked using FastQC v0.11.3 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapter sequences and low quality bases (Q < 20) were trimmed from 3′ ends using trim-galore (ver: 0.4.0; http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), which implements cutadapt v0.9.5[48]. Reads with average quality scores lower than 20 and read pairs in which either forward or reverse reads were trimmed to fewer than 35 nucleotides were discarded. Remaining reads were assessed again using FastQC, to ensure adapter sequences are eliminated.

Further filtering of poor quality reads was conducted using Trimmomatic v0.30[49] with settings "LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 AVGQUAL:30 MINLEN:35". Nucleotides with quality scores lower than 3 were trimmed from the 3′ and 5′ read ends. Reads with an average quality score lower than 30, and reads with a total length of fewer than 35 nucleotides after trimming were removed. Probabilistic error correction was performed on trimmed/filtered reads using SEECER[18] with default parameters. Both corrected and uncorrected reads were subjected to *de novo* assembly.

***De novo* transcriptome assembly.** Reads were assembled using either SOAPdenovo-Trans v1.03[27], Trinity[15] package r20140413p1 and Trinity package v2.3.2 (available at https://sourceforge.net/projects/trinityrnaseq/files/PREV_CONTENTS/previous_releases/ and http://trinityrnaseq.github.io), or Velvet v1.2.10[50]/Oases v0.2.08[28] with default parameters, except where indicated. The single-kmer assemblies were performed with and

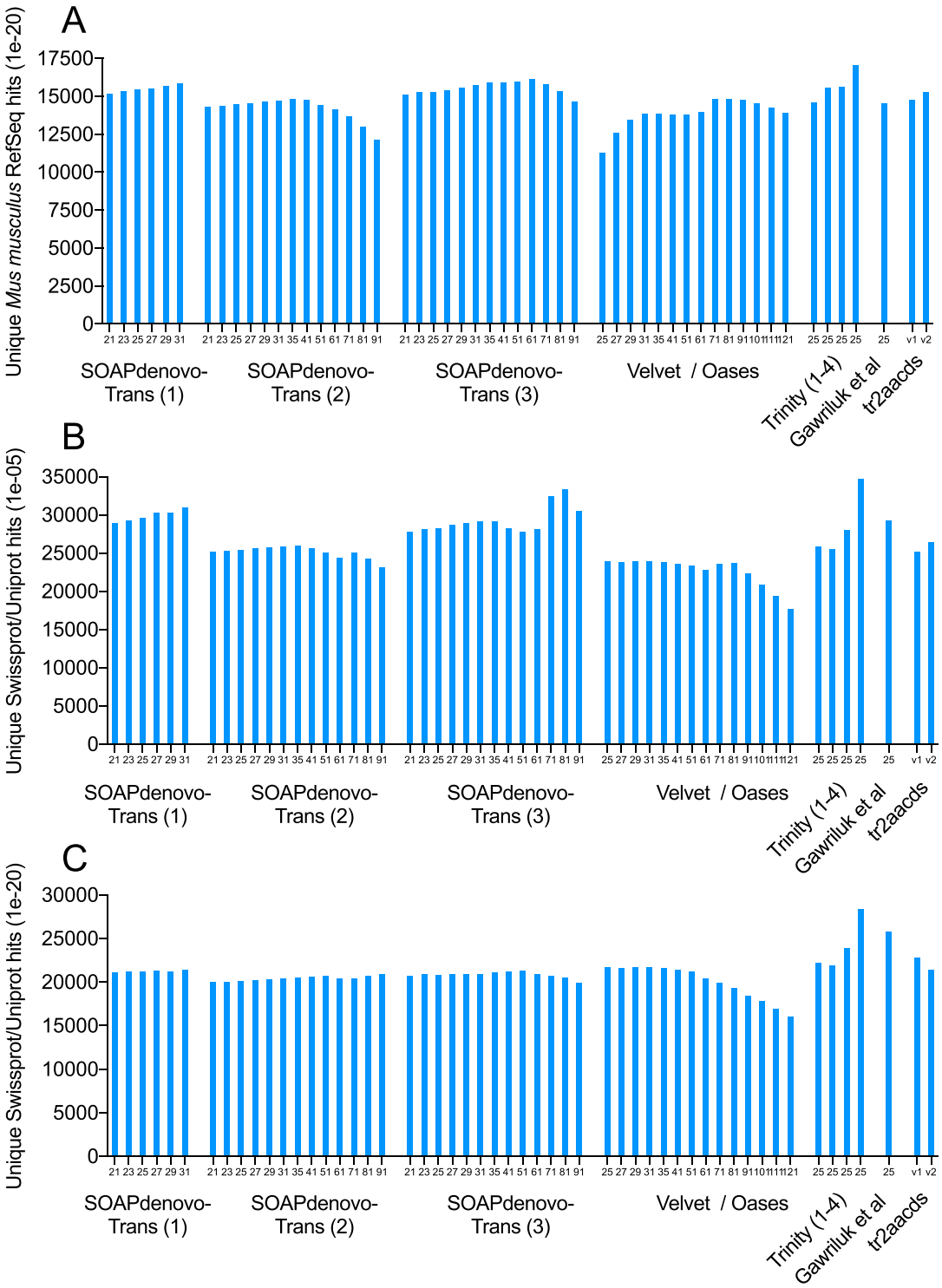



**Figure 9.** Unique BLAST hits for each assembly when aligned to: (**A**) *Mus musculus* RefSeq transcript database (e-value $\leq 1 \times 10^{-20}$), (**B**) UniProtKB/SwissProt database (e-value $\leq 1 \times 10^{-5}$), (**C**) UniProtKB/SwissProt database (e-value $\leq 1 \times 10^{-20}$). SOAPdenovo-Trans (1): no filtering; (2): light filtering of poor quality read pairs, combined with *in silico* normalization; (3): moderate filtering of poor quality read pairs; Trinity (1): light filtering of poor quality read pairs, combined with *in silico* normalization; (2): moderate filtering of poor quality read pairs; (3): light filtering of poor quality read pairs, combined with SEECER error correction; (4) Trinity v2.3.2 after moderate filtering of poor quality read pairs; tr2aacds merged assemblies (v1): pre-v2.3.2 Trinity; (v2) including v2.3.2 Trinity output.

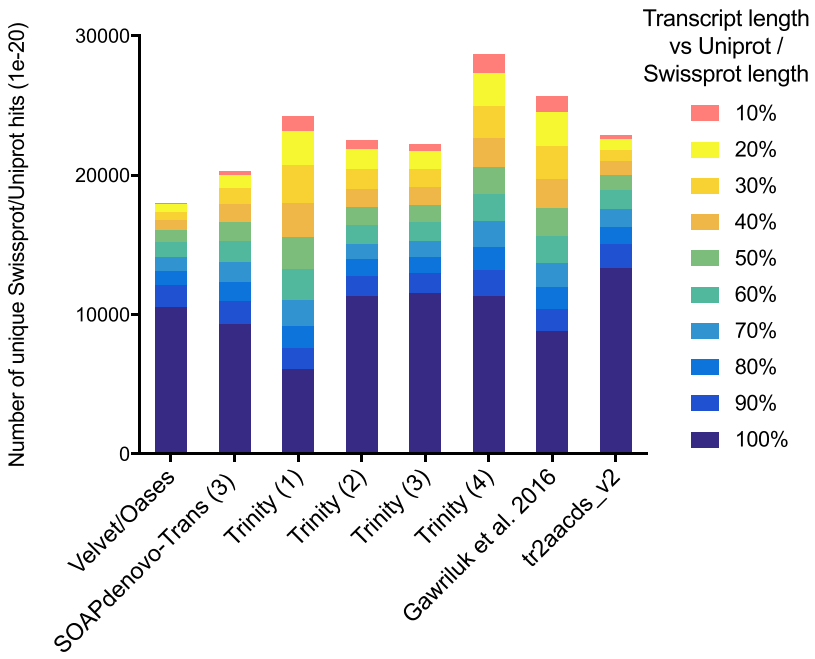



**Figure 10.** Assessing the number of full-length protein-coding transcripts within transcriptome assemblies, with UniProtKB/SwissProt BLAST hits (e-value $\leq 1 \times 10^{-20}$) categorised by length of alignment (nucleotides) with the UniProtKB/SwissProt reference transcript. Transcripts are considered 'full-length' if they are >90% of the reference transcript length. SOAPdenovo-Trans (3): moderate filtering of poor quality read pairs; Trinity (1): light filtering of poor quality read pairs, combined with *in silico* normalization; (2): moderate filtering of poor quality read pairs; (3): light filtering of poor quality read pairs, combined with SEECER error correction; (4) Trinity v2.3.2 after moderate filtering of poor quality read pairs; tr2aacds merged assemblies (v1): pre-v2.3.2 Trinity; (v2) including v2.3.2 Trinity output.

without digital normalization and error correction as described in Fig. 1. Reads were subjected to digital normalization using the "--normalize_reads" feature in Trinity.

SOAPdenovo-Trans parameters: "max_rd_len = 150, rd_len_cutof = 150, avg_ins = 192, reverse_seq = 0, asm_flags = 3" with kmer lengths: 21, 23, 25, 27, 29, 31, 35, 41, 51, 61, 71, 81, 91. Trinity was used at kmer length 25, with parameters: "--normalize_reads --seqType fq --JM 100 G --CPU 20 --min_kmer_cov 2". Reads were assembled with Velvet at kmer lengths 21, 23, 25, 27, 29, 31, 35, 41, 51, 61, 71, 81, 91, 101, 111, 121. Velvet was compiled with parameters "MAXKMERLENGTH = 141 BIGASSEMBLY = 1 LONGSEQUENCES = 1 OPENMP = 1". Velveth was run using "25,33,2 -shortPaired -fastq -separate", "35 -shortPaired -fastq –separate" and "41,131,10 -shortPaired -fastq -separate". Insert lengths of the fragments were estimated with CollectInsertSizeMetrics in Picard Tools version 1.90 (http://broadinstitute.github.io/picard/). Velvetg was run with parameters "-read_trkg yes -ins_length 215". Oases was run with parameters "-min_trans_lgth 100 -ins_length 215".

Assembly statistics were computed using the TrinityStats.pl from the Trinity package, and summary statistics are provided in log files produced by SOAPdenovo-Trans and Oases (Supplementary dataset 1).

**Collating non-redundant transcripts from multiple assemblies.** The *tr2aacds* pipeline from the EvidentialGene package was used to identify and collate non-redundant transcripts from each individual transcriptome assembly. The *tr2aacds* pipeline predicts amino acid sequences and transcript coding sequences, removes transcript redundancy based on coding potential, removes sequence fragments, clusters highly similar sequences together into loci, and classifies non-redundant transcripts as 'primary' or 'alternative'. Transcripts that scored poorly were removed, with remaining 'primary' and 'alternative' transcripts from each single-kmer assembly merged. This process was conducted twice: first ("tr2aacds_v1") with all SOAPdenovo-Trans, Velvet/Oases, and Trinity r20140413p1 assemblies ("Trinity 1–3"), and again to incorporate the Trinity v2.3.2 assembly to create "tr2aacds_v2".

Accuracy and completeness was assessed in all assemblies (single-kmer and tr2aacds) using BUSCO v1.1b1[13] to establish the presence or absence of universal single copy orthologs common to vertebrates and eukaryotes. Accuracy was assessed by the proportion of original sequence reads mapped ('backmapping') to each assembly using Bowtie v0.12.9[51] with settings: '-q --phred33-quals -n 2 -e 99999999 -l 25 -I 1-X 1000 -p 12 -a -m 200 --chunkmbs 256'. Independent RNA-Seq reads were obtained from the NCBI sequence read archive (SRA): datasets SRR636836, SRR636837, and SRR636838 were obtained from project PRJNA184055, and datasets SRR2146799–SRR2146807 from project PRJNA292021. These reads were generated from liver[10] and skin[11] and neither tissue was subjected to treatment - they are 'control' groups in their corresponding experiments. The independent RNA-Seq reads were aligned using Bowtie to each draft transcriptome assembly, with settings as specified above. The proportion of mapped reads was calculated using samtools flagstat with default parameters[52].

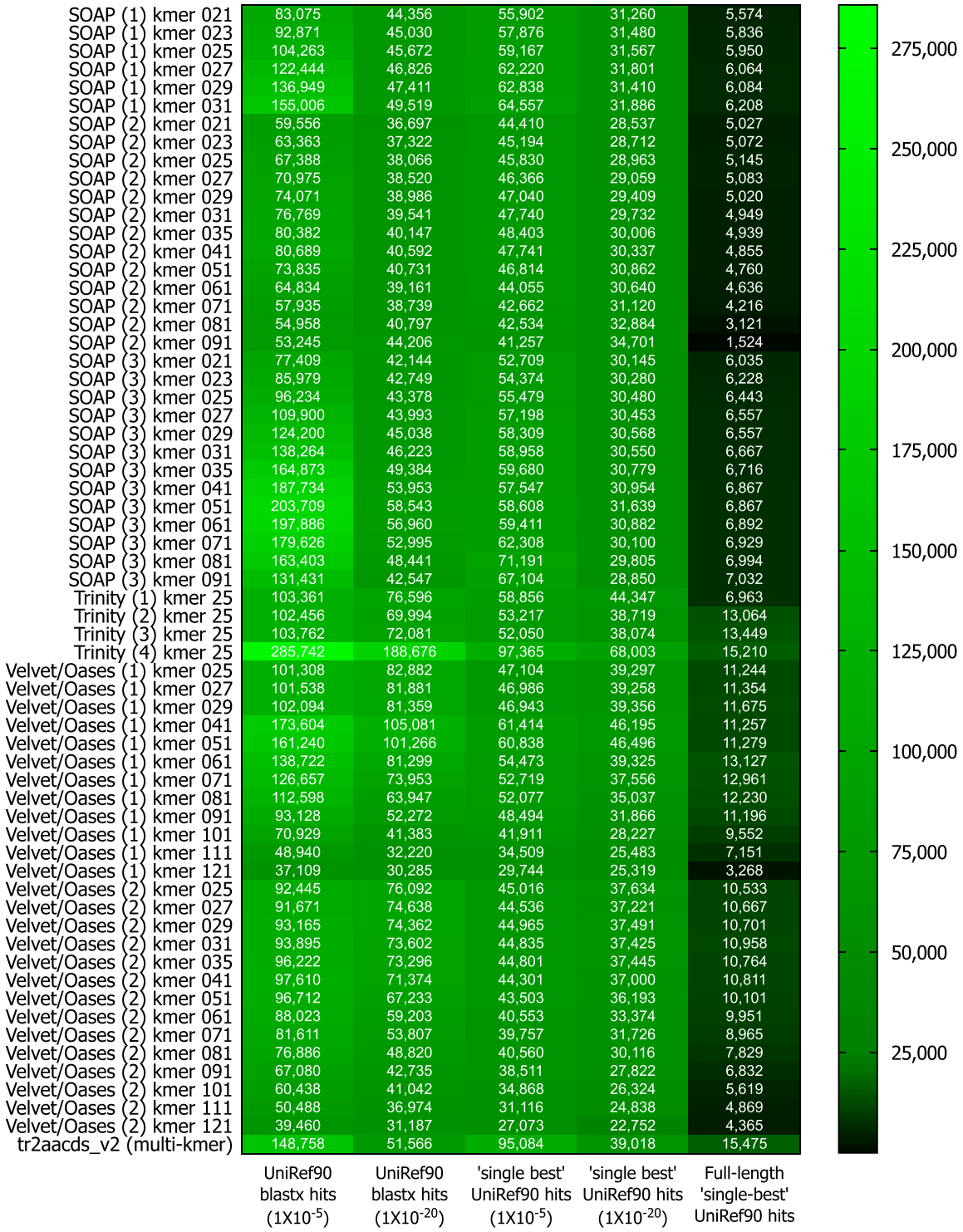



**Figure 11.** Diamond BLASTx hits from the UniRef90 database. SOAPdenovo-Trans (1): no filtering; (2): light filtering of poor quality read pairs, combined with *in silico* normalization; (3): moderate filtering of poor quality read pairs; Trinity (1): light filtering of poor quality read pairs, combined with *in silico* normalization; (2): moderate filtering of poor quality read pairs; (3): light filtering of poor quality read pairs, combined with SEECER error correction; (4) Trinity v2.3.2 after moderate filtering of poor quality read pairs; Velvet/Oases (1): moderate filtering of poor quality read pairs; Velvet/Oases (2): moderate filtering of poor quality read pairs combined with *in silico* normalization.

Structural integrity was examined using TransRate v1.0.3[14] with default settings, in which Salmon v0.6.0[53,54] and SNAP-aligner v1.0 18beta[55] were implemented. Redundancy in assembled transcripts was assessed by the proportion of highly similar contiguous sequences (contigs), clustered using CD-HIT-EST v4.6.5[16,17] with settings '-c 0.95 -n 8 -p 1 -g 1 -M 200000 -T 8 -d 40'. Further clustering at 90%, 95%, and 100% similarity was conducted on a representative single-kmer assembly "Trinity (2)" to assess contig redundancy.

**Annotation and identification of non-coding RNAs.** The best performing assembly was annotated using the Trinotate pipeline (ver2.0.2, http://trinotate.github.io/). In brief, *de novo* transcripts were aligned against the UniProtKB/SwissProt database (ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/complete/uniprot_sprot.fasta.gz; accessed 7th January 2016) using NCBI BLAST + BLASTx (for nucleotide sequences) and BLASTp (for protein sequences)[56]. Transdecoder v2.0.1 (https://transdecoder.github.io/) was used to predict ORFs, with BLASTp performed using translations of predicted ORFs as the query and UniProtKB/SwissProt database as the target. HMMER v3.1b1 and Pfam v27 databases[57] were used to predict protein domains. SignalP v4.1[58] was used to predict signal peptides, and RNAmmer v1.2[59] to predict rRNAs. Annotations were loaded into an SQL database (packaged with Trinity: Supplementary dataset 2). Gene Ontology (GO) terms linked to the UniProtKB/SwissProt entry for each BLAST hit were used for ontology annotation. GO functional classifications were summarised using the Web Gene Ontology (WEGO) annotation plot (http://wego.genomics.org.cn)[60]. *De novo* transcripts were also aligned to *Mus musculus* RefSeq RNA transcripts (ftp://ftp.ncbi.nlm.nih.gov/refseq/M_musculus/mRNA_Prot/mouse.1.rna.fna.gz; accessed 3rd Feb, 2017) using NCBI BLAST + BLASTn with settings '-num_threads 32 -max_target_seqs. 1 -evalue 1e-20 -outfmt 6'.

Alignment of transcripts from each assembly to the UniRef90 database was conducted using DIAMOND v0.8.36[61]. This program is significantly faster than NCBI BLAST + for aligning nucleotide sequences to a protein database (up to 20,000X increase in speed). 'DIAMOND BLASTx' was used instead of 'BLAST+ BLASTx' to offset the larger size of the database: UniRef90 = ~53 million sequences; UniProtKB/SwissProt database = ~550 thousand sequences. Default settings were used, with the addition of parameters '--sensitive -p 40 -k 1 -e 1e-05 -b 40 -c 1'.

Non-coding RNA analysis was conducted using the Coding-Non-Coding Index (CNCI) signature identification tool (version 2, Feb 28th 2014, https://github.com/www-bioinfo-org/CNCI; 0fa252b) profiling adjoining nucleotide triplets and classifying transcripts as protein-coding or non-coding, independent of known annotations[62]. Non-coding transcripts were identified using the "vertebrate species model", and a threshold cutoff of −0.05[63–65].

Figures were produced using R software v3.3.2 and GraphPad Prism 7.

## References


1. Wilson, D. E. & Reeder, D. M. *Mammal Species of the World: A Taxonomic and Geographic Reference* (JHU Press, 2005).
2. Brunjes, P. C. The precocial mouse, *Acomys cahirinus*. *Psychobiology* **18**, 339–350 (1990).
3. Dickinson, H., Walker, D. W., Cullen-McEwen, L., Wintour, E. M. & Moritz, K. The spiny mouse (*Acomys cahirinus*) completes nephrogenesis before birth. *Am. J. Physiol. Renal Physiol.* **289**, F273–9 (2005).
4. Lamers, W. H. *et al.* Hormones in perinatal rat and spiny mouse: relation to altricial and precocial timing of birth. *Am. J. Physiol.* **251**, E78–85 (1986).
5. Quinn, T. A. *et al.* Ontogeny of the adrenal gland in the spiny mouse, with particular reference to production of the steroids cortisol and dehydroepiandrosterone. *Endocrinology* **154**, 1190–1201 (2013).
6. Quinn, T. A., Ratnayake, U., Dickinson, H., Castillo-Melendez, M. & Walker, D. W. Ontogenetic Change in the Regional Distribution of Dehydroepiandrosterone-Synthesizing Enzyme and the Glucocorticoid Receptor in the Brain of the Spiny Mouse (*Acomys cahirinus*). *Dev. Neurosci.* **38**, 54–73 (2016).
7. Bellofiore, N. *et al.* First evidence of a menstruating rodent: the spiny mouse (*Acomys cahirinus*). *Am. J. Obstet. Gynecol.* **216**, 40.e1–40.e11 (2017).
8. Emera, D., Romero, R. & Wagner, G. The evolution of menstruation: a new model for genetic assimilation: explaining molecular origins of maternal responses to fetal invasiveness. *Bioessays* **34**, 26–35 (2012).
9. Hadid, Y. *et al.* Sympatric incipient speciation of spiny mice Acomys at 'Evolution Canyon,' Israel. *Proc. Natl. Acad. Sci. USA* **111**, 1043–1048 (2014).
10. Fushan, A. A. *et al.* Gene expression defines natural changes in mammalian lifespan. *Aging Cell* **14**, 352–365 (2015).
11. Gawriluk, T. R. *et al.* Comparative analysis of ear-hole closure identifies epimorphic regeneration as a discrete trait in mammals. *Nat. Commun.* **7**, 11164 (2016).
12. Garber, M., Grabherr, M. G., Guttman, M. & Trapnell, C. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat. Methods* **8**, 469–477 (2011).
13. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
14. Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M. & Kelly, S. TransRate: reference-free quality assessment of *de novo* transcriptome assemblies. *Genome Res.* **26**, 1134–1144 (2016).
15. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
16. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
17. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
18. Le, H.-S., Schulz, M. H., McCauley, B. M., Hinman, V. F. & Bar-Joseph, Z. Probabilistic error correction for RNA sequencing. *Nucleic Acids Res.* **41**, e109 (2013).
19. The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45, D158–D169 (2017).
20. Boutet, E. *et al.* UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View. *Methods Mol. Biol.* **1374**, 23–54 (2016).
21. Suzek, B. E. *et al.* UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**, 926–932 (2015).

22. Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R. & Wu, C. H. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **23**, 1282–1288 (2007).
23. Gilbert, D. EvidentialGene: Evidence Directed Gene predictions for Eukaryotes (2010).
24. Gilbert, D. "Perfect(ing) Arthropod Genes with Next Gen Informatics. *4th Arthropod Genomics Symposium* (2010).
25. Luo, H. *et al*. Identification and function annotation of long intervening noncoding RNAs. *Brief. Bioinform*., doi:10.1093/bib/bbw046 (2016).
26. Zhao, Y. *et al*. NONCODE 2016: an informative and valuable data source of long non-coding RNAs. *Nucleic Acids Res.* **44**, D203–8 (2016).
27. Xie, Y. *et al*. SOAPdenovo-Trans: *de novo* transcriptome assembly with short RNA-Seq reads. *Bioinformatics* **30**, 1660–1666 (2014).
28. Schulz, M. H., Zerbino, D. R., Vingron, M. & Birney, E. Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**, 1086–1092 (2012).
29. Khudyakov, J. I., Champagne, C. D., Meneghetti, L. M. & Crocker, D. E. Blubber transcriptome response to acute stress axis activation involves transient changes in adipogenesis and lipolysis in a fasting-adapted marine mammal. *Sci. Rep.* **7**, 42110 (2017).
30. Blande, D., Halimaa, P., Tervahauta, A. I., Aarts, M. G. M. & Kärenlampi, S. O. *de novo* transcriptome assemblies of four accessions of the metal hyperaccumulator plant Noccaea caerulescens. *Sci Data* **4**, 160131 (2017).
31. MacManes, M. D. Establishing evidenced-based best practice for the *de novo* assembly and evaluation of transcriptomes from non-model organisms., doi:10.1101/035642 (2015).
32. Priyam, A. *et al*. Sequenceserver: a modern graphical user interface for custom BLAST databases, doi:10.1101/033142 (2015).
33. Ireland, Z., Russell, A. P., Wallimann, T., Walker, D. W. & Snow, R. Developmental changes in the expression of creatine synthesizing enzymes and creatine transporter in a precocial rodent, the spiny mouse. *BMC Dev. Biol.* **9**, 39 (2009).
34. O'Connell, B. A., Moritz, K. M., Roberts, C. T., Walker, D. W. & Dickinson, H. The placental response to excess maternal glucocorticoid exposure differs between the male and female conceptus in spiny mice. *Biol. Reprod.* **85**, 1040–1047 (2011).
35. Ellery, S. J. *et al*. Creatine pretreatment prevents birth asphyxia-induced injury of the newborn spiny mouse kidney. *Pediatr. Res.* **73**, 201–208 (2013).
36. Robertson, L. S. & Cornman, R. S. Transcriptome resources for the frogs Lithobates clamitans and Pseudacris regilla, emphasizing antimicrobial peptides and conserved loci for phylogenetics. *Mol. Ecol. Resour.* **14**, 178–183 (2014).
37. Maudhoo, M. D., Madison, J. D. & Norgren, R. B. Jr. *de novo* assembly of the chimpanzee transcriptome from NextGen mRNA sequences. *Gigascience* **4**, 18 (2015).
38. Athanasiadis, A., Rich, A. & Maas, S. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol.* **2**, e391 (2004).
39. Du, L. *et al*. First insights into the giant panda (Ailuropoda melanoleuca) blood transcriptome: a resource for novel gene loci and immunogenetics. *Mol. Ecol. Resour.* **15**, 1001–1013 (2015).
40. Zieliński, P., Stuglik, M. T., Dudek, K., Konczal, M. & Babik, W. Development, validation and high-throughput analysis of sequence markers in nonmodel species. *Mol. Ecol. Resour.* **14**, 352–360 (2014).
41. Bens, M. *et al*. FRAMA: from RNA-seq data to annotated mRNA assemblies. *BMC Genomics* **17**, 54 (2016).
42. Francischetti, I. M. B. *et al*. The 'Vampirome': Transcriptome and proteome analysis of the principal and accessory submaxillary glands of the vampire bat Desmodus rotundus, a vector of human rabies. *J. Proteomics* **82**, 288–319 (2013).
43. Carneiro, M. *et al*. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* **345**, 1074–1079 (2014).
44. Shimoyama, M. *et al*. The Chinchilla Research Resource Database: resource for an otolaryngology disease model. *Database* **2016** (2016).
45. Haas, B. J. *et al*. *de novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
46. Robertson, G. *et al*. *de novo* assembly and analysis of RNA-seq data. *Nat. Methods* **7**, 909–912 (2010).
47. Surget-Groba, Y. & Montoya-Burgos, J. I. Optimization of *de novo* transcriptome assembly from next-generation sequencing data. *Genome Res.* **20**, 1432–1440 (2010).
48. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EM Bnet. journal* **17**, 10 (2011).
49. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
50. Zerbino, D. R. & Birney, E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).
51. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
52. Li, H. *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
53. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides accurate, fast, and bias-aware transcript expression estimates using dual-phase inference, doi:10.1101/021592 (2015).
54. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods*, doi:10.1038/nmeth.4197 (2017).
55. Zaharia, M. et al. Faster and more accurate sequence alignment with SNAP. *ArXiv preprint*, doi:arXiv:1111.5572 (2011).
56. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
57. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–37 (2011).
58. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).
59. Lagesen, K. *et al*. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108 (2007).
60. Ye, J. *et al*. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* **34**, W293–7 (2006).
61. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
62. Sun, L. *et al*. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* **41**, e166 (2013).
63. Nam, J.-W. & Bartel, D. P. Long noncoding RNAs in C. elegans. *Genome Res.* **22**, 2529–2540 (2012).
64. Zhou, Z.-Y. *et al*. Genome-wide identification of long intergenic noncoding RNA genes and their potential association with domestication in pigs. *Genome Biol. Evol.* **6**, 1387–1392 (2014).
65. Zhou, Q.-Z., Zhang, B., Yu, Q.-Y. & Zhang, Z. BmncRNAdb: a comprehensive database of non-coding RNAs in the silkworm, Bombyx mori. *BMC Bioinformatics* **17**, 370 (2016).


## Acknowledgements

The authors would like to thank the following people for their contribution to the study: Daniel Cameron, Gerry Tonkin-Hill and all members of the Papenfuss lab; Steve Androulakis, and all members of the Monash Bioinformatics Platform; Vivien Vasic and members of the MHTP Medical Genomics Facility; and Walter and

Eliza Hall Institute Department of Bioinformatics. This work was supported by the Victorian Government Operational Infrastructure Support Scheme and an Australian Research Council Discovery Grant to DWW and HD. JM is in receipt of a Faculty of Medicine, Nursing and Health Sciences Postgraduate Research Scholarship to undertake his PhD studies. HD is supported by a NHMRC Career Development Fellowship. DWW is supported by Cerebral Palsy Alliance.

## Author Contributions

H.D., P.T.S., D.P., T.S. and A.T.P. designed the project. H.D. collected the tissues for sequencing and funded the project. T.W. conducted the library preparation and sequencing. A.T.P. and J.M. designed the assembly pipeline. J.M. conducted the assembly, analysis, and validation with guidance and advice from A.T.P. and R.L. J.M. wrote the manuscript. All authors were involved in interpretation of results and all authors read and approved the final manuscript.

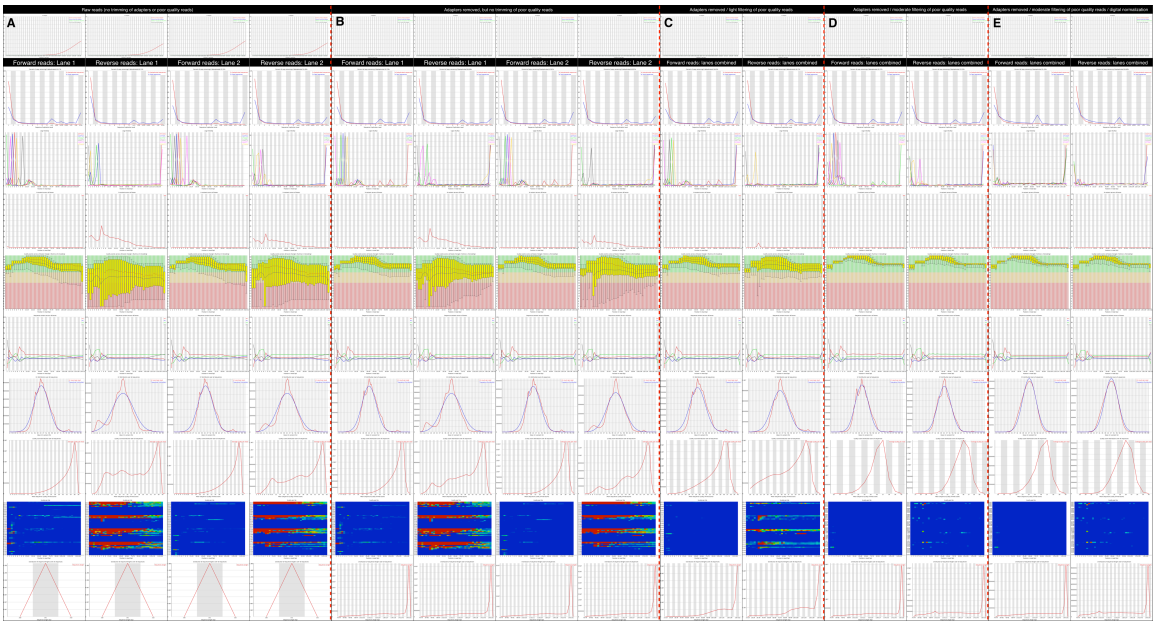
## Additional Information

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-09334-7

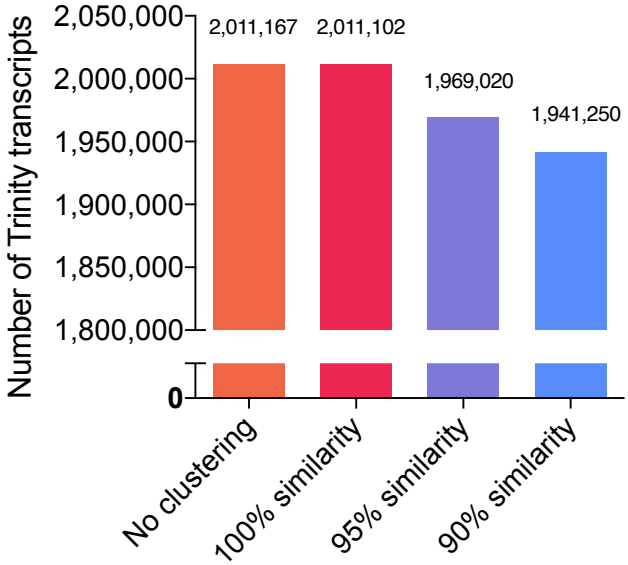
**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

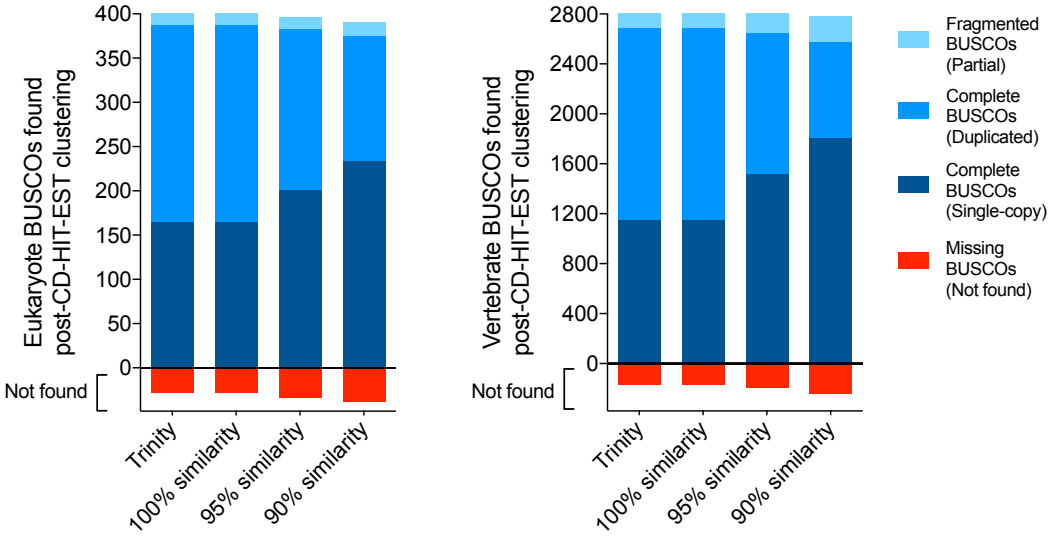
## 3.2 Supplementary figures



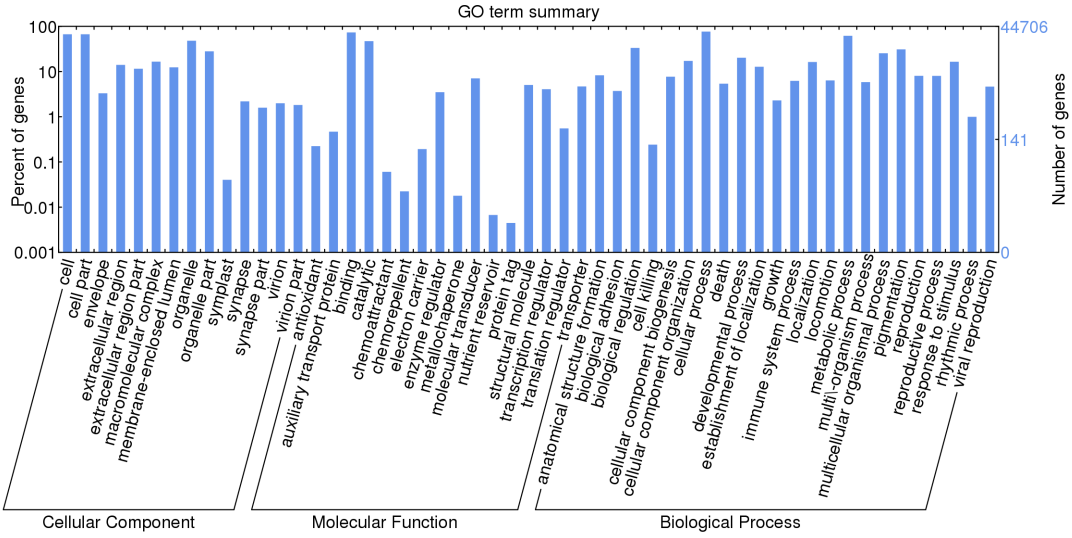

Supplementary Figure S1: FastQC reports for RNA-Seq reads: (A) unfiltered, (B) light filtering of poor quality reads, (C) moderate filtering of concatenated reads, (D) Strict quality filtering, (E) In silico normalization.



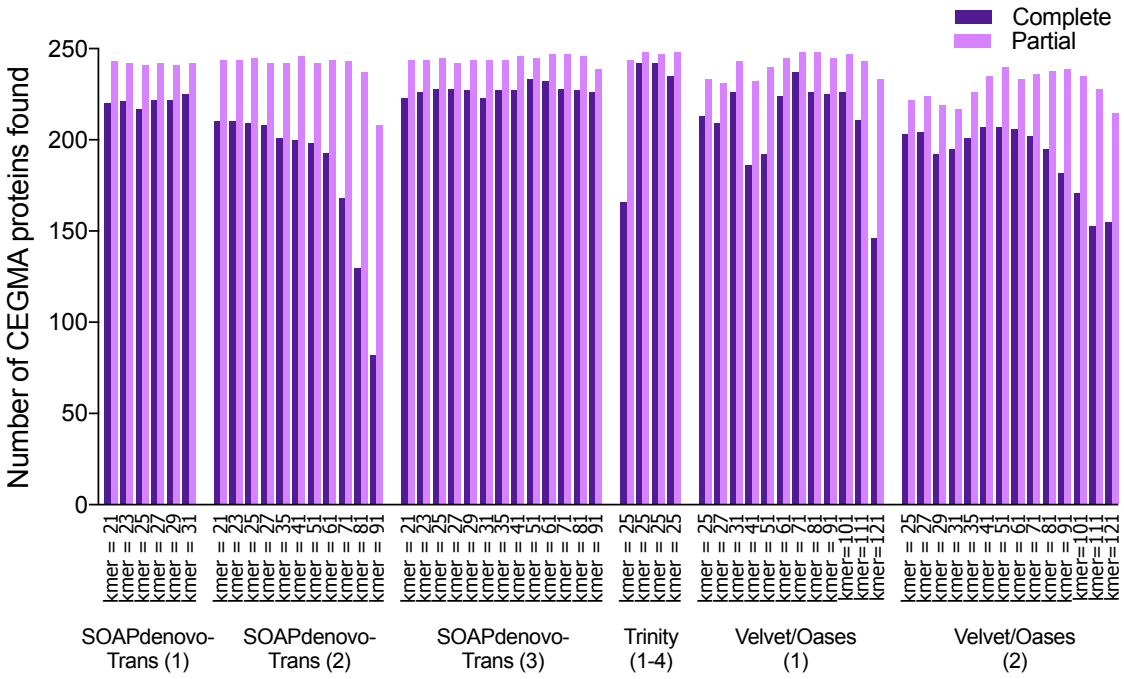

Supplementary Figure S2: Number of Trinity transcripts after CD-HIT-EST clustering at specified thresholds (100%, 95%, 90%).




Supplementary Figure S3: BUSCO analysis after CD-HIT-EST clustering at specified thresholds (100%, 95%, 90%).




Supplementary Figure S4: Summary of gene ontology (GO) terms associated with SwissProt/UniProt blast hits.




Supplementary Figure S5: Number of complete/partial CEGMA proteins found in each assembly.

## 3.3 Detailed protocol

This manuscript describes the protocol used for assembly and annotation of a transcriptome for the spiny mouse. It has been published by protocols.io; it has not been peer reviewed. Original source for this work:

Mamrot, J. (2017). De novo transcriptome assembly workflow. DOI: https://dx.doi.org/10.17504/protocols.io.ghebt3e

# De novo transcriptome assembly workflow

**Jared Mamrot, Roxane Legaie, Stacey J Ellery, Trevor Wilson, Torsten Seemann, David R Powell, David Gardner, David W Walker, Peter Temple-Smith, Anthony T Papenfuss, Hayley Dickinson**

**Abstract**

This protocol describes the production of a reference-quality *de novo* transcriptome assembly for the spiny mouse (Acomys cahirinus). These methods can be applied to other RNA-Seq datasets to generate high-quality transcriptome assemblies in other species. Validation and description of assembly output is described in 'De novo transcriptome assembly for the spiny mouse (*Acomys cahirinus*)' *bioRxiv*, p.076067.

**Authored by:** Jared Mamrot

**Citation:** Jared Mamrot, Roxane Legaie, Stacey J Ellery, Trevor Wilson, Torsten Seemann, David R Powell, David Gardner, David W Walker, Peter Temple-Smith, Anthony T Papenfuss, Hayley Dickinson. De novo transcriptome assembly workflow. protocols.io, https://dx.doi.org/10.17504/protocols.io.ghebt3e

**Published:** 06 Mar 2017

## Protocol

### **Step 1:** Import and organise raw data

Download raw data from the NCBI to working directory and archive a copy (read-only). To efficiently transfer data the NCBI recommends using Aspera connect, a FASP® transfer program which facilitates high-speed data transfer.

Many commands in this protocol take hours/days to complete: to avoid processes being killed if connection to the server is lost, employ the 'nohup' command and/or run processes in the background ('&') and disown them from the terminal ('disown %1'). Where possible, follow good scientific practices eg. Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L. and Teal, T.K., 2016. Good Enough Practices in Scientific Computing. *arXiv preprint arXiv:1609.00037*.

Aspera connect:
Download - http://downloads.asperasoft.com/en/downloads/8?list
(ver3.6.2) Documentation -
https://www.ncbi.nlm.nih.gov/books/NBK242625/ Requirements - NCBI
SRA toolkit

NCBI SRA toolkit:
Download - https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software (ver2.8.1-3)
Documentation - https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=toolkit_doc

DATASET

**Illumina HiSeq 1500**
**NCBI BioProject Accession:**
**PRJNNA342864**

```
#Create working directory and directory for installed software
mkdir $HOME/projects $HOME/projects/spiny_mouse
export WORKDIR=$HOME/projects/spiny_mouse/
cd $WORKDIR && mkdir user_installed_software
export PROGRAMDIR=$WORKDIR/user_installed_software

#Download, unpack, and install aspera connect
cd $PROGRAMDIR
wget http://download.asperasoft.com/download/sw/connect/3.6.2/aspera-
connect-3.6.2.117442-linux-64.tar.gz -O aspera.tar.gz
tar zxvf aspera.tar.gz && rm aspera.tar.gz
bash aspera-connect*
cd ~/.aspera/connect/bin
#add binaries to a directory contained in PATH, or add current directory to
PATH
echo export PATH=\$PATH:`pwd`\ >> ~/.bashrc && source ~/.bashrc

#Download reads from the NCBI
cd $WORKDIR
ascp -i ~/.aspera/connect/etc/asperaweb_id_dsa.openssh -T anonftp@ftp-
trace.ncbi.nlm.nih.gov:/sra/sra-
instant/reads/ByRun/sra/SRR/SRR427/SRR4279903 anonftp@ftp-
trace.ncbi.nlm.nih.gov:/sra/sra-
instant/reads/ByRun/sra/SRR/SRR427/SRR4279904 .

#Obtain reads in fastq format using the ncbi SRA Toolkit
find . -name "*.sra" -exec fastq-dump --split-spot --split-files --skip-
technical -I -F -Q 33 -W -T -R pass '{}' \;
cd SRR4279903/pass/1 && mv fastq Lane1_R1.fastq
cd ../2 && mv fastq Lane1_R2.fastq
cd ../../../SRR4279904/pass/1 && mv fastq Lane2_R1.fastq
cd ../2 && mv fastq Lane2_R2.fastq

#Move fastq files to WORKDIR
cd $WORKDIR
find . -name "Lane*" -exec mv '{}' $WORKDIR \; && rm -R SRR427990*

#Delete sra files
cd $WORKDIR/
find . -name "*.sra" -delete

#Archive a read-only copy of the raw data
cd $WORKDIR/
mkdir protected_data && cd protected_data
cp ../Lane* .
chmod 444 Lane*
```

## **Step 2:** Decompress and assess RNA-Seq read quality

Decompress gzipped files (*.gz), and use FastQC for preliminary read quality assessment. GNU zip (gzip) is a popular compression utility free from patented algorithms. FastQC is a quality control tool for high throughput sequence data which assesses multiple metrics and provides a QC report.

gzip:
Download - https://www.gnu.org/software/gzip/ (ver1.2.4)
Documentation - http://www.math.utah.edu/docs/info/gzip_toc.html

fastQC:
Download - http://www.bioinformatics.babraham.ac.uk/projects/download.html#fastqc (ver0.11.15)
Documentation - http://www.bioinformatics.babraham.ac.uk/projects/fastqc/Help/
Requirements - A suitable Java Runtime Environment (https://www.java.com/en/) and the Picard BAM/SAM Libraries (included in download)

COMMANDS (UNIX - BASH)

```
#Download and install GZIP and FastQC
cd $PROGRAMDIR
wget http://www.gzip.org/gz124src.zip
wget
http://www.bioinformatics.babraham.ac.uk/projects/fastqc/fastqc_v0.11.5
.zip
#unpack
unzip *.zip
#install
cd gzip124src
./configure --prefix=`pwd`
make
make install
#add binaries to a directory contained in PATH, or add current
directory to PATH
cd bin && echo export PATH=\$PATH:`pwd`\ >> ~/.bashrc && source
~/.bashrc

#Set file permissions for FastQC
cd $PROGRAMDIR/FastQC
chmod 755 fastqc
#add binaries to a directory contained in PATH, or add current
directory to PATH
echo export PATH=\$PATH:`pwd`\ >> ~/.bashrc && source ~/.bashrc

#If required, decompress fastq files downloaded from the NCBI (may not
be necessary depending on SRAtoolkit version)
cd $WORKDIR
gunzip *.fastq.gz

#Run FastQC on each sample to assess read quality
for f in Lane*; do fastqc $f; done
```

## **Step 3:** Trim adapters and re-examine read quality

Trim_galore is a tool that implements cutadapt to consistently apply quality and adapter trimming to FastQ files: it seeks out and removes adapter sequences from RNA-Seq data.

Trim_galore:
Download - https://github.com/FelixKrueger/TrimGalore/releases (ver0.4.2) Documentation - http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/Trim_Galore_User_Guide_v0.4.2.pdf
Requirements: Cutadapt and FastQC

Cutadapt:
Download - https://pypi.python.org/pypi/cutadapt (ver1.9.1)
Documentation - https://media.readthedocs.org/pdf/cutadapt/stable/cutadapt.pdf
Reference - Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal, 17(1), pp.pp-10.

*Minimum read length parameter ('--length 80') is dependant on length of original reads (in this case 150bp, paired-end)

COMMANDS (UNIX - BASH)

```
#Download and install cutadapt using pip
cd $PROGRAMDIR
pip install --user --upgrade cutadapt

#Download and install cutadapt using anaconda
(https://www.continuum.io/downloads)
cd $PROGRAMDIR
conda install -c bioconda cutadapt && python setup.py install --user

#Download trim_galore
cd $PROGRAMDIR
wget https://github.com/FelixKrueger/TrimGalore/archive/0.4.2.tar.gz -O
trim_galore.tar.gz
#unpack
tar zxvf trim_galore.tar.gz && cd TrimGalore-0.4.2
#add binary to a directory contained in PATH, or add current directory to
PATH
echo export PATH=\$PATH:`pwd`\ >> ~/.bashrc && source ~/.bashrc

#Run trim_galore on each lane
mkdir $WORKDIR/adapter_trimmed_reads && cd $WORKDIR
trim_galore --paired --retain_unpaired --length 80 --output
./adapter_trimmed_reads Lane1_R1.fastq Lane1_R2.fastq
1>trim_galore_lane_1_reads.log 2>trim_galore_lane_1_reads.err &

trim_galore --paired --retain_unpaired --length 80 --output
./adapter_trimmed_reads Lane2_R1.fastq Lane2_R2.fastq
1>trim_galore_lane_2_reads.log 2>trim_galore_lane_2_reads.err &

#Stdout ('1>') and stderr ('2>') are redirected to .log/.err files to aid
in 'disowning' processes from the terminal session ('disown %1')
```

## **Step 4:** Remove poor quality reads

Remove poor quality reads and trim poor quality nucleotides from read ends using trimmomatic. Trimmomatic is a flexible read trimming tool for Illumina NGS data. Re-assess 'cleaned' reads using FastQC to check read metrics have been improved.

Using a reletively high quality threshold improves assembler performance and reduces memory requirements, however lowly expressed transcripts may be lost as proportionately more reads are excluded. Conversely, a relatively low quality threshold may improve retention of lowly expressed transcripts, but negatively affect assembler performance (https://doi.org/10.3389/fgene.2014.00013). The optimal degree of read trimming and removal can differ for each dataset.

Several read quality thresholds were used in the publication associated with this protocol (Mamrot, J., Legaie, R., Ellery, S.J., Wilson, T., Gardner, D., Walker, D.W., Temple-Smith, P., Papenfuss, A.T. and Dickinson, H., 2016. De novo transcriptome assembly for the spiny mouse (Acomys cahirinus). *bioRxiv*, p.076067). Each dataset was assembled (Steps 5-9), validated and compared for accuracy and completeness. For this dataset a relatively low quality threshold (as outlined in the COMMANDS section below) produced higher quality assemblies.

In addition to differing levels of trimming/filtering, read error correction was conducted using SEECER (http://sb.cs.cmu.edu/seecer/). Assembly metrics were slightly improved when error-corrected reads were assembled with Trinity (this 'error-corrected' assembly was included in downstream analyses), however there were no noticeable improvements assembling error-corrected reads with SOAPdenovo-Trans and Velvet/Oases with this dataset. Despite this unexpected outcome, error correction is still recommended for best-practice transcriptome assembly (http://oyster-river-protocol.readthedocs.io/en/master/; http://dx.doi.org/10.1101/035642; https://doi.org/10.7717/peerj.113)

Trimmomatic:
Download - http://www.usadellab.org/cms/?page=trimmomatic (ver0.36) Documentation - http://www.usadellab.org/cms/uploads/supplementary/Trimmomatic/TrimmomaticManual_V0.32.pdf
Reference - Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics, btu170.

SEECER:
Download - http://sb.cs.cmu.edu/seecer/install.html#Download (ver1.3)
Documentation - http://sb.cs.cmu.edu/seecer/downloads/manual.pdf
Reference - Hai-Son Le, Marcel H. Schulz, Brenna M. McCauley, Veronica F. Hinman and Ziv Bar-Joseph (2013). Probabilistic error correction for RNA sequencing. Nucleic Acids Research /
Requirements - GNU Scientific Library, SeqAn, Jellyfish, OPENMP API

COMMANDS (UNIX - BASH)

```
#There is substantial overlap in the capabilities of trim_galore and
trimmomatic; each program was used here for different purposes: trim galore
to seek out and remove adapters, trimmomatic for trimming poor quality
bases/reads.

#Download and install trimmomatic using linuxbrew (http://linuxbrew.sh/)
cd $PROGRAMDIR
brew tap homebrew/science
brew install trimmomatic
#alternatively, download trimmomatic from source
cd $PROGRAMDIR
```

```
wget
http://www.usadellab.org/cms/uploads/supplementary/Trimmomatic/Trimmomatic
-0.36.zip
unzip Trimmomatic-0.36.zip && rm Trimmomatic-0.36.zip && cd Trimmomatic-
0.36/
#add binary to a directory contained in PATH, or add current directory to
PATH
echo export PATH=\$PATH:`pwd`\ >> ~/.bashrc && source ~/.bashrc

cd $WORKDIR/adapter_trimmed_reads
trimmomatic PE -phred33 Lane1_R1_val_1.fq Lane1_R2_val_2.fq
lane1_R1_pairedout.fq lane1_R1_unpairedout.fq lane1_R2_pairedout.fq
lane1_R2_unpairedout.fq LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 AVGQUAL:20
MINLEN:36 1>trimmomatic_1.log 2>trimmomatic_1.err &
trimmomatic PE -phred33 Lane2_R1_val_1.fq Lane2_R2_val_2.fq
lane2_R1_pairedout.fq lane2_R1_unpairedout.fq lane2_R2_pairedout.fq
lane2_R2_unpairedout.fq LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 AVGQUAL:20
MINLEN:36 1>trimmomatic_2.log 2>trimmomatic_2.err &

mkdir $WORKDIR/cleaned_trimmed_reads
mv *_pairedout.fq $WORKDIR/cleaned_trimmed_reads/

#Assess read quality using FastQC, and compare with untrimmed/unfiltered
reads and concatenate lanes
cd $WORKDIR/cleaned_trimmed_reads
for f in *pairedout.fq; do fastqc $f; done
cat lane1_R1_pairedout.fq lane2_R1_pairedout.fq > R1_pairedout.fastq
cat lane1_R2_pairedout.fq lane2_R2_pairedout.fq > R2_pairedout.fastq
#optionally, unpairedout.fq reads can be added to R2_pairedout.fq prior to
assembly

#Download and install SEECER
cd $PROGRAMDIR
wget http://sb.cs.cmu.edu/seecer/downloads/SEECER-0.1.3.tar.gz
tar zxvf SEECER-0.1.3.tar.gz && rm SEECER-0.1.3.tar.gz
cd SEECER-0.1.3
./configure --prefix=`pwd`
make
make install
#add binary to a directory contained in PATH, or add current dir to PATH
cd bin && echo export PATH=\$PATH:`pwd`\ >> ~/.bashrc && source ~/.bashrc

#Conduct error correction on trimmed/filtered reads using SEECER;
'run_seecer.sh' on testdata to check install, then symlink real data into
./testdata directory and run the 'run_seecer.sh' script
cd $PROGRAMDIR/SEECER-0.1.3/testdata.
ln -s $WORKDIR/cleaned_trimmed_reads/R1_pairedout.fastq R1_pairedout.fastq
ln -s $WORKDIR/cleaned_trimmed_reads/R2_pairedout.fastq R2_pairedout.fastq
cd $PROGRAMDIR/SEECER-0.1.3/
bash ./bin/run_seecer.sh -t tmpDirectory ./testdata/R1_pairedout.fastq
./testdata/R2_pairedout.fastq 1>seecer.log 2>seecer.err &
```

## **Step 5:** Prepare for SOAPdenovo-Trans assembly

SOAPdenovo-Trans is a de Bruijn graph-based assembler for transcriptome data, derived from the SOAPdenovo2 genome assembler. It incorporates the innovative error-removal model from Trinity and combines this with the robust heuristic graph traversal method for solving isoform-related sub-graphs from Oases (Trinity and Oases are also employed in this protocol). Further details on the algorithms and applications of SOAPdenovo-Trans are available at https://doi.org/10.1093/bioinformatics/btu077

SOAPdenovo-Trans requires user-specified parameters to be listed in a 'config' file.

SOAPdenovo-Trans:
Download - http://soap.genomics.org.cn/SOAPdenovo-Trans.html (ver1.03)
Documentation - http://soap.genomics.org.cn/SOAPdenovo-Trans.html / https://github.com/aquaskyline/SOAPdenovo-Trans
Reference - Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S. and Zhou, X., 2014. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics, 30(12), pp.1660-1666.

COMMANDS (UNIX - BASH)

```
#Download
cd $PROGRAMDIR
mkdir SOAPdenovo-Trans
wget https://downloads.sourceforge.net/project/soapdenovotrans/SOAPdenovo-Trans/bin/v1.03/SOAPdenovo-Trans-bin-v1.03.tar.gz -O SOAPdenovo-Trans/SOAPdenovo-Trans-bin-v1.03.tar.gz
cd SOAPdenovo-Trans/
tar zxvf SOAPdenovo-Trans-bin-v1.03.tar.gz
#add binary to a directory contained in PATH, or add current directory to PATH
echo export PATH=\$PATH:`pwd`\ >> ~/.bashrc && source ~/.bashrc

#Create config file- options are described/explained in the SOAPdenovo-Trans documentation
cd $WORKDIR/cleaned_trimmed_reads
cat > SOAP.config <<END_TEXT
max_rd_len=150
[LIB]
avg_ins=192
reverse_seq=0
asm_flags=3
q1=R1_pairedout.fastq
q2=R2_pairedout.fastq
END_TEXT
```

## **Step 6:** Assemble reads using SOAPdenovo-Trans

Assemble reads at multiple kmer sizes using SOAPdenovo-Trans.

Use SOAPdenovo-Trans31mer for kmer sizes 21 - 31.
Use SOAPdenovo-Trans127mer for kmer sizes 41 - 121.

After assembly is complete, statistics are contained in the *.scafStat file, and scaffolded contigs in the *.scafSeq file.

COMMANDS (UNIX - BASH)

```
cd $WORKDIR/cleaned_trimmed_reads
mkdir SOAPdenovo-Trans_output
for ((n=21; n<33; n=n+2)); do SOAPdenovo-Trans-31mer all -K $n -p 32 -s
SOAP.config -o SOAPdenovo-Trans_output/SOAP_$n 1>SOAP_k$n.log
2>SOAP_k$n.err; done
for ((n=41; n<131; n=n+10)); do SOAPdenovo-Trans-127mer all -K $n -p 32
-s SOAP.config -o SOAPdenovo-Trans_output/SOAP_$n 1>SOAP_k$n.log
2>SOAP_k$n.err; done
for ((n=35; n<135; n=n+10)); do SOAPdenovo-Trans-127mer all -K $n -p 32
-s SOAP.config -o SOAPdenovo-Trans_output/SOAP_$n 1>SOAP_k$n.log
2>SOAP_k$n.err; done

#SOAPdenovo-Trans is a relatively fast assembler, however assembly may
require a large amount of RAM (>500GB)
```

## **Step 7:** Assemble reads with Trinity

Assemble reads using Trinity (k-mer size set at 25). Trinity is the most widely used de novo transcriptome assembler: it is a de Bruijn graph-based assembler that implements an error removal model when constructing transcripts. It retains and clusters transcript isoforms, and is often as a benchmark to evaluate novel assemblers.

*In silico* normalization can be employed in the Trinity pipeline to reduce read counts before assembly. The normalized reads can be obtained (from /trinity_out_dir/insilico_read_normalization/*.fq) and assembled using other assemblers (SOAPdenovo-Trans and Velvet/Oases; Repeating steps 5-9).

Assembly statistics can be generated using the bundled 'TrinityStats.pl' script.

Trinity:
Download - https://github.com/trinityrnaseq/trinityrnaseq/releases (ver2.3.2)
Documentation - https://github.com/trinityrnaseq/trinityrnaseq/wiki
Requirements - Trinity comes bundled with required programs, including fastool, jellyfish, parafly, samtools and trimmomatic.
Reference - Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q. and Chen, Z., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology, 29(7), pp.644-652.

```
#Install Trinity
cd $PROGRAMDIR
wget https://github.com/trinityrnaseq/trinityrnaseq/archive/Trinity-
v2.3.2.tar.gz -O Trinity-v2.3.2.tar.gz
tar zxvf Trinity-v2.3.2.tar.gz
cd trinityrnaseq
make
make plugins
#test installation
cd sample_data/test_Trinity_Assembly/ && ./runMe.sh
#if successful, add trinityrnaseq, trinity-plugins, util, misc, and
support_scripts directories to PATH
echo export PATH=$PATH$( find $PROGRAMDIR/trinityrnaseq/ -type d -printf
":%p" ) >> ~/.bashrc && source ~/.bashrc

#Use Trinity to assemble error-corrected (fasta) and non-error-corrected
(fastq) reads. Some user-defined parameters for Trinity are version-
specific: check documentation before use

cd $WORKDIR/cleaned_trimmed_reads
Trinity --seqType fq --JM 40G --left R1_pairedout.fastq --right
R2_pairedout.fastq --CPU 12 --min_kmer_cov 2 --output Trinity-DN
1>trinity.log 2>trinity.err

#Move Trinity assembly to Trinity-DN directory
mv $WORKDIR/cleaned_trimmed_reads/Trinity-
DN/trinity_out_dir/Trinity.fasta ..

#Basic assembly statistics can be generated using the TrinityStats.pl
script included in the package
cd $WORKDIR/cleaned_trimmed_reads/Trinity-DN/
TrinityStats.pl Trinity.fasta > Stats.txt
```

## **Step 8:** Align reads to Trinity.fasta

Align reads to assembled transcripts using Bowtie to provide a measure of transcript 'completeness', and to calculate the average insert size if unknown (required for Velvet / Oases). Bowtie is a fast, memory-efficient short read aligner and Picard tools is a collection of command line utilities for processing next-gen sequencing data.

Bowtie:
Download - https://sourceforge.net/projects/bowtie-bio/files/bowtie/1.2.0/ (ver1.2)
Documentation - http://bowtie-bio.sourceforge.net/index.shtml
Reference - Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25.

Picard tools:
Download - https://broadinstitute.github.io/picard/ (ver2.1.1)
Documentation - https://broadinstitute.github.io/picard/command-line-overview.html
Reference - https://broadinstitute.github.io/picard/

```
#Install Bowtie with anaconda
conda install bowtie
#install bowtie from tarball
cd $PROGRAMDIR
wget https://downloads.sourceforge.net/project/bowtie-
bio/bowtie/1.2.0/bowtie-1.2-linux-x86_64.zip -O bowtie-1.2.zip
#unpack
unzip bowtie-1.2.zip && cd bowtie-1.2
#add binaries to a directory contained in PATH, or add current
directory to PATH
echo export PATH=\$PATH:`pwd`\ >> ~/.bashrc && source ~/.bashrc

#Install Picard tools using linuxbrew
brew install picard-tools
#install from source
wget
https://github.com/broadinstitute/picard/releases/download/2.1.1/picard
-tools-2.1.1.zip -O picard-tools-2.1.1.zip
unzip picard-tools-2.1.1.zip
#add java binaries to a directory contained in PATH, or add current
directory to PATH
echo export PATH=\$PATH:`pwd`\ >> ~/.bashrc && source ~/.bashrc

#Build a bowtie alignment database
cd $WORKDIR/cleaned_trimmed_reads/Trinity-DN/
bowtie-build Trinity.fasta Trinity_bowtie

#Run bowtie to align fastq reads to the assembly
bowtie -q --phred33-quals -n 2 -e 99999999 -l 25 -I 1 -X 1000 -p 20 -a
-m 200 --chunkmbs 128 -S Trinity_bowtie -1 R1_pairedout.fastq -2
R2_pairedout.fastq trinity_backmap_paired.sam
1>bowtie_trinity_backmapping.log 2>bowtie_trinity_backmapping.err

#Convert alignment file in SAM format (human readable) to BAM format
(binary)
samtools view -bS trinity_backmap_paired.sam >
trinity_backmap_paired.bam && rm trinity_backmap_paired.sam

#Classify alignments as properly paired (i.e. on the same contig with a
reasonable insert length), aligned, or not aligned using the samtools
flagstat tool
samtools flagstat trinity_backmap_paired.bam >
trinity_backmap_paired_flagstat.txt

#Calculate the average insert size
java -jar picard.jar CollectInsertSizeMetrics
I=Trinity_backmap_paired.bam O=insert_size_output.txt
H=insert_size_output.pdf M=0.5
```

## **Step 9:** Assembling reads with Velvet / Oases

Velvet is a de Bruijn graph based short-read assembler designed for de novo genome assembly. Oases takes contigs assembled using Velvet and solves sub-graphs caused by transcript isoforms to build transcripts.

Use Velvet to assemble reads at multiple kmer sizes (21, 23, 25, 27, 29, 31, 35, 41, 51, 61, 71, 81, 91, 101, 111, 121), then use Oases to assemble contigs into transcripts.

Assembly statistics are contained in the file 'stats.txt'.

Velvet:
Download - https://www.ebi.ac.uk/~zerbino/velvet/ (ver1.2.10)
Documentation - http://www.ebi.ac.uk/~zerbino/velvet/Manual.pdf
Reference - Velvet: algorithms for de novo short read assembly using de Bruijn graphs. D.R. Zerbino and E. Birney. Genome Research 18:821-829.

Oases:
Download - http://www.ebi.ac.uk/~zerbino/oases/ (ver0.2.08)
Documentation - http://www.ebi.ac.uk/~zerbino/oases/OasesManual.pdf

Reference -M.H. Schulz, D.R. Zerbino, M. Vingron and Ewan Birney. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics, 2012. DOI: 10.1093/bioinformatics/bts094.

COMMANDS (UNIX - BASH)

```
#Download velvet
cd $PROGRAMDIR
wget http://www.ebi.ac.uk/~zerbino/velvet/velvet_1.2.10.tgz -O
velvet_1.2.10.tar.gz
#unpack
tar zxvf velvet_1.2.10.tar.gz && cd velvet_1.2.10
#specify number of threads to use for assembly and set variables for
velvet
export OMP_NUM_THREADS=31
export OMP_THREAD_LIMIT=32
make 'MAXKMERLENGTH=141' 'BIGASSEMBLY=1' 'OPENMP=1' 'LONGSEQUENCES=1'
#add binaries to a directory contained in PATH, or add current directory
to PATH
echo export PATH=\$PATH:`pwd`\ >> ~/.bashrc && source ~/.bashrc

#Download Oases
cd $PROGRAMDIR
git clone --recursive https://github.com/dzerbino/oases.git
cd oases
export OMP_NUM_THREADS=31
export OMP_THREAD_LIMIT=32
make 'MAXKMERLENGTH=141' 'BIGASSEMBLY=1' 'OPENMP=1' 'LONGSEQUENCES=1'
#add binary to a directory contained in PATH, or add current directory to
PATH
echo export PATH=\$PATH:`pwd`\ >> ~/.bashrc && source ~/.bashrc
```

```
#Run velveth
cd $WORKDIR/cleaned_trimmed_reads/
mkdir velvet-oases_output && cd velvet-oases_output/
~/velvet_1.2.10/velveth dir 21,37,2 -shortPaired -fastq -separate
../R1_pairedout.fastq ../R2_pairedout.fastq 1>velveth_21-35.log
2>velveth_21-35.err
~/velvet_1.2.10/velveth dir 41,131,10 -fastq -shortPaired -separate
../R1_pairedout.fastq ../R2_pairedout.fastq 1>velveth_41-121.log
2>velveth_41-121.err

#Run velvetg
cd $WORKDIR/cleaned_trimmed_reads/velvet-oases_output/
for f in dir_*; do velvetg $f -read_trkg yes -ins_length 215 1>$f.log
2>$f.err; done

#Run Oases (include insert length calculated in step 8)
cd $WORKDIR/cleaned_trimmed_reads/velvet-oases_output/
for f in dir_*; do oases $f -min_trans_lgth 100 -ins_length 215
1>oases_$f.log 2>oases_$f.err; done
```

## **Step 10:** Remove redundant transcript

Reduce transcript redundancy using the CD-HIT algorithm. CD-HIT is used to cluster similar biological sequences together, reducing sequence redundancy and improving the performance and accuracy of specific downstream sequence analyses.

De novo assembly often produces highly similar sequences. These can be biological, for example transcript isoforms and transcripts from the same family, or artifacts of the assembly process, such as chimeric transcripts, unsupported insertions, incomplete/fragmented/misassembled or duplicate transcripts. The threshold for clustering depends on the specific analysis being performed: to cluster exact duplicates a threshold of '1.0' is used, to cluster transcripts that are share 99% similarity a threshold of '0.99' is used, and so on.

CD-HIT-EST:
Download - https://github.com/weizhongli/cdhit (ver4.6.4)
Documentation - http://weizhongli-lab.org/lab-wiki/doku.php?
Reference - 'Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences', Weizhong Li & Adam Godzik Bioinformatics, (2006) 22:1658-9. / Limin Fu, Beifang Niu, Zhengwei Zhu, Sitao Wu and Weizhong Li, CD-HIT: accelerated for clustering the next generation sequencing data. Bioinformatics, (2012), 28 (23): 3150-3152. doi: 10.1093/bioinformatics/bts565.

*All analyses and validation steps described in this protocol use 'un-clustered' assembly .fasta files.

COMMANDS (UNIX - BASH)

```
#Install CD-HIT-EST
cd $PROGRAMDIR
git clone --recursive https://github.com/weizhongli/cdhit.git
cd cdhit
make
#add binaries to a directory contained in PATH, or add current directory
to PATH
echo export PATH=\$PATH:`pwd`\ >> ~/.bashrc && source ~/.bashrc

#for SOAPdenovo-Trans
cd $WORKDIR/cleaned_trimmed_reads/SOAPdenovo-Trans_output/ && mkdir
SOAP_cdhit
for f in SOAP*/*.contig; do cd-hit-est -i $f -o
SOAP_cdhit/clustered_95_$f -c 0.95 -n 8 -p 1 -g 1 -M 200000 -T 12 -d 40
1>SOAP_cdhit/$f.log 2>SOAP_cdhit/$f.err; done

#for Trinity
cd $WORKDIR/cleaned_trimmed_reads/Trinity-DN/ && mkdir trinity_cdhit
cd-hit-est -i Trinity.fasta -o trinity_cdhit/Trinity_clustered_0.95 -c
0.95 -n 8 -p 1 -g 1 -M 200000 -T 12 -d 40 1>trinity_cdhit/Trinity_cd-hit-
est_0.95.log 2>trinity_cdhit/Trinity_cd-hit-est_0.95.err

#for Velvet/Oases
cd $WORKDIR/cleaned_trimmed_reads/velvet-oases_output/ && mkdir velvet-
oases_cdhit
for f in dir_*; do cd-hit-est -i $f/transcripts.fa -o velvet-
oases_cdhit/clustered_95_$f -c 0.95 -n 8 -p 1 -g 1 -M 200000 -T 12 -d 40
1>velvet-oases_cdhit/$f.log 2>velvet-oases_cdhit/$f.err; done
```

## **Step 11:** Remove SOAPdenovo-Trans 'gaps'

Run GapCloser on the SOAPdenovo-Trans assemblies. GapCloser is bundled with SOAPdenovo2, and removes strings of 'N' introduced in transcripts during scaffolding by SOAPdenovo-Trans. Removal of these gaps can improve accuracy and performance in specific downstream analysis. GapCloser is designed to remove assembly artifacts (N's) using the abundant pair relationships of short reads.

GapCloser:
Download - https://sourceforge.net/projects/soapdenovo2/files/GapCloser/ (ver1.12-r6)
Documentation - Read the GapCloser_Manual.pdf in GapCloser-bin-v1.12-r6.tgz
Reference - Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y. and Tang, J., 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience, 1(1), p.18.

*Use of 'non-GapClosed' or 'post-GapClosed' SOAPdenovo-Trans assemblies depends on the analyses applied. In many cases the inclusion of long strings of 'N' will cause errors or poor performance: if this occurs, use the GapClosed assemblies.

```
#Download and install GapCloser
cd $PROGRAMDIR
wget
https://sourceforge.net/projects/soapdenovo2/files/GapCloser/src/r6/
GapCloser-src-v1.12-r6.tgz/download --no-check-certificate -O
GapCloser.tar.gz
tar zxvf GapCloser.tar.gz
cd v1.12-r6 && make
#add binaries to a directory contained in PATH, or add current
directory to PATH
cd bin && echo export PATH=\$PATH:`pwd`\ >> ~/.bashrc && source
~/.bashrc

#Copy assemblies to a single directory and rename to *.fasta
cd $WORKDIR/cleaned_trimmed_reads/SOAPdenovo-Trans_output/
mkdir SOAP_assembly_fasta_files
for f in SOAP*; do cp $f/*.scafSeq SOAP_assembly_fasta_files; done
cd SOAP_assembly_fasta_files
for f in *.scafSeq; do mv $f $(echo $f | sed 's/.scafSeq/.fasta/g');
done

#Create a config file (similar to Step 5)
cd $WORKDIR/cleaned_trimmed_reads/SOAPdenovo-
Trans_output/SOAP_assembly_fasta_files
cat > GapCloser.config <<END_TEXT
max_rd_len=150
[LIB]
avg_ins=192
reverse_seq=0
asm_flags=3
q1=$WORKDIR/cleaned_trimmed_reads/R1_pairedout.fastq
q2=$WORKDIR/cleaned_trimmed_reads/R2_pairedout.fastq
END_TEXT

#Run GapCloser
cd $WORKDIR/cleaned_trimmed_reads/SOAPdenovo-
Trans_output/SOAP_assembly_fasta_file
for f in *.fasta; do GapCloser -b GapCloser.config -a $f -o
GapClosed_$f -l 150 -t 24 1>GapCloser_{$f}.log 2>GapCloser_{$f}.err;
done
```

## **Step 12:** Merge single k-mer assemblies

Move all transcriptome assemblies to a single directory. Use the EvidentialGene tr2aacds pipeline and/or Transfuse to identify high-quality, non-redundant transcripts from all single-kmer assemblies, and combine these to form a 'merged' or 'clustered' assembly. EvidentialGene tr2aacds.pl processes de novo assemblies with different kmer sizes (and from different assemblers), into the most biologically useful 'best' set of mRNA, classified into primary and alternate transcripts.
Similarly, Transfuse intelligently merges multiple de novo transcriptome assemblies, combining 'high-grade' transcripts into a single high quality transcriptome.

EvidentialGene tr2aacds
Download - http://arthropods.eugenes.org/genes2/about/EvidentialGene_trassembly_pipe.html
Documentation - http://arthropods.eugenes.org/EvidentialGene/evigene/
Reference - Gilbert, Donald (2013) Gene-omes built from mRNA seq not genome DNA. 7th annual arthropod genomics symposium. Notre Dame.
http://arthropods.eugenes.org/EvidentialGene/about/EvigeneRNA2013poster.pdf and
http://globalhealth.nd.edu/7th-annual-arthropod-genomics-symposium/ and
doi:10.7490/f1000research.1112594.1 / Gilbert D. (2016) Accurate & complete gene construction with EvidentialGene.
Talk at Galaxy Community Conference 2016, Bloomington IN. F1000Research, 5:1567 (slide set).
doi:10.7490/f1000research.1112467.1
http://eugenes.org/EvidentialGene/about/evigene_bothgalmod1606iu.pdf (Galaxy+GMOD full slide set)

Transfuse
Download - https://github.com/cboursnell/transfuse/releases/ (ver0.5.0)
Documentation - https://github.com/cboursnell/transfuse
Requirements - dependencies are bundled with the transfuse tarball
Reference - https://github.com/cboursnell/transfuse

COMMANDS (UNIX - BASH)

```
#Download and install EvidentialGene tr2aacds
wget ftp://arthropods.eugenes.org/evigene.tar
tar xvf evigene.tar
cd evigene/
#add executables from scripts directory and subdirectories to PATH
echo export PATH=$PATH$( find $PROGRAMDIR/evigene/scripts/ -type d -
printf ":%p" ) >> ~/.bashrc

#Concatenate all fasta assembly files
cd $WORKDIR/cleaned_trimmed_reads
mkdir all_assemblies
cd SOAPdenovo-Trans_output/SOAP_assembly_fasta_files
cp SOAP_* ../../all_assemblies; done
cd ../../velvet-oases_output
mkdir velvet-oases_fasta_files
for f in dir_*; do mv $f/transcripts.fa $f/velvet-oases_{$f}.fasta;
done
for f in dir_*; do cp $f/velvet-oases* velvet-oases_fasta_files; done
```

```
cd velvet-oases_fasta_files && cp *.fasta ../../all_assemblies
cd ../Trinity-DN
cp Trinity.fasta ../all_assemblies
cd ../all_assemblies
cat *.fasta > All_assemblies.fasta
mkdir tr2aacds_merge && mv All_assemblies.fasta tr2aacds_merge/

#Rename fasta headers
cd $WORKDIR/cleaned_trimmed_reads/all_assemblies/tr2aacds_merge/
perl -ane 'if(/\>/){$a++;print ">Locus_$a\n"}else{print;}'
All_assemblies.fasta > All_renamed.fasta

#Reformat fasta headers, then run the tr2aacds pipeline
trformat.pl -output All_assemblies.tr -input All_renamed.fasta
rm All_assemblies.fasta All_renamed.fasta
tr2aacds.pl -mrnaseq All_assemblies.tr -NCPU=40 1>tr2aacds.log
2>tr2aacds.err

#Download and install transfuse using Ruby
gem install transfuse

#or, alternatively, from source
wget
https://github.com/cboursnell/transfuse/releases/download/v0.5.0/
transfuse-0.5.0-linux-x86_64.tar.gz -O transfuse.tar.gz
tar -zxvf transfuse.tar.gz

#add binary to a directory contained in PATH, or add current
directory to PATH
echo export PATH=\$PATH:`pwd`\ >> ~/.bashrc && source ~/.bashrc

#Run transfuse on the fasta assemblies to merge non-redundant and
read-supported contigs together
cd $WORKDIR/cleaned_trimmed_reads/all_assemblies
mkdir transfuse
LIST=`ls -m *.fasta | tr -d " \t\n\r"`
transfuse -a $LIST -l R1_pairedout.fastq -r R2_pairedout.fastq -o
transfuse/transfused_assembly -t 40 1>tfuse.log 2>tfuse.err

#If transfuse fails due to insufficient MCP, symlink snap-aligner
from TransRate into the transfuse /bin/ directory (or
/ruby*/gems/transrate/ directory if install using Ruby) after
editing the snap.rb file to increase MCP as described in Step 15.

#Further problems/issues should be directed to
https://github.com/cboursnell/transfuse/issues or to
https://gitter.im/cboursnell/transfuse
```

## **Step 13:** Search for BUSCOs

Search all assemblies for the presence/absence of Benchmarking Universal Single-Copy Orthologs (BUSCOs). BUSCO provides a quantitative measure of transcriptome quality and completeness, based on evolutionarily-informed expectations of gene content from near-universal single-copy orthologs selected from OrthoDB v9.

Aligned BUSCOs are classified as 'complete', 'fragmented', or 'missing'. 'Complete' genes are further categorised as 'single' or 'duplicate', with genomes ideally containing a single copy of each gene (transcriptomes often have multiple copies of transcripts). This tool provides a genome-free/reference-free validation of transcriptome assembly quality, and allows comparison between multiple assemblies.

BUSCO:
Download - https://gitlab.com/ezlab/busco (ver1.22)
Documentation - http://busco.ezlab.org/
Requirements - species-specific genesets (from http://busco.ezlab.org/)
Reference - Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, p.btv351.

COMMANDS (UNIX - BASH)

```
#For easy install use homebrew / linuxbrew (http://linuxbrew.sh/)
cd $PROGRAMDIR
brew install busco
#to download and install from source
cd $PROGRAMDIR
wget https://gitlab.com/ezlab/busco/repository/archive.tar.gz?ref=master
-O BUSCO.tar.gz
tar -zxvf BUSCO.tar.gz && rm BUSCO.tar.gz && mv busco* BUSCO_v1.22
cd BUSCO_v1.22
echo export PATH=\$PATH:`pwd`\ >> ~/.bashrc && source ~/.bashrc

#Download the required BUSCO catalogue (eg vertebrata and/or eukaryota)
and unpack
cd $PROGRAMDIR/BUSCO_v1.22/
wget http://busco.ezlab.org/v1/files/eukaryota_buscos.tar.gz -O
eukaryota.tar.gz
wget http://busco.ezlab.org/v1/files/vertebrata_buscos.tar.gz -O
vertebrata.tar.gz
tar -zxvf *.tar.gz

#Run busco
cd $WORKDIR/cleaned_trimmed_reads/all_assemblies
for f in *.fasta; do python BUSCO_v1.22.py -o busco_vert_$f -i $f -l
$PROGRAMDIR/BUSCO_v1.22/vertebrata -m tran -c 32 -f 1>{$f}_busco_vert.log
2>{$f}_busco_vert.err; done

for f in *.fasta; do python BUSCO_v1.22.py -o busco_euk_$f -i $f -l
$PROGRAMDIR/BUSCO_v1.22/eukaryota -m tran -c 32 -f 1>{$f}_busco_euk.log
2>{$f}_busco_euk.err; done
```

## **Step 14:** Search for CEGMAs

Run CEGMA on assemblies to identify the presence/absence of core eukaryotic genes. The most accurate and complete transcriptome assemblies are expected to contain a greater number of CEGMA core genes. CEGMA is considered deprecated software: it is no longer supported, and has effectively been superseded by BUSCO, however CEGMA scores may still be useful for comparing new transcriptome assemblies to previously established assemblies.

CEGMA is difficult to install; a guide can be found at
http://korflab.ucdavis.edu/datasets/cegma/ubuntu_instructions_1.txt
Additional information on compiling genewise is available at
http://seqanswers.com/forums/archive/index.php/t-24027.html

CEGMA:
Download - http://korflab.ucdavis.edu/datasets/cegma/#SCT3 (ver2.5)
Documentation - http://korflab.ucdavis.edu/datasets/cegma/README / http://korflab.ucdavis.edu/Datasets/cegma/faq.html
Requirements - there are a number of dependancies, as detailed in the install guide.
Reference - G. Parra, K. Bradnam and I. Korf. 'CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes.'
Bioinformatics, 23: 1061-1067 (2007) / Genis Parra, Keith Bradnam, Zemin Ning, Thomas Keane, and Ian Korf. Assessing the gene space in draft genomes' Nucleic Acids Research, 37(1): 298-297 (2009)

COMMANDS (UNIX - BASH)

```
#Download and install CEGMA
cd $PROGRAMDIR
wget http://korflab.ucdavis.edu/datasets/cegma/CEGMA_v2.5.tar.gz
tar zxvf CEGMA_v2.5.tar.gz
rm CEGMA_v2.5.tar.gz && CEGMA_v2.5
make
#add executables to a directory contained in PATH, or add current
directory to PATH
cd bin && echo export PATH=\$PATH:`pwd`\ >> ~/.bashrc && source ~/.bashrc

###Ensure all required dependencies are installed as detailed in the
installation guide

#Run CEGMA on each assembly
cd $WORKDIR/cleaned_trimmed_reads/all_assemblies
mkdir cegma_output
for f in *.fasta; do mkdir cegma_output/dir_$f && cp $f
cegma_output/dir_$f; done
cd $WORKDIR/cleaned_trimmed_reads/all_assemblies/cegma_output
export CEGMA=$PROGRAMDIR/cegma
for f in
$WORKDIR/cleaned_trimmed_reads/all_assemblies/cegma_output/dir_*/*.fasta;
do ~/cegma/bin/cegma --genome $f -o
$WORKDIR/cleaned_trimmed_reads/all_assemblies/cegma_output/dir_*/cegma_$f
-T 32; done
#If the 'for loop' fails, run CEGMA on assemblies individually
```

## **Step 15:** Run TransRate

TransRate is software for de-novo transcriptome quality analysis. It examines each assembly in detail and reports quality scores for contigs and assemblies, allowing you to identify the optimal assembly, filter out contigs from an assembly that are not supported by the reads, and help decide when to stop trying to improve the assembly.

TransRate:
Download - https://github.com/blahah/transrate/releases (ver1.2)
Documentation - http://hibberdlab.com/transrate/index.html
Requirements - snap-aligner (v1.0beta18), bam-read (v1.0), salmon (v0.6.0), vsearch (ver1.8.0): bundled with transrate tarball
Reference - TransRate: reference free quality assessment of de-novo transcriptome assemblies (2016). Richard D Smith-Unna, Chris Boursnell, Rob Patro, Julian M Hibberd, Steven Kelly. Genome Research doi: http://dx.doi.org/10.1101/gr.196469.115

COMMANDS (UNIX - BASH)

```
#To install using ruby
cd $PROGRAMDIR
gem install transrate
transrate --install-deps all
#to install from source
cd $PROGRAMDIR
wget https://bintray.com/artifact/download/blahah/generic/transrate-
1.0.2-linux-x86_64.tar.gz -O transrate.tar.gz
tar -zxvf transrate.tar.gz
cd transrate-1.0.2-linux-x86_64/bin
#add executables to a directory contained in PATH, or add current
directory to PATH
echo export PATH=\$PATH:`pwd`\ >> ~/.bashrc && source ~/.bashrc

#Prior to run, edit /transrate/lib/app/lib/transrate/snap.rb to increase
MCP

#edit the function “build_paired_cmd”, by finding the line “> cmd << " -
omax 10" # max alignments per pair/read” and inserting “> cmd << " -mcp
10000000" # maximum candidate pool size” after it.

#Run TransRate
cd $WORKDIR/cleaned_trimmed_reads/all_assemblies/
for f in *.fasta; do transrate --assembly $f --left R1_pairedout.fastq --
right R2_pairedout.fastq --output ./transrate_$f --threads 32
1>trate_$f.log 2>trate_$f.err; done
```

## **Step 16:** Align reads to each assembly

Use Bowtie to align reads to each assembly. The proportion of properly aligned reads is stored in the *_flagstat.txt files. This provides a general assessment of assembly completeness: if many reads align properly (i.e. on the same contig), it suggests the assembly quality is high.

Bowtie:
Download - https://sourceforge.net/projects/bowtie-bio/files/bowtie/1.2.0/ (ver1.2)
Documentation - http://bowtie-bio.sourceforge.net/index.shtml
Reference - Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25.

COMMANDS (UNIX - BASH)

```
#Bowtie download and install detailed at Step 8

#Build an alignment database for each assembly
cd $WORKDIR/cleaned_trimmed_reads/all_assemblies
for f in *.fasta; do bowtie-build $f $f_bowtie; done

#Run bowtie and convert output from SAM format (human readable) to BAM
format (binary)
for f in *.fasta; do bowtie -q --phred33-quals -n 2 -e 99999999 -l 25
-I 1 -X 1000 -p 20 -a -m 200 --chunkmbs 128 -S $f_bowtie -1
$WORKDIR/cleaned_trimmed_reads/R1_pairedout.fq -2
$WORKDIR/cleaned_trimmed_reads/R2_pairedout.fq $f.sam
1>bowtie_$f_backmapped.log 2>bowtie_$f_backmapping.err && samtools
view -bS $f.sam > $f.bam && rm $f.sam; done

#Classify alignments as properly paired (i.e. on the same contig with
a reasonable insert length), aligned, or not aligned using the
samtools flagstat tool
for f in *.bam; do samtools flagstat $f > $f_flagstat.txt; done
```

## **Step 17:** Annotation

Idenitfy the most complete and accurate de novo transcriptome assemblies based on BUSCO/CEGMA scores, transrate scores, and properly paired read alignment, and annotate them using the Trinotate or Dammit pipelines. Annotation of transcripts aids inferrence of biological function, for instance identifying transcripts that have functional domains (Pfam domains), signaling proteins, and transmembrane proteins.

Trinotate is a suite of tools for functional annotation of de novo assembled transcriptomes. Trinotate employs sequence homology search to known transcripts (BLAST+/SwissProt), protein domain identification (HMMER/PFAM), protein signal peptide and transmembrane domain prediction (signalP/tmHMM), and leveraging various annotation databases (eggNOG/GO/Kegg databases).

Dammit performs essentially the same task: it is a simple, yet comprehensive, de novo transcriptome annotator born from the observations that annotation is time-consuming, mundane and often frustrating, with many existing solutions being overly complicated or relying on non-free software.

First, install perl module URI::Escape using cpan or cpanm (curl -L https://cpanmin.us | perl - App::cpanminus). Download the swissprot/uniprot database and prepare for blast.

Trinotate:
Download - https://github.com/Trinotate/Trinotate/releases (ver2.0.2)
Documentation - https://trinotate.github.io/
Requires - transdecoder, sqlite, ncbi-blast+, HMMER/PFAM, signalP, tmhmm, RNAmmer (further details available at https://trinotate.github.io/)
Reference - https://trinotate.github.io/

Dammit:
Download - https://github.com/camillescott/dammit
Documentation - http://dammit.readthedocs.io/en/latest/ ;
http://dammit.readthedocs.io/en/latest/installing.html
Requires - many, many dependencies, run within a virtual python environment. Further explained in the doumentation above
Reference - Camille Scott (2016) 'dammit: an open and accessible de novo transcriptome annotator', in preparation; www.camillescott.org/dammit

COMMANDS (UNIX - BASH)

```
#A guide for installing and using Trinotate is available at
https://trinotate.github.io/
cd $PROGRAMDIR
cpan URI::Escape
wget https://github.com/Trinotate/Trinotate/archive/v3.0.1.tar.gz -O
Trinotate_v3.0.1.tar.gz
tar -zxvf Trinotate_v3.0.1.tar.gz && rm Trinotate_v3.0.1.tar.gz
cd Trinotate_v3.0.1/sample_data
./runMe.sh
#ensure 'runMe.sh' test completed successfully, then run:
./cleanMe.pl
cd $PROGRAMDIR/Trinotate_v3.0.1/admin/
Build_Trinotate_Boilerplate_SQLite_db.pl  Trinotate

#add executables from Trinotate directory and subdirectories to PATH
echo export PATH=$PATH$( find $PROGRAMDIR/Trinotate_v3.0.1/ -type d
-printf ":%p" ) >> ~/.bashrc

#Trinotate
#Identify ORFs using TransDecoder
cd $WORKDIR/cleaned_trimmed_reads/Trinity-DN/
TransDecoder.LongOrfs -t Trinity.fasta

#Download, unpack, and build the Uniprot/Swissprot database
cd $PROGRAMDIR
mkdir databases && cd databases/
```

```
wget
ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgeb
ase/complete/uniprot_sprot.fasta.gz && gunzip uniprot_sprot.fasta.gz
makeblastdb -in uniprot_sprot.fasta -dbtype prot -parse_seqids

#Blast transcripts against the database to identify potential
orthologs
cd $WORKDIR/cleaned_trimmed_reads/Trinity-DN/
blastx -query Trinity.fasta -db $PROGRAMDIR/databases/uniprot_sprot -
evalue 1e-5 -num_threads 40 -max_target_seqs 1 -outfmt 6 >
blastx_uniprot_1e05.outfmt6
blastp -query longest_orfs.pep -db $PROGRAMDIR/databases/uniprot_sprot
-evalue 1e-5 -num_threads 40 -max_target_seqs 1 -outfmt 6 >
blastp_uniprot_1e05.outfmt6

#Use an HMM search to identify PFAM domains in the longest ORFs
hmmscan --cpu 20 --domtblout TrinotatePFAM.out Pfam-A.hmm
longest_orfs.pep

#Use an HMM search to identify PFAM domains in uniprot_sprot blastp
hits
TransDecoder.Predict -t target_transcripts.fasta --retain_pfam_hits
TrinotatePFAM.out --retain_blastp_hits blastp_uniprot_1e05.outfmt6

#Identify signalling proteins in the longest ORFs using SignalP
signalp -f short -n signalp.out longest_orfs.pep &

#Identify rRNA sequences within the transcriptome using RNAmmer
RnammerTranscriptome.pl --transcriptome Trinity.fasta --
path_to_rnammer $PROGRAMDIR/rnammer-1.2/rnammer

#Identify transmembrane proteins using tmhmm
tmhmm --short < transdecoder.pep > tmhmm.out

#Build the boilerplate trinotate sqlite database
Trinotate Trinotate.sqlite init --gene_trans_map
Trinity.fasta.gene_trans_map --transcript_fasta Trinity.fasta --
transdecoder_pep Trinity.fasta.transdecoder_dir/longest_orfs.pep

#Load annotation results into the SQLite database
Trinotate Trinotate.sqlite LOAD_swissprot_blastp blastp_uniprot_1e05.outfmt6
Trinotate Trinotate.sqlite LOAD_swissprot_blastx blastx_uniprot_1e05.outfmt6
Trinotate Trinotate.sqlite LOAD_pfam TrinotatePFAM.out
Trinotate Trinotate.sqlite LOAD_signalp signalp.out
Trinotate Trinotate.sqlite LOAD_tmhmm tmhmm.out
Trinotate Trinotate.sqlite LOAD_rnammer Trinity.fasta.rnammer.gffTrinotate

#Generate annotation reports
Trinotate Trinotate.sqlite report > trinotate_annotation_report.xls
Trinotate Trinotate.sqlite report -E 0.00001 >
trinotate_annotation_report_evalue00001.xls
```

```
#Extract GO terms from the report
extract_GO_assignments_from_Trinotate_xls.pl --Trinotate_xls
trinotate_annotation_report.xls -G --include_ancestral_terms >
go_annotations.txt

#Use trinotate to import transcript name information into the report
import_transcript_names.pl Trinotate.sqlite Trinotate_report.xls

#Identify and collate all columns containing a value for each loaded
result
awk '$3 ~ !/./' trinotate_report_ex_GO.xls >
trinotate_report_uniprot_hits.xls
awk '$8 ~ !/./' trinotate_report_ex_GO.xls >
trinotate_report_uniprot_protein_hits.xls
awk '$10 ~ !/./' trinotate_report_ex_GO.xls >
trinotate_report_pfam_hits.xls
awk '$11 ~ !/./' trinotate_report_ex_GO.xls >
trinotate_report_signalling_protein_hits.xls

#Include feature names/IDs
Trinotate_get_feature_name_encoding_attributes.pl Trinotate_report.xls
> Trinotate_report.xls.annotate_ids

#Extract GO term assignments for further analysis
extract_GO_assignments_from_Trinotate_xls.pl Trinotate_report.xls >
Trinotate_GO_assignments.txt

#Dammit
#Download and install
cd $PROGRAMDIR
wget https://github.com/camillescott/dammit/archive/v0.3.2.tar.gz -O
dammit.tar.gz
tar -zxvf dammit.tar.gz & rm dammit.tar.gz

cd dammit-0.3.2
make
sudo apt-get update
sudo apt-get install python-pip python-dev python-numpy git ruby hmmer
unzip infernal ncbi-blast+ liburi-escape-xs-perl emboss liburi-perl
build-essential libsm6 libxrender1 libfontconfig1 parallel
sudo gem install crb-blast

cd $PROGRAMDIR
curl -LO
https://github.com/TransDecoder/TransDecoder/archive/2.0.1.tar.gz -O
TransDecoder_2.0.1.tar.gz
tar -xvzf TransDecoder_2.0.1.tar.gz && rm TransDecoder_2.0.1.tar.gz
cd TransDecoder-2.0.1
make
echo 'export PATH=$PATH:$PROGRAMDIR/TransDecoder-2.0.1' >> ~/.bashrc
```

```
cd $PROGRAMDIR
curl -LO http://last.cbrc.jp/last-658.zip
unzip last-658.zip
cd last-658
make
echo 'export PATH=$PATH:$PROGRAMDIR/last-658/src' >> ~/.bashrc

cd $PROGRAMDIR
curl -LO http://busco.ezlab.org/v1/files/BUSCO_v1.22.tar.gz
tar -xvzf BUSCO_v1.22.tar.gz
chmod +x BUSCO_v1.22/*.py
echo 'export PATH=$PROGRAMDIR/BUSCO_v1.22:$PATH' >> ~/.bashrc

sudo pip install -U setuptools
sudo pip install dammit
dammit databases --install --all --busco-group vertebrata

#to get the latest version of dammit: 'pip install
git+https://github.com/camillescott/dammit.git'

#load dependencies and virtual python environment per
http://dammit.readthedocs.io/en/latest/installing.html then activate
environment
cd $PROGRAMDIR
source activate dammit

#Run dammit
cd $WORKDIR/cleaned_trimmed_reads/Trinity-DN/
dammit annotate Trinity.fasta --busco-group
eukaryota --n_threads 32 1>dammit_Trinity.log
2>dammit_Trinity.err &
```

## **Step 18:** Counting full-length transcripts

Take all blastx hits from annotated transcriptomes with evalue <1e-20, and use the 'analyze_blastPlus_topHit_coverage.pl' script bundled with Trinity to align assembled transcripts to corresponding entries in the Swissprot/Uniprot database. This metric provides a measure of assembly completeness: if many transcripts are near-full-length compared to the reference transcript/protein, it suggests the transcripts are well assembled. If few are near-full-length, it suggests transcripts are fragmented/poorly assembled. This metric is useful for comparing multiple assemblies produced using the same dataset.

Only the single best matching Trinity transcript is reported for each top matching database entry. If a target protein matches multiple Trinity transcripts as their best hits, that target protein is counted only once along with that Trinity transcript that provides the highest BLAST bit score and longest match length. Also compare distribution and size of contigs using bundled script 'trinity_component_distribution'.

```
#Parse blastx output to retain transcripts with evalues <=1e-20
cd $WORKDIR/cleaned_trimmed_reads/Trinity-DN/
awk '$11 <=1e-20' blastx_uniprot_1e-05.outfmt6 > blastx_uniprot_1e-
20.outfmt6

#Use the script bundled with Trinity to analyse blast hit coverage for
each transcript
analyze_blastPlus_topHit_coverage.pl blastx.outfmt6_1e-20 Trinity.fasta
$WORKDIR/databases/uniprot_sprot

#Reformat SOAPdenovo-Trans assemblies for
'trinity_component_distribution.pl'
cd $WORKDIR/cleaned_trimmed_reads/SOAPdenovo-
Trans_output/SOAP_assembly_fasta_files
for f in *.fasta; do perl -ane 'if(/\>/){$a++;print
">c$a\n"}else{print;}' $f > tmp.fasta && sed 's/>.*/&_g1/' tmp.fasta >
{$f}_trinity_format.fasta && rm tmp.fasta; done
for f in *_trinity_format.fasta; do trinity_component_distribution.pl
$f -o dist_$f 1>comp_distr.log 2>comp_distr.err; done

#Reformat Velvet-oases assemblies for
'trinity_component_distribution.pl'
cd $WORKDIR/cleaned_trimmed_reads/velvet-oases_output/velvet-
oases_fasta_files
for f in *.fasta; do perl -ane 'if(/\>/){$a++;print
">c$a\n"}else{print;}' $f > tmp.fasta && sed 's/>.*/&_g1/' tmp.fasta >
{$f}_trinity_format.fasta && rm tmp.fasta; done
for f in *_trinity_format.fasta; do trinity_component_distribution.pl
$f -o dist_$f 1>comp_distr.log 2>comp_distr.err; done
```

## Step 19: Count full-length transcripts for all assemblies

Blastx is the recommended tool for aligning transcripts to the Uniprot/Swissprot database, however alignment takes a relatively long time for large transcriptome assemblies (>1 week with 40 threads). A considerably faster tool is DIAMOND ('DIAMOND blastx' is 20,000X faster than NCBI Blast+ 'blastx').

Download - https://github.com/bbuchfink/diamond/releases/ (ver0.8.36)
Documentation - https://github.com/bbuchfink/diamond
Reference - B. Buchfink, Xie C., D. Huson (2015), 'Fast and sensitive protein alignment using DIAMOND', Nature Methods 12, 59-60.

COMMANDS (UNIX - BASH)

```
#Download and install DIAMOND
cd $PROGRAMDIR
mkdir Diamond && cd Diamond
wget
http://github.com/bbuchfink/diamond/releases/download/v0.8.36/diamond-
linux64.tar.gz -O diamond.tar.gz
tar -zxvf diamond.tar.gz && rm diamond.tar.gz
#add executables to a directory contained in PATH, or add current
directory to PATH
echo export PATH=\$PATH:`pwd`\ >> ~/.bashrc && source ~/.bashrc

#Prepare the Swissprot/Uniprot database
cd $PROGRAMDIR/databases/
diamond makedb --in uniprot_sprot.fasta -d uniprot_sprot

#run DIAMOND blastx on each assembly
cd $WORKDIR/cleaned_trimmed_reads/all_assemblies/
for f in *.fasta; do diamond blastx -d
$PROGRAMDIR/databases/uniprot_sprot -q $f -o diamond_$f --sensitive -p
32 -k 1 -e 1e-05 -b 10 -c 1; done

#parse DIAMOND blastx output, retaining transcripts with evalues <=1e-
20
mkdir blast_hits_1e-20
for f in diamond*; do awk '$11 <=1e-20' $f > blast_hits_1e-20/1e-
20_$f; done

#analyse transcript length compared to the length of the reference
transcript
cd $WORKDIR/cleaned_trimmed_reads/all_assemblies
for f in *.fasta; do analyze_blastPlus_topHit_coverage.pl
./blast_hits_1e-20/1e-20_diamond_$f $f
$PROGRAMDIR/databases/uniprot_sprot; done

#count unique blastx hits (i.e. the number of reference transcripts
that align to one or more assembled transcripts)
cd $WORKDIR/cleaned_trimmed_reads/all_assemblies/
mkdir number_of_unique_blastx_hits
for f in diamond_*; do cut -f2 $f | sort | uniq -c | wc -l >
./number_of_unique_blastx_hits/count_$f; done

cd $WORKDIR/cleaned_trimmed_reads/all_assemblies/blast_hits_1e-20/
mkdir number_of_unique_blastx_hits_1e-20
for f in 1e-20_diamond_*; do cut -f2 $f | sort | uniq -c | wc -l >
./number_of_unique_blastx_hits_1e-20/count_$f; done
```

NOTES

12 Mar 2017

Diamond blastx is less sensitive than NCBI BLAST+ blastx, however the significant gain in speed facilitates comparison with larger databases, such as UniRef90 and nr.

## **Step 20:** Compare transcripts to a closely-related species

Use the NCBI blast+ utility to identify Mus musculus orthologs within spiny mouse transcriptome assemblies. Download the Mus RefSeq protein and RNA catalogues and convert them to blast databases. Run blastn (transcripts vs transcripts) and blastx (transcripts vs proteins), parse the output to retain hits <=1e-20, then identify which transcripts/proteins in the RefSeq catalogue correspond to one (or more) transcripts in the spiny mouse transcriptome assembly used.

Download - ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/ (ver2.6.0)
Documentation - https://www.ncbi.nlm.nih.gov/books/NBK279690/
Reference - Altschul et al. (1990) Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology. 1990;215(3):403–410. doi: 10.1016/S0022-2836(05)80360-2.
Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L., 2009. BLAST+: architecture and applications. BMC bioinformatics, 10(1), p.421.
Madden (2002) Madden T. The BLAST sequence analysis tool. In: McEntyre J, Ostell J, editors. The NCBI Handbook. Bethesda: National Center for Biotechnology Information; 2002. Chapter 16. https://www.ncbi.nlm.nih.gov/books/NBK153387/

COMMANDS (UNIX - BASH)

```
#Obtain, unpack, and build blast databases for Mus musculus RefSeq
sequences
cd $WORKDIR/databases
ascp -i ~/.aspera/connect/etc/asperaweb_id_dsa.openssh -QT -l 100M
anonftp@ftp-
private.ncbi.nlm.nih.gov:/refseq/M_musculus/mRNA_Prot/mouse.1.rna.fna.
gz anonftp@ftp-
private.ncbi.nlm.nih.gov:/refseq/M_musculus/mRNA_Prot/mouse.1.protein.
faa.gz .
gunzip *.gz
makeblastdb -in mouse.1.rna.fna -dbtype nucl -parse_seqids
makeblastdb -in mouse.1.protein.faa -dbtype prot -parse_seqids

#Blast transcripts against the RefSeq databases
cd $WORKDIR/cleaned_trimmed_reads/Trinity-DN
blastn -query Trinity.fasta -db $WORKDIR/databases/mouse.1.rna.fna -
num_threads 32 -max_target_seqs 1 -outfmt 6 >
blastn_Trinity.fasta_vs_mus_musculus_refseq.outfmt6
blastx -query Trinity.fasta -db mouse.1.protein.faa -num_threads 32 -
max_target_seqs 1 -outfmt 6 >
blastx_Trinity.fasta_vs_mus_musculus_refseq.outfmt6

#Retain hits at <=1e-20
cd $WORKDIR/cleaned_trimmed_reads/Trinity-DN
awk '$11 <=1e-20' blastn_Trinity.fasta_vs_mus_musculus_refseq.outfmt6
> blastn_Trinity.fasta_vs_mus_musculus_refseq_1e-20.outfmt6
awk '$11 <=1e-20' blastx_Trinity.fasta_vs_mus_musculus_refseq.outfmt6
> blastx_Trinity.fasta_vs_mus_musculus_refseq_1e-20.outfmt6
```

```
#Identify the number of unique RefSeq transcripts with >=1 blast hit
in the transcriptome assembly
cd $WORKDIR/cleaned_trimmed_reads/Trinity-DN
cut -f2 blastn_Trinity.fasta_vs_mus_musculus_refseq_1e-20.outfmt6 |
sort | uniq -c | wc -l > count_uniq_blastn_vs_refseq_1e-20_hits.txt
cut -f2 blastx_Trinity.fasta_vs_mus_musculus_refseq_1e-20.outfmt6 |
sort | uniq -c | wc -l > count_uniq_blastx_vs_refseq_1e-20_hits.txt
```

## **Step 21:** Identify non-coding RNAs

Use the Coding-Non-Coding Index (CNCI) to identify non-coding transcripts. CNCI profiles adjoining nucleotide triplets to distinguish protein-coding from non-coding sequences, independent of known annotations. Align transcripts to the NONCODE database for *Mus musculus* and identify unique single best blast hits.

CNCI:
Download - https://github.com/www-bioinfo-org/CNCI (version 2)
Documentation - https://github.com/www-bioinfo-org/CNCI
Reference - Liang Sun, Haitao Luo, Dechao Bu, Guoguang Zhao, Kuntao Yu, Changhai Zhang, Yuanning Liu, RunSheng Chen and Yi Zhao* Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Research (2013), doi: 10.1093/nar/gkt646

NONCODE:
Download - http://www.noncode.org/download.php (ver NONCODE2016_mouse.fa.gz)
Documentation - http://www.noncode.org/
Reference - Zhao Y, Li H, Fang S, Kang Y, Hao Y, Li Z, Bu D, Sun N, Zhang MQ, Chen R. NONCODE 2016: an informative and valuable data source of long non-coding RNAs. Nucleic acids research. (2015) Nov 19:gkv1252.

COMMANDS (UNIX - BASH)

```
#Download and install CNCI
cd $PROGRAMDIR
git clone https://github.com/www-bioinfo-org/CNCI.git
cd CNCI
unzip libsvm-3.0 && cd libsvm-3.0 && make && cd ..
chmod 755 *.py *.pl
#add scripts to a directory contained in PATH, or add current directory
to PATH
echo export PATH=\$PATH:`pwd`\ >> ~/.bashrc && source ~/.bashrc

#Run the CNCI script
cd $WORKDIR/cleaned_trimmed_reads/Trinity-DN/
python CNCI.py -f Trinity.fasta -o Trinity_ncRNA -p 32 -m ve

#Identify 'confident' non-coding transcripts (cutoff <= -0.05)
cd $WORKDIR/cleaned_trimmed_reads/Trinity-DN/Trinity_ncRNA/
awk '$3 <=-0.05' CNCI.index > CNCI_Trinity_ncRNA.txt
```

```
cd $WORKDIR/databases
wget http://www.noncode.org/datadownload/NONCODE2016_mouse.fa.gz
gunzip NONCODE2016_mouse.fa.gz
makeblastdb -in NONCODE2016_mouse.fa -dbtype nucl -parse_seqids

#Blast each assembly against the NONCODE database
cd $WORKDIR/cleaned_trimmed_reads/all_assemblies
for f in *.fasta; do blastn -query $f -db
$WORKDIR/databases/NONCODE2016_mouse.fa -num_threads 40 -max_target_seqs
1 -evalue 1e-05 -outfmt 6 > blastn_vs_noncode_$f; done
mkdir noncode_output && mv blastn_vs_noncode* noncode_output

#Parse blast output
cd $WORKDIR/cleaned_trimmed_reads/all_assemblies/noncode_output
mkdir evalue_1e-20
for f in blastn*; do awk '$11 <=1e-20' $f > 1e-20_$f; done
mv 1e-20_* evalue_1e-20 && cd evalue_1e-20
for f in 1e-20*; do cut -f2 | sort | uniq -c | wc -l >
count_best_blast_hits_$f; done
```

# Chapter 4: Embryonic genome activation

## 4.1 Overview

This manuscript describes the analysis of gene transcription in the spiny mouse embryos during the embryonic genome activation. A preprint has been posted to bioRxiv and the publication has been submitted to the journal Development.

Mamrot, J., Gardner, D. K., Temple-Smith, P., and Dickinson, H. (2018). Embryonic gene transcription in the spiny mouse (*Acomys cahirinus*): a new model of embryonic genome activation. *bioRxiv*.

bioRχiv
THE PREPRINT SERVER FOR BIOLOGY

PREPRINT ARTICLE

# Embryonic gene transcription in the spiny mouse (*Acomys cahirinus*): an investigation into the embryonic genome activation.

Jared Mamrot[1,2]*, David K. Gardner[3], Peter Temple-Smith[2,4], Hayley Dickinson[1,2].

**Our understanding of genetic mechanisms driving early embryonic development is primarily based on experiments conducted on mice, however translation of findings can be limited by physiological differences between mice and humans. To address this, we investigated whether the spiny mouse (*Acomys cahirinus*) is a closer model of early human embryonic development due to their more human-like endocrine profile. We therefore characterised the initiation of gene transcription in the spiny mouse embryo and compared the pattern of gene expression during the embryonic genome activation (EGA) with common mouse and human embryos. Naturally-mated spiny mouse embryos were obtained at the 2-cell, 4-cell and 8-cell stages of development (n=4 biological replicates per stage). RNA-Seq of these samples produced 709.1M paired-end reads in total. *De novo* assembly of reads was conducted using Trinity. Embryo-specific transcripts were extracted from the *de novo* assembly and added to the reference spiny mouse transcriptome. Transcription was first detected between the 2-cell and 4-cell stages for the majority of gene iso-groups (n=3,428), with fewer gene iso-groups first transcribed between the 4-cell and 8-cell stages (n=1,150). The pattern of gene expression in spiny mouse embryos during this period of development is more human-like than common mouse embryos. This is the first evidence the spiny mouse may provide a more suitable model of gene expression in early human embryonic development. The improved reference *Acomys cahirinus* transcriptome is publically accessible, further increasing the value of this tool for ongoing research. Further investigation into early development in the spiny mouse is warranted.**

The spiny mouse (*Acomys cahirinus*) is a small rodent native to regions of the Middle East and Africa (Nowak, 1999; Wilson and Reeder, 2005). It displays several unique physiological traits, including the capacity to regenerate skin without fibrotic scarring (Gawriluk et al., 2016; Seifert et al., 2012; Simkin et al., 2017) and a human-like endocrine profile, including cortisol as the primary glucocorticoid and production of DHEA (Quinn et al., 2013; Quinn et al., 2016). Unlike most eutherian mammals, the Egyptian spiny mouse ('common spiny mouse') has a menstrual cycle (Bellofiore et al., 2017). It is the only known species of rodent that menstruates and there are important differences in early embryonic development and implantation in menstruating species compared to those with an oestrus cycle such as mice, rats, cows, sheep and pigs (Brevini et al., 2006; Brosens et al., 2009; Emera et al., 2012; Graf et al., 2014; Memili & First, 2000; Niakan et al., 2012; Telford et al., 1990). Differences such as the polarity of apical attachment and cellular communication between the embryo and the endometrium can be identified before an embryo has implanted, and we may better understand the underlying mechanisms determining pregnancy success or failure by using a menstruating mammal to model human embryonic development in place of the mouse (*Mus musculus*) (Aplin & Ruane 2017; Brosens et al., 2009; Wang & Dey, 2006; Whitby et al., 2017). Little is known about early development in the spiny mouse, however preliminary evidence suggests it may overcome limitations of other species for modelling embryo development in humans.

[1] The Ritchie Centre, Hudson Institute of Medical Research, Melbourne, Australia.
[2] Department of Obstetrics and Gynaecology, Monash University, Melbourne, Australia.
[3] School of BioSciences, University of Melbourne, Melbourne, Australia.
[4] Education Program in Reproduction and Development, Monash University, Melbourne, Australia.
*Corresponding author: Jared Mamrot (email: jared.mamrot@hudson.org.au, twitter: @Jared_Mamrot)

One publication exists on spiny mouse embryo development in which the authors established methods for producing and culturing spiny mouse embryos *in vitro* (Pasco et al., 2012). One of the key challenges identified in this study was the presence of a '4-cell block', with embryos unable to develop past 4-cells when cultured outside of the reproductive tract. Embryos developed *in vivo* obtained at the 8-cell stage are able to be cultured successfully *in vitro* through to the implantation stage, however the timing of the cell block is an example of differences between *Mus musculus* and spiny mouse embryos at the molecular level (Taft, 2008). Mouse embryos exhibit a 2-cell block when exposed to inadequate culture conditions, whereas human embryos exhibit a 4- to 8-cell block (Braude et al., 1988; Goddard & Pratt, 1983). The cellular environment is a major influence on gene expression in preimplantation embryos (Gardner & Kelley, 2017; Mantikou et al., 2017); characterising gene expression profiles during embryogenesis may therefore help direct future research efforts to overcome the 4-cell block in the spiny mouse and promote its use as a model of human embryo development.

Embryogenesis is a complex process regulated by diverse, interdependent physiological mechanisms. Successful development from a single cell (zygote) to live offspring requires coordinated changes in cell cycle, chromatin state, DNA methylation and genome conformation. Cellular machinery for transcription and translation must be successfully assembled, and transcription of the incipient genome must take place. Failure to successfully attain any of these developmental milestones results in death of the organism. The first major developmental transition in eukaryotic embryos is the maternal-to-zygotic transition (MZT), which involves clearance of maternally-inherited transcripts and transcription of the newly formed embryonic genome (the embryonic genome activation, 'EGA') (Ivanova et al., 2017; Schier, 2007; Tadros & Lipshitz, 2009). Next Generation Sequencing (NGS) can be used to comprehensively characterise this event. In mammals the MZT typically occurs between the 1-cell stage and 16-cell stages of development, however the timing and pattern of embryonic gene expression is species-specific (Tadros & Lipshitz, 2009). In mice, the MZT occurs predominantly between the 1-cell and 4-cell stages, with the EGA beginning at the 2-cell stage (Flach et al., 1982; Wang and Dey, 2006). In comparison, in human embryos the EGA begins at the 4- to 8-cell stage (Braude et al., 1988; Tesarik et al., 1988). The timing of these events coincides with the timing of the 'cell block' previously described in these species (Braude et al., 1988; Goddard & Pratt, 1983).

Recent studies on human embryos have identified ~150 genes upregulated from the oocyte to 1-cell stage, followed by ~1,000 genes upregulated from the 2-cell to 4-cell stage (Xue et al., 2013; Yan et al., 2013), and >2,500 genes first transcribed between the 4-cell and 8-cell stage. The specific genes activated during each stage of the EGA have been shown to differ significantly between mice and humans, with reports of only ~40% concurrence between these two species (Heyn et al., 2014; Xie et al., 2010). Despite this, expression of specific genes driving the EGA are similar between humans and mice, and the overall pattern of transcription follows a similar pattern in mammals such as the cow, sheep, rabbit and other primates, occurring in 'waves' with different genes transcribed at different timepoints (Dobson et al., 2004; Taylor et al., 1997; Tesarı´k et al., 1987; Vassena et al., 2011). Although the pattern of EGA in rodents is similar to humans, conspicuous differences exist (Christians et al., 1994; Crosby et al., 1988; De Sousa et al., 1998; Frei et al., 1989; Schramm and Bavister, 1999; Telford et al., 1990) and the search for a more suitable model continues.

The aims of this study were to characterise early embryonic gene expression, identify the period of EGA, and to compare the pattern of global gene expression in the spiny mouse to existing datasets from human and mouse embryos. We hypothesise the EGA in the spiny mouse embryo will more closely reflect the EGA in human embryos than mouse embryos.

# METHODS

## Sample preparation and RNA sequencing

Embryos were collected from female spiny mice (n=12) in accordance with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes with approval from the Monash Medical Centre Animal Ethics Committee. Female dams were staged from delivery of their previous litter (spiny mice conceive their next litter approximately 12h postpartum) and culled at specific time-points for embryo retrieval at the required stage: 2-cell at 48h postpartum (n=4), 4-cell at 52h postpartum ('early' 4-cell; n=2) or at 68h postpartum ('late 4-cell'; n=2), and 8-cell at 72h postpartum (n=4). Embryos were flushed from the excised reproductive tract using warmed G-MOPS PLUS handling medium containing 5 mg/ml human serum albumin (Vitrolife, Göteborg, Sweden), washed through warmed sterile $Ca^{2+}$/$Mg^{2+}$-free PBS three times using sterile pulled glass pipettes, and grouped into biological replicates (n=4 for each stage: 12 samples total). Embryos were snap frozen using liquid nitrogen in a minimal volume of cell lysis solution (~1µl) comprised of lysis buffer, dithiothreitol (DTT) and RNase inhibitors per NuGEN SoLo RNA-Seq kit (NuGEN Technologies, Inc; San Carlos, CA, USA). To reduce the impact of embryo collection and freezing on gene transcription this process was conducted as quickly as possible: embryos were snap-frozen in lysis solution using liquid nitrogen and stored at -80°C in less than 5 minutes post-mortem.

To aid lysis, two freeze-thaw cycles were conducted on a slurry of dry ice and ethanol. Samples were processed per the Nugen SoLo protocol (version M01406v3; NuGEN). After ligation of cDNA, qPCR was performed on all samples to determine the number of amplification cycles required to ensure that amplification was in the linear range. Based on these results, each sample was amplified using 24 cycles to obtain a suitable amount of cDNA for sequencing. Final libraries were quantitated by Qubit and size profile determined by the Agilent Bioanalyzer.

Custom 'AnyDeplete' rRNA depletion probes were designed and produced by NuGEN Technologies, Inc (San Carlos, CA, USA) using rRNA sequences from the spiny mouse transcriptome (Mamrot et al., 2017). Prior to use, efficacy and off-target effects of the rRNA depletion probes were examined *in silico* by NuGEN. Samples were loaded using c-Bot (200pM per library pool) and run on 2 lanes of an Illumina HiSeq 3000 8-lane flow-cell. PhiX spike-in was not used directly due to incompatibility with the custom rRNA depletion probes, however it was incorporated into other lanes of the same run. RNA-Seq data (100bp, paired-end reads) were uploaded to the NCBI under Bioproject PRJNA436818 (SRA : SRP133894).

The quality of RNA-Seq reads was assessed using FastQC v0.11.6 (https://github.com/s-andrews/FastQC; 50f0c26), with MultiQC v1.4 (https://github.com/ewels/MultiQC; baefc2e) report available from Github (https://github.com/jpmam1) (Ewels et al., 2016). Adapter sequences were trimmed from the reads using trim-galore v0.4.2 (https://github.com/FelixKrueger/TrimGalore; d6b586e), implementing cutadapt v1.12 (https://github.com/marcelm/cutadapt; 98f0e2f). Reads with a quality scores lower than 20 and read pairs in which either forward or reverse reads were trimmed to fewer than 35 nucleotides were discarded. Further trimming was conducted using Trimmomatic v0.36 (http://www.usadellab.org/cms/index.php?page=trimmomatic) with settings "`LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 AVGQUAL:25 MINLEN:35`" (Bolger et al., 2014). Nucleotides with quality scores lower than 3 were trimmed from the 3' and 5' read ends. Reads with an average quality score lower than 25 or with a length of fewer than 35 nucleotides after trimming were removed. Error correction of trimmed reads was performed using Rcorrector v1.0.2 (https://github.com/mourisl/Rcorrector; 144602f) (Song & Florea, 2015). FastQC was used to assess the improvement in read quality after trimming adapter removal; MultiQC report is available from Github (https://github.com/jpmam1).

## *De novo* transcriptome assembly and read alignment

Error corrected reads were assembled using Trinity v2.4.0 (https://github.com/trinityrnaseq/trinityrnaseq; 1603d80) with settings "`--max_memory 400G, --CPU 32 and --full_cleanup`" (Haas et al., 2013). Assembly statistics were computed using the TrinityStats.pl script from the Trinity package and are provided in Table S1. All reads were aligned to the assembled 'embryo' transcriptome using Bowtie2 v2.2.5 (https://github.com/BenLangmead/bowtie2; e718c6f) with settings: "`--end-to-end, --score-min L,-0.1,-0.1, --no-mixed, --no-discordant, -k 100, -X 1000, --time, -p 24`" (Langmead & Salzberg, 2012).

Read-supported contigs were identified within the 'embryo' transcriptome assembly using samtools "idxstats" v1.5. (https://github.com/samtools/samtools; f510fb1) (Li et al., 2009). Read support was defined as >=1 reads aligned. Read-supported contigs from the embryo-specific assembly were added to the reference spiny mouse transcriptome assembly previously described by Mamrot et al. (2017) and samples were aligned to this 'updated' transcriptome using Bowtie2 with settings described above.

Trinity contigs were aligned to the UniProtKB/SwissProt protein sequence database (ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/complete/uniprot_sprot.fasta.gz accessed 14th October 2017) using BLASTx v2.5.0+ (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.5.0/) (Altschul et al., 1997). Confident BLAST hits were retained, transcripts were annotated using the single-best hit based on e-value, and Gene Ontology (GO) terms were obtained for further analysis. Transcript annotations are included (Supplementary Table 4 and from https://doi.org/10.5281/zenodo.1308820). Trinity-normalized reads were aligned to the NCBI nr protein database (ftp://ftp.ncbi.nih.gov/blast/db/FASTA/nr.gz accessed 10th February 2018) using DIAMOND v0.9.17 blastx (Buchfink et al., 2015) with taxonomic and functional annotation of reads aligning to eukaryotic and prokaryotic lineages conducted using MEGAN6 Community Edition v6.10.10 (https://ab.inf.uni-tuebingen.de/software/megan6) (Huson et al., 2016). MEGAN6 files were accessed using MeganServer v1.0.1 (https://ab.inf.uni-tuebingen.de/software/meganserver) (Beier et al., 2017). Reads were also aligned to mouse and human RefSeq rRNA sequences (accessions: NR_003279.1, NR_003278.3, NR_003280.2, NR_046144.1, NR_003285.2, NR_003287.2, NR_003286.2, X71802.1) and mouse tRNAs within the GtRNAdb database (http://gtrnadb.ucsc.edu/genomes/eukaryota/Mmusc10/mm10-tRNAs.fa) (Chan & Lowe, 2016).

## Transcript clustering and differential gene iso-group expression

Read alignments to the updated transcriptome generated using Bowtie2 were clustered with Corset v1.0.7 (https://github.com/Oshlack/Corset; cf4d4fb) to reduce the impact of redundant transcripts and

transcript isoforms when assessing gene expression (Davidson and Oshlack, 2015). Variance in RNA-Seq data was explored using Varistran v1.0.3 (https://github.com/MonashBioinformaticsPlatform/varistran; ff90258), which implements Anscombe's variance stabilizing transformation (1948) to equalize noise across all samples before assessing gene expression levels (Harrison, 2017). Differential gene expression was explored using the Degust web application (http://degust.erc.monash.edu/). Further investigation was conducted using EdgeR v3.20.8 (https://bioconductor.org/packages/release/bioc/html/edgeR.html) (Robinson et al., 2010). Correlations were calculated using Corrplot v0.83 (https://github.com/taiyun/corrplot; d7ba847) (Wei & Simko, 2017). Confidence bounds for effect sizes were calculated using TopConfects v1.0.1 (https://github.com/pfh/topconfects; 43cd006) (Harrison, 2018).

### Profiling gene expression during the EGA

Gene expression data were accessed for mouse and human embryos from the NCBI Gene Expression Omnibus (GEO) project GSE44183 (accessed 22/02/2018) (Xue et al., 2013). This dataset contains both human and mouse embryos collected at the same developmental stages (mouse: 3X2-cell, 3X4-cell, 3X8-cell; human: 3X2-cell, 4X4-cell and 10X8-cell). Gene expression profiles were generated from Log2 fold changes in Fragments Per Kilobase of transcript per Million mapped reads (FPKM) extracted from expression matrices provided by the authors (ftp://ftp.ncbi.nlm.nih.gov/geo/series/GSE44nnn/GSE44183/suppl/). All figures were produced using R software v3.4.0 and GraphPad Prism 7.

## RESULTS

### RNA sequencing and quality control

In total, 701.9 million reads passed filtering across 12 samples (Table 1) with a relatively high proportion of >Q30 reads (95.2%). The error rate was 0.2% (expected <0.5%) and phasing/prephasing was 0.13/0.08 (expected <0.4/<0.2), indicating high-quality sequencing with minimal technical errors. Read error correction resulted in 266 million repairs (~0.1% of all nucleotides). Quality metrics obtained using FastQC before and after read processing are available from Github:
https://rawgit.com/jpmam1/multiQC_reports/master/pre-trimming_multiqc_report.html
https://rawgit.com/jpmam1/multiQC_reports/master/post-trimming_multiqc_report.html.

### *De novo* transcriptome assembly and read alignment

All trimmed and error-corrected reads were assembled into an 'embryo' transcriptome (assembly metrics: Table S1) to detect transcripts specific to early development not present in our reference spiny mouse transcriptome. The proportion of reads mapping to this embryo-specific transcriptome, the number of unique reads per sample, and proportion of reads from each sample aligned to human/mouse rRNA sequences are shown in Figure 1 (no reads aligned to the tRNA database). Transcripts from the embryo assembly were aligned to the UniProtKB / SwissProt protein database using BLASTx: ~70% of transcripts aligned to *Mus musculus*, *Homo sapiens* and *Rattus norvegicus*, and ~30% aligned to other eukaryotic and prokaryotic taxa (interactive summary: https://public.flourish.studio/visualisation/20088/).

### Clustering and differential gene iso-group expression

Transcripts from the *de novo* assembly (n=595,435) were clustered together based on read mapping to form 309,543 representative gene clusters (referred to from here on as iso-groups). This clustering facilitated use of gene-level methods for quantification and analysis. Average read count for each sample library and hierarchical clustering of samples based on average iso-group abundance in counts-per-million (cpm) are shown in Figure 2A (further sample correlations are shown in Figure S3). Hierarchical clustering revealed clear differentiation between 2-cell embryos and the 4-cell/8-cell embryos, with less clear differentiation between 4-cell and 8-cell embryos (Figure 2B).

Application of Anscombe's variance stabilizing transformation tempered dispersion across all samples (average dispersion = 0.0784). Library sizes before and after 'trimmed mean of M' (TMM) normalization (Robinson & Oshlack, 2010) are listed in Supplementary Table 2. Read counts were clustered in two dimensions to examine group differences in iso-group abundance. Multi-dimensional scaling (MDS) analysis suggests two of the samples ("2cell_C" and "8cell_A") have atypical profiles compared to the other samples (Figure 3).

Further investigation revealed significant differences between samples '2-cell_C' and '8-cell_A' compared to the other 10 samples (Figure 4). BLASTx alignment of differentially expressed (DE) iso-groups to the UniProtKB/SwissProt database revealed evidence of contamination, with >800 bacteria-associated iso-groups highly expressed in these two samples and no expression detected in the other samples (Table S3). Metatranscriptomic analysis of read alignments to the NCBI nr database confirmed significant prokaryotic contamination in samples '2-cell_C' and '8-cell_A' (Figure S1). These samples were not able to be salvaged due to the level of contamination and were excluded from further analysis, reducing statistical power (Figure S2). With these two samples removed the expression profiles of spiny mouse embryos are comparable to other mammals at this stage of development (Figure 5).

Table 1: Summary of paired-end (100bp) Illumina sequencing output for each sample

| ULN | Sample name | Biosample | I7 Index | Mean Library Size (bp)* | Identifier sequence | Reads Passed Filter (Million) |
|---|---|---|---|---|---|---|
| 17---04796 | Sample1 (“2cell_A”) | SRR6804613 | C02 | 338 | GACTACGA | 58.1 |
| 17---04797 | Sample2 (“2cell_B”) | SRR6804612 | D02 | 318 | ACTCCTAC | 62.1 |
| 17---04798 | Sample3 (“2cell_C”) | SRR6804607 | E02 | 335 | CTTCCTTC | 61.9 |
| 17---04799 | Sample4 (“2cell_D”) | SRR6804606 | F02 | 338 | ACCATCCT | 59.2 |
| 17---04800 | Sample5 (“4cell_A”) | SRR6804609 | G02 | 338 | CGTCCATT | 59.6 |
| 17---04801 | Sample6 (“4cell_B”) | SRR6804608 | H02 | 354 | AACTTGCC | 57.2 |
| 17---04802 | Sample7 (“4cell_C”) | SRR6804611 | A03 | 342 | GTACACCT | 53.2 |
| 17---04803 | Sample8 (“4cell_D”) | SRR6804610 | B03 | 343 | ACGAGAAC | 54.5 |
| 17---04804 | Sample9 (“8cell_A”) | SRR6804617 | C03 | 323 | CGACCTAA | 64.4 |
| 17---04805 | Sample10 (“8cell_B”) | SRR6804616 | D03 | 321 | TACATCGG | 55.9 |
| 17---04806 | Sample11 (“8cell_C”) | SRR6804615 | E03 | 327 | ATCGTCTC | 56.8 |
| 17---04807 | Sample12 (“8cell_D”) | SRR6804614 | F03 | 365 | CCAACACT | 59.8 |
| | | | | | Total reads | 701.9 |

* The expected sample library size range was ~320-360 bp.

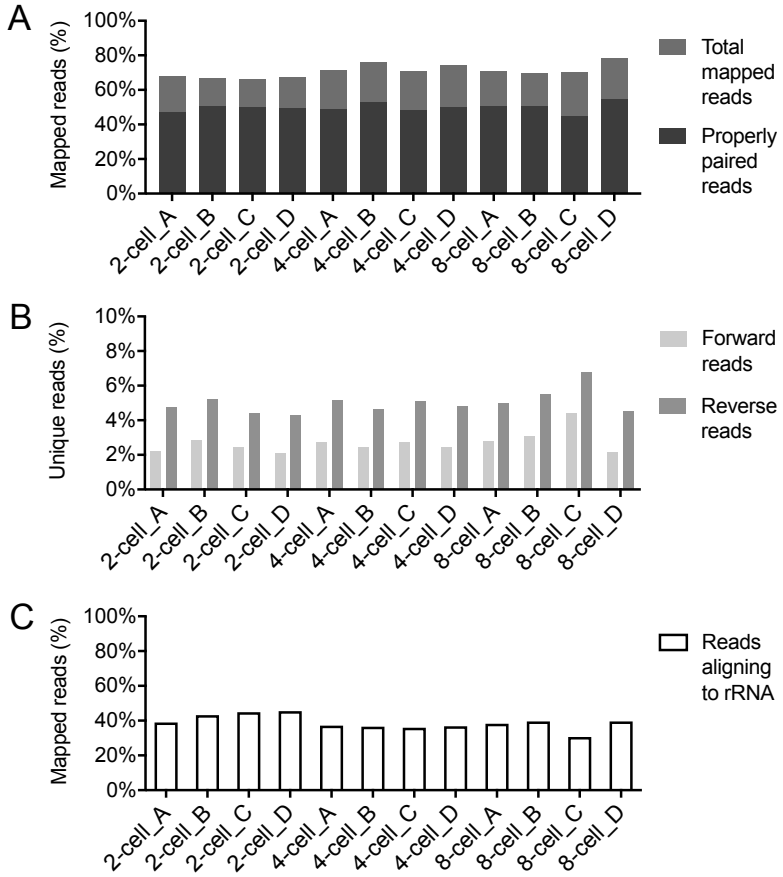



Figure 1: (A) Proportion of reads mapping to the 'embryo' spiny mouse transcriptome assembly using the conservative settings specified in the methods. “Properly paired reads” both align to the same transcript, “Total mapped reads” represent either forward or reverse reads mapped to a transcript. (B) The proportion of unique reads per sample, and (C) reads mapping to human / mouse RefSeq rRNA sequences.

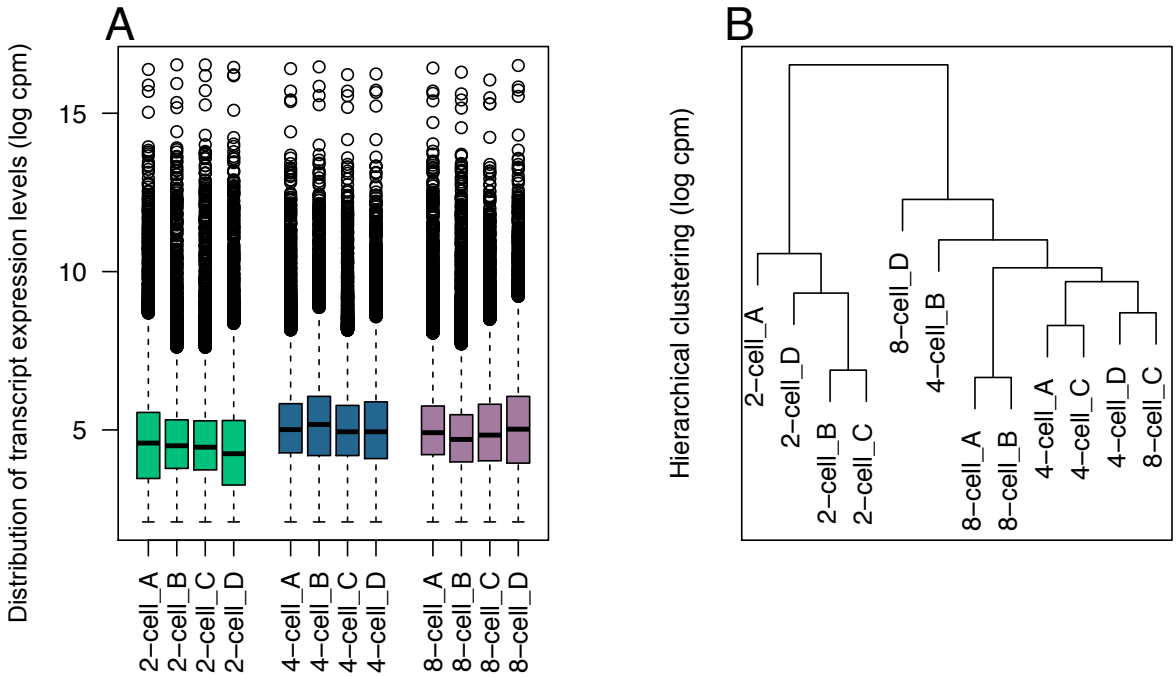



Figure 2: (A) Average read count per sample and (B) hierarchical clustering of samples based on iso-group abundance in each library (log cpm = counts per million after log transformation). Ambiguous clustering of samples indicates an underlying bias in overall iso-group expression levels.

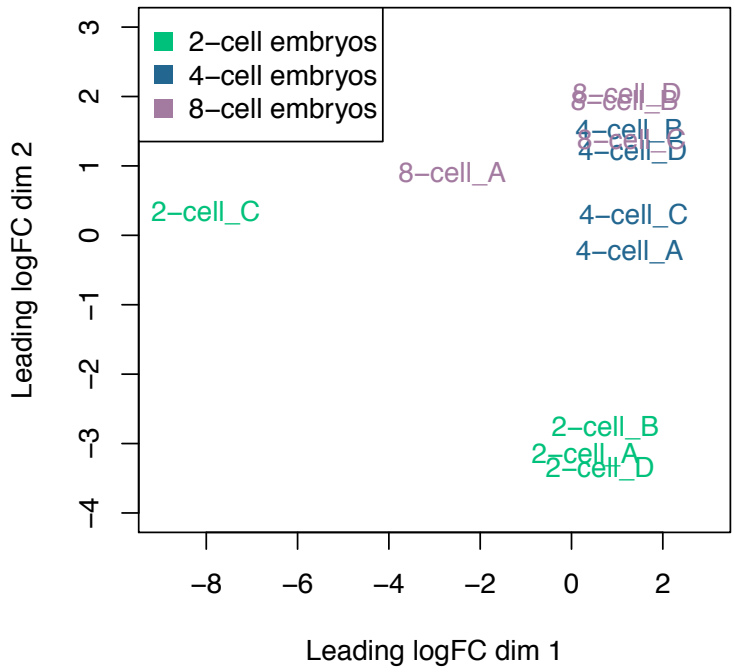



Figure 3: MDS plot illustrating differences in average iso-group expression levels between samples for the top 500 differentially-expressed iso-groups (n=12).

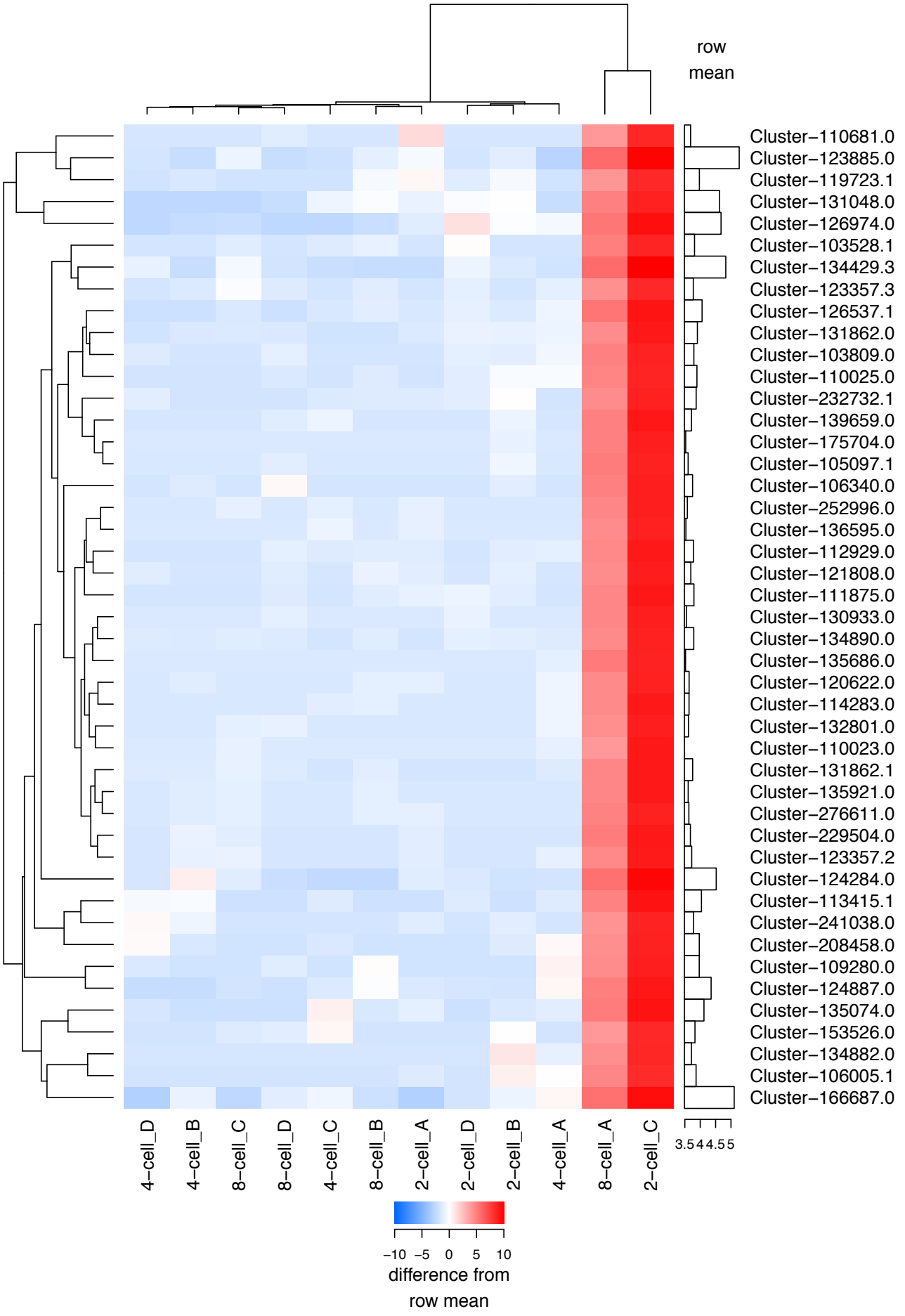



Figure 4: Heatmap of top 50 differentially expressed ‘Clusters’ (iso-groups) (n=12). Expression in samples 2-cell_C and 8-cell_A is highly abnormal due to the presence of prokaryotic contamination.

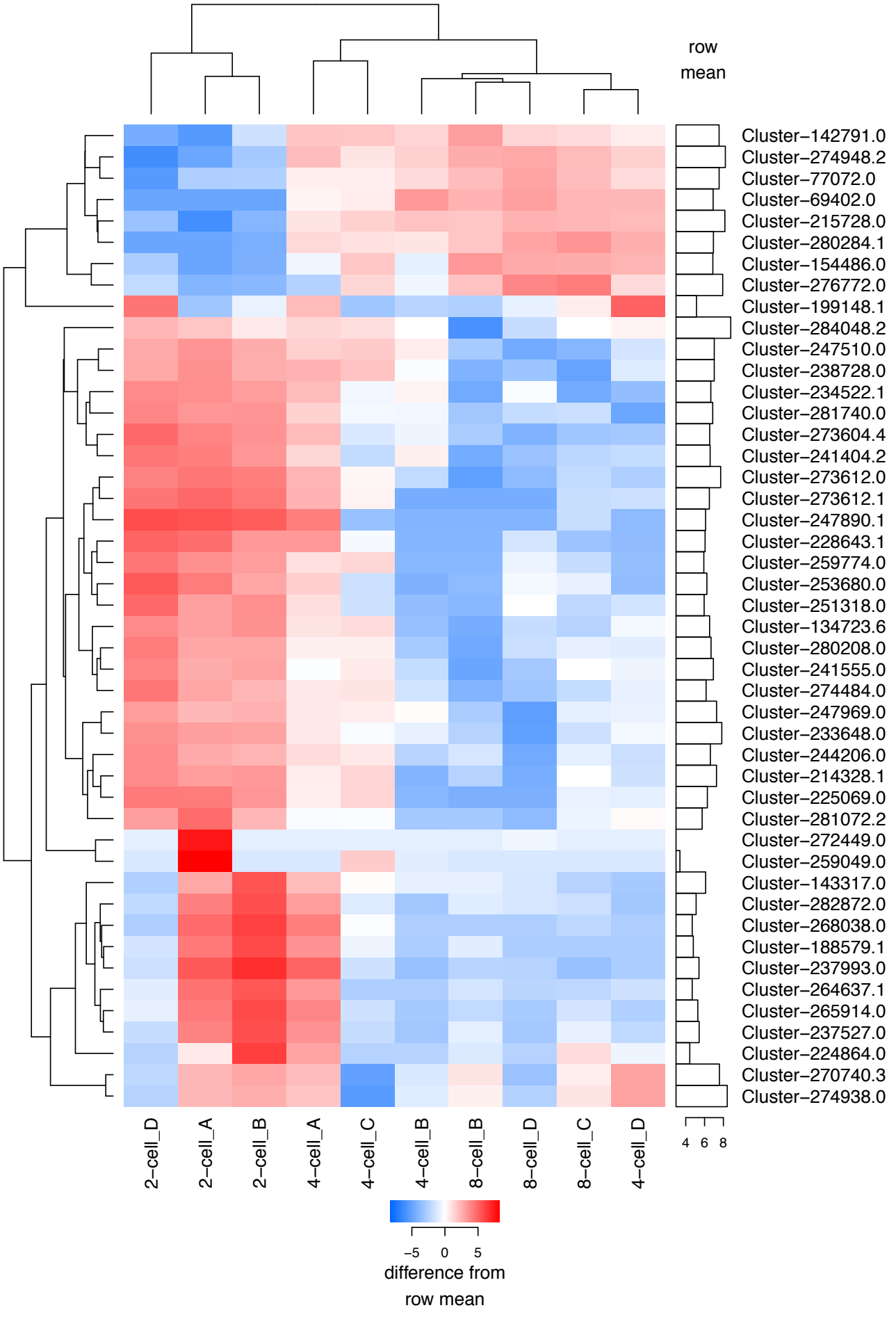



Figure 5: Heatmap of the top 50 differentially expressed iso-groups with samples 2-cell_C and 8-cell_A excluded from analysis (n=10). Gene cluster iso-group IDs: Table S4; https://doi.org/10.5281/zenodo.1308820

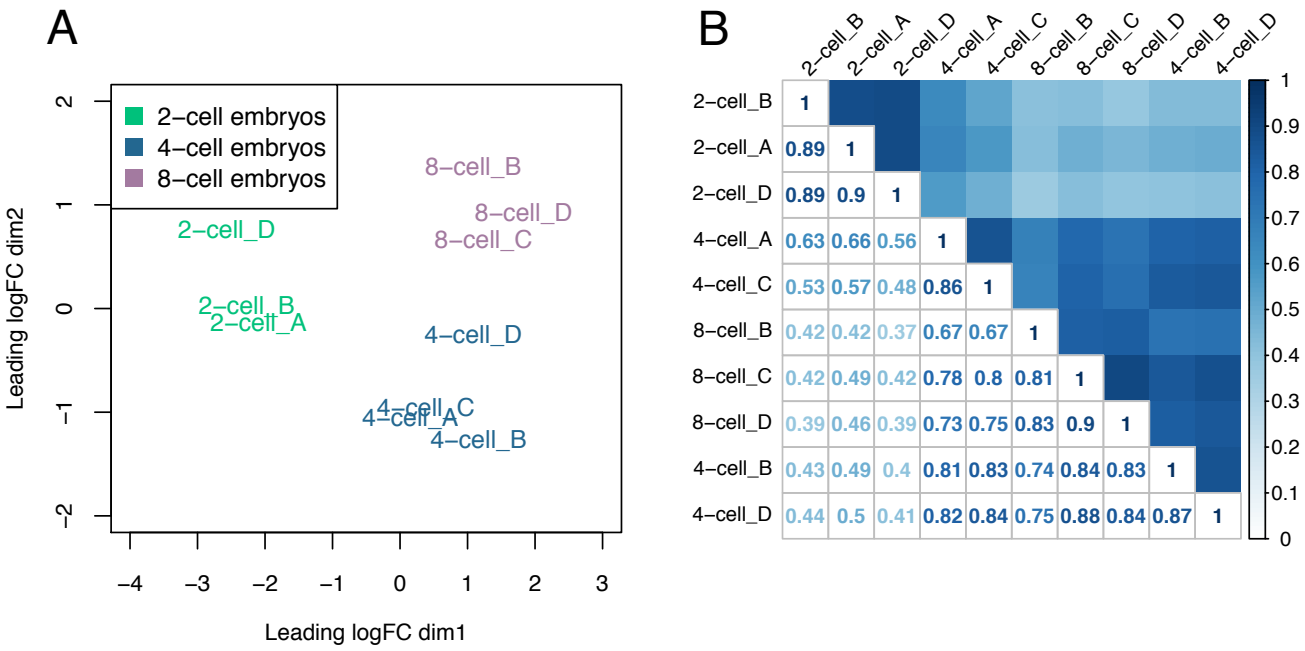



Figure 6: (A) MDS plot for top 1000 iso-groups in remaining uncontaminated samples after reanalysis (n=10), and (B) corresponding Spearman rank correlations of iso-group abundance (n=10). Samples 4-cell_A and 4-cell_C were collected at 52h post-partum ("4-cell early"), and samples 4-cell_B and 4-cell_D were collected at 68h postpartum ("4-cell late"). These samples cluster in dimension 1, however there is overlap in dimension 2.

Alignments from the 10 uncontaminated samples to the 'embryo' transcriptome were re-examined. All transcripts with >=1 reads aligned were extracted (54,660 read-supported contigs in total; 441.23 Mb of sequence data) and added to the 'reference' transcriptome assembled by Mamrot et al. (2017). Alignment, clustering and iso-group expression analysis were performed against the 'updated' reference transcriptome. Transcripts from the updated assembly (n=2,274,638) were clustered based on read mapping using Corset; the number of clusters (iso-groups) produced using the updated reference assembly (n=253,449) was fewer than the number produced using the *de novo* 'embryo' assembly (n=309,543). Exclusion of contaminated samples resulted in stronger correlations within developmental stages (Figure 5, Figures S4 & S5) and increased delineation between developmental stages, with 'early' 4-cell embryo samples (4-cell_A and 4_cell_C) clustering more closely to the 2-cell embryos, and the 'late' 4-cell samples (4-cell_B and 4_cell_D) clustering more closely to the 8-cell embryos (Figure 6).

Fit of the negative binomial distribution to transcript counts (Figures S6 & S7), biological coefficient of variation / quasi-likelihood dispersion (Figure S8), and mean-difference of each sample against combined samples (Figure S9) support the use of quasi-likelihood F-tests to determine differential expression. In total, differentially expression was detected in 3,428 gene iso-groups between the 2-cell and 4-cell stages and 1,150 gene iso-groups between the 4-cell and 8-cell stages of embryo development in the spiny mouse (Figures 7 and 8).

Differential expression is first detected in the majority of embryonic iso-groups at the 2-cell to 4-cell stage (Figure 7). Effect sizes and confidence intervals were calculated for all DE iso-groups revealing relatively large differences between developmental stages. Iso-groups with the largest effect sizes were predominantly upgregulated at the 2- to 4-cell stage with a more even ratio of upregulated / downregulated gene iso-groups at the 4- to 8-cell stage (Figure 8). The ratio of total upregulated and downregulated DE iso-groups was similar between developmental stages (Figure 9). This pattern of genome activation in spiny mouse embryos resembles that of the mouse embryo, however the expression of specific iso-groups corresponding to genes such as HSP70 (Figure 10F) and the overall profile of transcript expression (Figure 11) share commonalities with the EGA in humans.

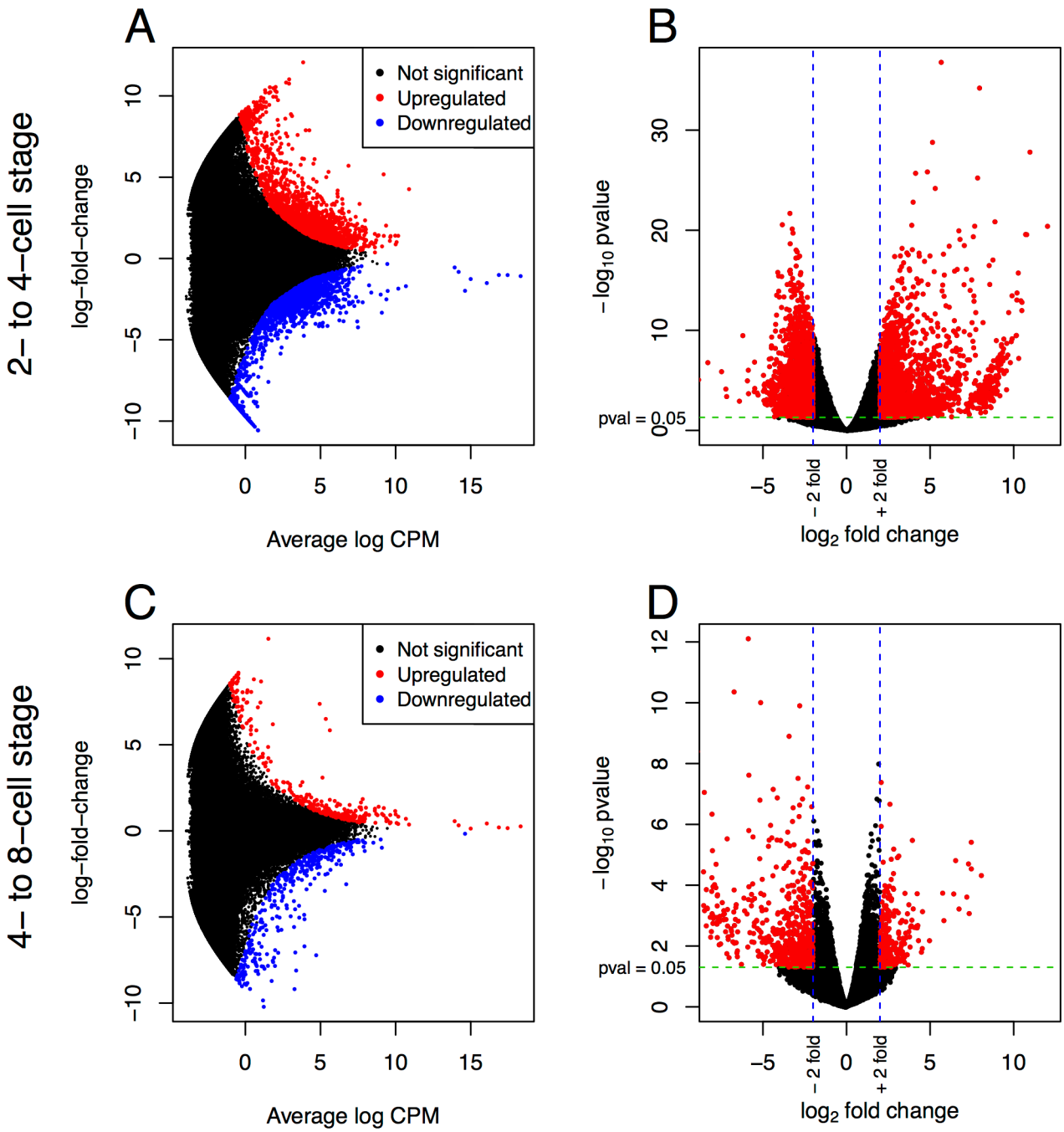



Figure 7: Differentially expressed iso-groups between the 2-cell and 4-cell stages of development (A & B) and the 4-cell and 8-cell stages of development (C & D). Coloured dots represent individual differentially expressed iso-groups. Smear plots: FDR<0.05 (A & C). Volcano plots: p-value <0.05 (B & D).

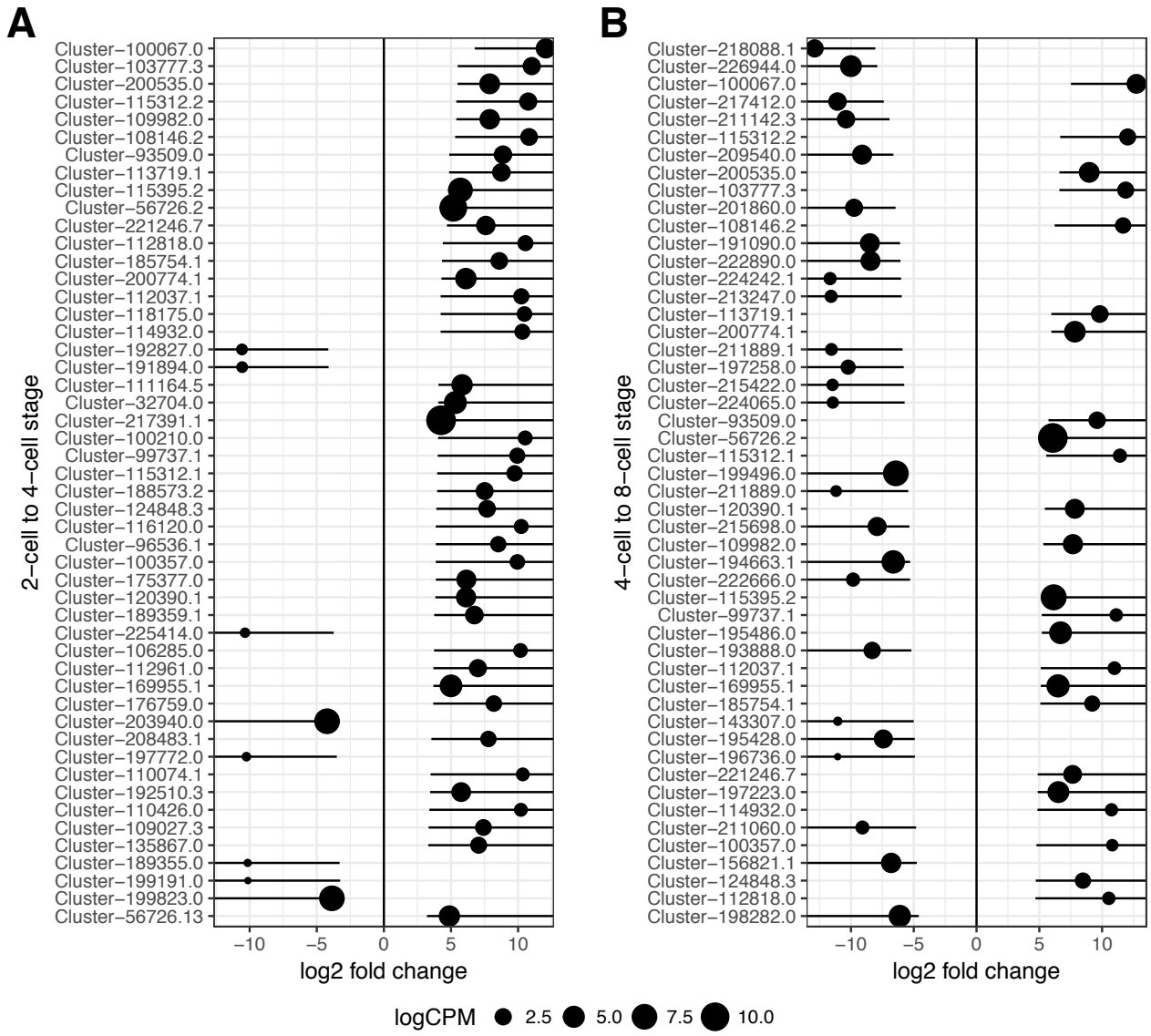
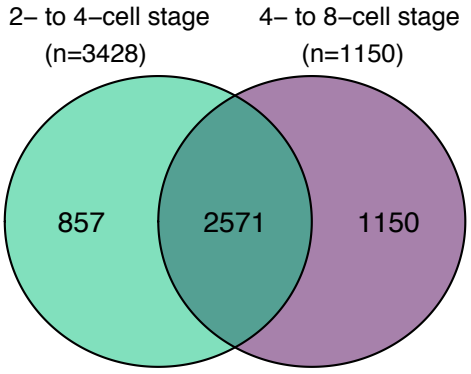



Figure 8: Iso-group clusters with highest effect sizes (including confidence bounds) for (A) the 2-cell to 4-cell stage and (B) the 4-cell to 8-cell stage; logCPM = log2(counts-per-million).




| | 2- to 4-cell | 4- to 8-cell |
|---|---|---|
| Upregulated | 1757 | 555 |
| Downregulated | 1671 | 595 |
| Total | 3248 | 1150 |

Figure 9: Number of differentially expressed iso-groups at each developmental stage (FDR <0.05). The overlap represents iso-groups common to both stages, but first transcribed at the 2- to 4-cell stage. The total number of differentially-expressed iso-groups is also quantified as upregulated and downregulated iso-groups.

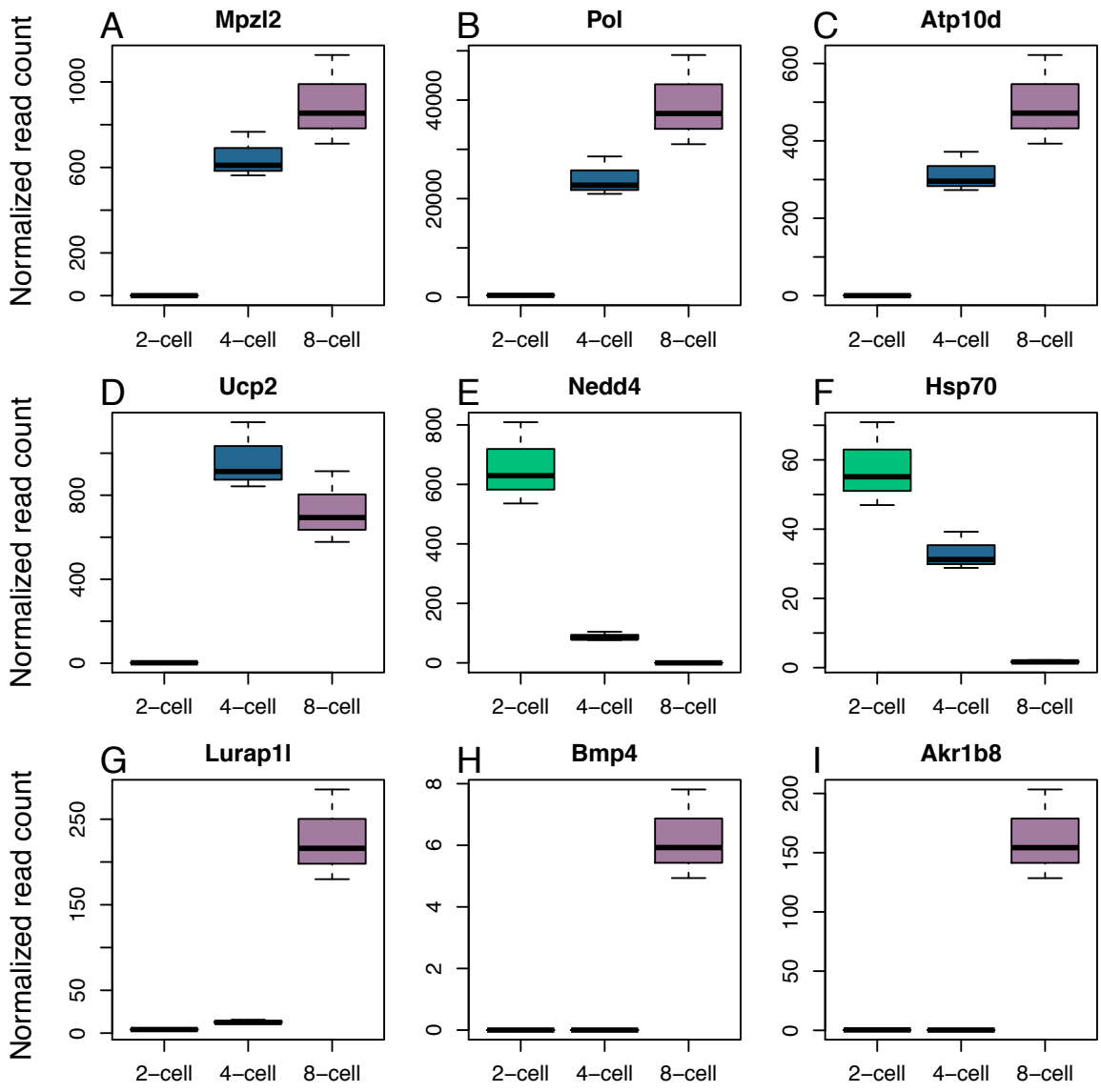



Figure 10: EdgeR negative binomial-fitted iso-group expression profiles for genes-of-interest identified by TopConfects. Expression patterns include increasing expression from the 2-cell to 8-cell stage (A, B & C), high-to-low expression (D, E & F) and expression initiated at the 4-cell to 8-cell stage (G, H & I).

## Profiling gene expression during the EGA

Known differences in specific genes activated during the EGA in mice and humans limit direct comparisons between these species and the spiny mouse, however analysis of overall patterns of transcription are used here to approximate similarities / dissimilarities between species. Profiles depicted in Figure 10 illustrate different patterns of expression seen in transcripts of interest identified by the effect size analysis. Except for Hsp70, these transcripts are not known to play an important role in the EGA in mammals; they are presented to illustrate representative expression patterns seen during the EGA (increasingly high expression, high-to-low expression, and delayed expression until the 4- to 8-cell stage). To determine whether the EGA in the spiny mouse embryo more closely reflects the EGA in human or mouse embryos, profiles were generated for mouse, spiny mouse and human embryos illustrating the pattern of expression changes between the 2-, 4- and 8-cell stages for each species (Figure 11). Expression changes are less extreme in the mouse embryo, and a smaller number of transcripts are differentially expressed between the 4- to 8-cell stage compared to spiny mouse and human embryos.

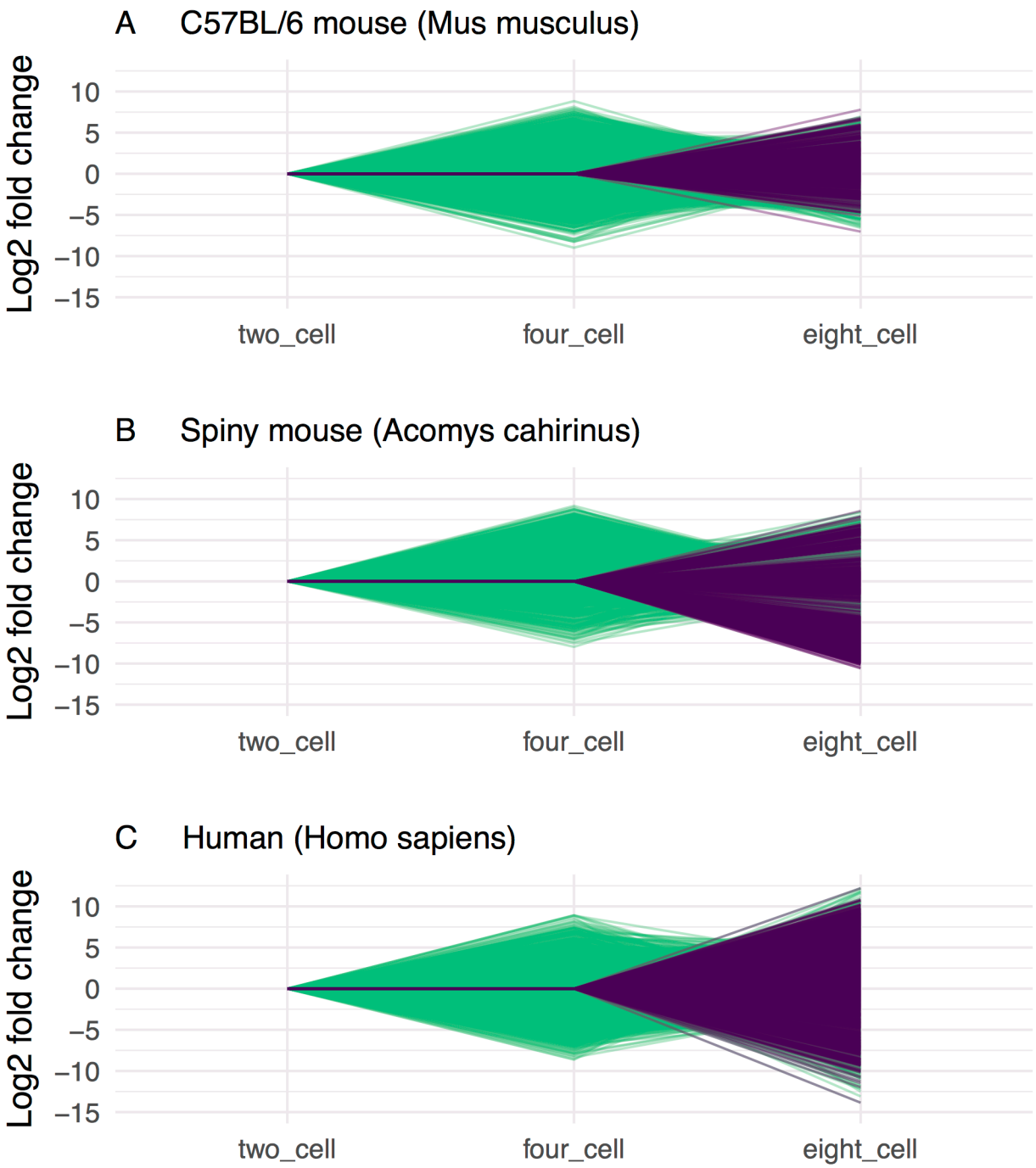



Figure 11: Transcript expression profiles for (A) the C57/BL6 'common' mouse, (B) the spiny mouse and (C) for human embryos during the EGA. Transcripts in which expression is first detected between the 2- to 4-cell stages are represented by green lines. Transcripts in which expression is first detected between the 4- to 8-cell stage are represented by purple lines. Fewer transcripts are first expressed between the 4- to 8-cell stage in mouse embryos compared to spiny mouse and human embryos. Differences in expression of transcripts activated at the 4- to 8-cell stage in mice are smaller than spiny mouse and human embryos, displaying less extreme log2 fold changes.

# DISCUSSION

Here we show that the embryonic genome activation (EGA) begins between the 2-cell and 4-cell stages of embryo development in the spiny mouse. This time-point had the greatest number of differentially expressed (DE) transcripts and transcripts for several genes reported to drive the EGA in other mammalian species were identified at this developmental stage for the first time, such as Hsp70 (Bensaude et al., 1983: Figure 10F), Eif4e (Yartseva & Giraldez, 2015), Eif1a (Lindeberg et al., 2004) and Elavl1 (Bell et al., 2008) (Figure S10). The pattern of transcription was similar to other mammals in which the EGA has been characterized (Svoboda, 2017), with massive changes in gene expression occurring within a relatively short time frame. Characteristics used to delineate between the common mouse, spiny mouse and human embryo include the expression of specific genes / iso-groups, the timing of EGA initiation and the 'burst' of transcription required for continued development (Richter & Sonenberg, 2005). By these criteria, findings from this study suggest the spiny mouse is a closer model of human embryonic gene expression than the common mouse. This is the first assessment of the spiny mouse for this purpose and these findings warrant further investigation.

An unexpected outcome of this study was sample contamination. Embryo collection was conducted very quickly to minimise the effect of stress on gene transcription and the increased speed of embryo collection resulted in two of the samples becoming compromised. Initial transcript expression analysis conducted using the DEGUST web platform (http://degust.erc.monash.edu) revealed this unexpected technical complication. These samples were unable to be salvaged as only ~30% of the reads they contained aligned to mammalian proteins in the NCBI nr database (Figure S1). This contamination limited our ability to use the *de novo* assembly as a reference for read alignment as a large proportion of the assembled transcripts were found to represent prokaryotic sequences rather than spiny mouse sequences. Use of this assembly would have resulted in erroneous quantification of transcript expression levels. Extracting read-supported embryo-specific transcripts from the 'embryo' assembly and adding them to the reference spiny mouse transcriptome (Mamrot et al., 2017) was a successful solution for avoiding transcripts derived from prokaryotic organisms. Several transcripts that are only expressed during early embryo development can now be found in the reference transcriptome, such as Oct3/4, Nanog, Oobox and H1foo. The updated assembly has been uploaded to a permanent data repository (https://doi.org/10.5281/zenodo.1188364) and is accessible via our BLAST search website: http://spinymouse.erc.monash.edu/sequenceserver/ ("Trinity_v2.3.2_plus_embryo-specific_transcripts") (Priyam et al., 2015). This approach significantly improved gene-level resolution and improved the reference transcriptome for future applications.

A downstream effect of sample contamination was reduced statistical power. Our preliminary power calculation predicted 4 samples per group would be required to accurately quantify differences in iso-group expression between developmental stages (Figure S2). Exclusion of two samples reduced our ability to resolve DE iso-groups, however the parameters used for the initial power calculation were found to be relatively conservative and the analysis was modified to mitigate against this confounding factor. Quasi-likelihood F-tests were used to establish differential expression (rather than likelihood ratio tests) to gain stricter error rate control by accounting for uncertainty in the original dispersion estimate (Chen et al., 2016). In addition, transcript expression was analysed at the 'gene' level to avoid potential biases previously reported in transcript-level analyses (Kanitz et al., 2015; Leshkowitz et al., 2016; Williams et al., 2017). This modified workflow was effective in identifying a large number of DE iso-groups, however the total number of DE iso-groups reported for each timepoint are likely to be underestimated. Using an unadjusted p-value (p=0.05) as a cutoff for statistical significance, rather than adjusting the p-value to reduce the false discovery rate (FDR=0.05), provides an indication of gene iso-groups that may have been detected as differentially expressed given full experimental power (Figures 7B & 7D). This suggests the number of iso-groups upregulated between the 2-cell and 4-cell stage and number of genes downregulated between the 4-cell and 8-cell stage are likely greater than reported.

The presence of rRNA reads was another unexpected outcome. Rather than sequence poly(A)+ RNA, we depleted rRNA using custom designed depletion probes manufactured by NuGEN (formerly known as Insert Dependent Adaptor Cleavage "InDA-C" probes) to obtain non-coding RNA transcripts and partially-degraded maternally-inherited transcripts in addition to mRNA (Bush et al., 2017; Schuierer et al., 2017). This approach was partially successful. Greater than 80% of total RNA is composed of rRNA in preimplantation embryos (Bush et al., 2017; O'Neil et al., 2013; Piko & Clegg, 1982), so levels detected in our samples (~30-40%: Figure 1) suggest the AnyDeplete rRNA probes worked, but were not fully effective. There are several potential explanations for this result; the most likely explanation is that our AnyDeplete probes were designed and tested using spiny mouse RNA-Seq derived transcripts whereas AnyDeplete probes are typically designed using a reference genome (a spiny mouse genome is not yet publicly available). The impact of rRNA levels on the ability to detect relative abundance of protein-coding RNA transcripts in preimplantation embryos is unknown.

Protein-coding transcripts known to regulate early development in mammalian embryos were detected at the 2- to 4-cell stage in the spiny mouse. These transcripts, including Yap1, RNA polymerase II, E3 ubiquitin-protein ligase, and the eukaryotic initiation factor family of transcripts have been implicated in the EGA in humans through various mechanisms of action (Ge, 2017; Svoboda 2017). One of the first proteins transcribed in mammalian embryos is the Heat Shock Protein 70kDa (known as Hsp70 / Hspa1a) (Bensaude et al., 1983). This protein performs several roles during the MZT, such as establishing chromatin structure, genome stability, and chaperoning O-linked glycosylated proteins into the cell nucleus (Abane & Mezger, 2010; Guinez et al., 2005; Nagaraj et al., 2017). In spiny mouse embryos Hsp70 expression is relatively high at the 2-cell stage followed by decreasing expression at the 4- and 8-cell stages. High expression of this gene during the first 'wave' of the EGA has been shown in many species, including the mouse, bovine and human embryo (Bettegowda et al., 2007; Christians et al., 1997; Lelièvre et al., 2017). Early transcription of HSP70 is crucial for successful cell cleavage and continued development, with compromised gene expression and protein levels correlated with embryo cell blocks. This pathway provides a potential target for understanding and overcoming the 4-cell block in the spiny mouse.

Direct comparison of DE transcripts between mouse, spiny mouse and human embryos at these stages of development was not conducted due to poor inter-species consensus reported by others (e.g. Heyn et al., 2014; Xie et al., 2010), however specific transcripts directly implicated in the EGA were investigated (Figure S10). Overall variation in EGA-related expression was found between the mouse, spiny mouse and human, with the results for these transcripts of interest (Eif4e, Elavl1, Pou5f1, Eif1a) representing typical inter-species differences. Although differences were identified in this study between the mouse, spiny mouse and human, further efforts to replicate and reproduce these results would increase the likelihood that these findings represent differences in the underlying mechanisms driving the EGA, rather than other confounding factors. A more robust inter-species comparison of the EGA is the overall changes in transcript expression patterns during these early developmental stages (Figure 11). This comparison revealed a closer relationship between spiny mouse and human embryos, compared to the common mouse, with a greater number of transcripts first expressed at the 4- to 8-cell stage and a larger range of expression changes during this period of development. These findings support use of the spiny mouse (*Acomys cahirinus*) as a model of the human EGA.

In conclusion, anatomy and physiology varies between all animal models of human reproduction and development. Primates are arguably the most accurate representation of human physiology, with similar anatomy and endocrine profiles, however ethical and logistical constraints limit their usefulness for basic research. Rodents offer an attractive alternative, as they have short breeding intervals and their anatomy and physiology has been comprehensively studied. Despite the advantages, translation of findings from mice to humans is not always successful, suggesting the common mouse may not be the best model for early human development. Conspicuously, the absence of a menstrual cycle in the common mouse is associated with key differences in how embryos are formed and develop. Here, we aimed to investigate the spiny mouse and assess its usefulness for modelling early human embryonic gene transcription. Methodological limitations impacted our ability to comprehensively address this aim, however the novel findings reported here support further investigation into other aspects of embryology in this species. Future directions for this work include further sequencing of spiny mouse embryos at the zygote, 16-cell stage, morula and blastocyst stages, and use of this RNA-Seq dataset to investigate the conditions required to overcome the 4-cell block in the spiny mouse embryo to facilitate further comparison of embryos developed *in vitro* and *in vivo* in this species.

## Acknowledgements

We acknowledge the support of the Victorian State Government Operational Infrastructure Scheme. We thank Vivien Vasic, Trevor Wilson and members of the MHTP Genomics platform for conducting the challenging library prep and RNA-Seq. Tony Papenfuss for providing access to WEHI facilities. David Powell for his assistance with the spiny mouse blast database website and for use of Monash Bioinformatics Platform resources. David Walker for his ongoing support. Ashleigh Clark and Nadia Bellofiore for maintaining the spiny mouse colony.

## Author contributions

JM and HD designed the study with advice from DKG and PTS. JM collected and prepared samples for sequencing, assembled and analysed the sequencing data, prepared figures and wrote the manuscript. All authors read and approved the final version of the manuscript.

## References

Abane, R., & Mezger, V. (2010). Roles of heat shock factors in gametogenesis and development. The FEBS journal, 277(20), 4150-4172.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of

protein database search programs. Nucleic Acids Research, 25(17), 3389-3402.

Anscombe, F. J. (1948). The transformation of Poisson, binomial and negative-binomial data. Biometrika, 35(3/4), 246-254.

Aplin, J. D., & Ruane, P. T. (2017). Embryo–epithelium interactions during implantation at a glance. Journal of Cell Science, 130(1), 15-22.

Beier, S., Tappu, R., & Huson, D. H. (2017). Functional Analysis in Metagenomics Using MEGAN 6. In Functional Metagenomics: Tools and Applications (pp. 65-74). Springer, Cham.

Bell, C. E., Calder, M. D., & Watson, A. J. (2008). Genomic RNA profiling and the programme controlling preimplantation mammalian development. Molecular Human Reproduction, 14(12), 691-701.

Bellofiore, N., Ellery, S. J., Mamrot, J., Walker, D. W., Temple-Smith, P., & Dickinson, H. (2017). First evidence of a menstruating rodent: the spiny mouse (Acomys cahirinus). American Journal of Obstetrics & Gynecology, 216(1), 40-e1.

Bensaude, O., Babinet, C., Morange, M., & Jacob, F. (1983). Heat shock proteins, first major products of zygotic gene activity in mouse embryo. Nature, 305(5932), 331.

Bettegowda, A., Lee, K. B., & Smith, G. W. (2007). Cytoplasmic and nuclear determinants of the maternal-to-embryonic transition. Reproduction, Fertility and Development, 20(1), 45-53.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114-2120.

Braude, P., Bolton, V., & Moore, S. (1988). Human gene expression first occurs between the four-and eight-cell stages of preimplantation development. Nature, 332(6163), 459.

Brevini, T. A., Cillo, F., Antonini, S., Tosetti, V., & Gandolfi, F. (2006). Temporal and spatial control of gene expression in early embryos of farm animals. Reproduction, Fertility and Development, 19(1), 35-42.

Brosens, J. J., Parker, M. G., McIndoe, A., Pijnenborg, R., & Brosens, I. A. (2009). A role for menstruation in preconditioning the uterus for successful pregnancy. American Journal of Obstetrics & Gynecology, 200(6), 615-e1.

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. Nature Methods, 12(1), 59.

Bush, S. J., McCulloch, M. E., Summers, K. M., Hume, D. A., & Clark, E. L. (2017). Integration of quantitated expression estimates from polyA-selected and rRNA-depleted RNA-seq libraries. BMC Bioinformatics, 18(1), 301.

Chan, P.P. & Lowe, T.M. (2016) GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Research, 44(Database issue): D184-D189.

Chen, Y., Lun, A. T., & Smyth, G. K. (2016). From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. F1000Research, 5.

Christians, E., Michel, E., Adenot, P., Mezger, V., Rallu, M., Morange, M., & Renard, J. P. (1997). Evidence for the involvement of mouse heat shock factor 1 in the atypical expression of the HSP70. 1 heat shock gene during mouse zygotic genome activation. Molecular and cellular biology, 17(2), 778-788.

Crosby, I. M., Gandolfi, F., & Moor, R. M. (1988). Control of protein synthesis during early cleavage of sheep embryos. Journal of reproduction and fertility, 82(2), 769-775.

Davidson, N. M., & Oshlack, A. (2014). Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. Genome biology, 15(7), 410.

De Sousa, P. A., Caveney, A., Westhusin, M. E., & Watson, A. J. (1998). Temporal patterns of embryonic gene expression and their dependence on oogenetic factors. Theriogenology, 49(1), 115-128.

Dobson, A. T., Raja, R., Abeyta, M. J., Taylor, T., Shen, S., Haqq, C., & Pera, R. A. R. (2004). The unique transcriptome through day 3 of human preimplantation development. Human molecular genetics, 13(14), 1461-1470.

Emera, D., Romero, R., & Wagner, G. (2012). The evolution of menstruation: A new model for genetic assimilation. Bioessays, 34(1), 26-35.

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics, 32(19), 3047-3048.

Flach, G., Johnson, M. H., Braude, P. R., Taylor, R. A., & Bolton, V. N. (1982). The transition from maternal to embryonic control in the 2-cell mouse embryo. The EMBO journal, 1(6), 681-686.

Frei, R. E., Schultz, G. A., & Church, R. B. (1989). Qualitative and quantitative changes in protein synthesis occur at the 8–16-cell stage of embryogenesis in the cow. Journal of reproduction and fertility, 86(2), 637-641.

Gardner, D. K., & Kelley, R. L. (2017). Impact of the IVF laboratory environment on human preimplantation embryo phenotype. Journal of Developmental Origins of Health and Disease, 8(4), 418-435.

Gawriluk, T. R., Simkin, J., Thompson, K. L., Biswas, S. K., Clare-Salzler, Z., Kimani, J.M., Kiama, S.G., Smith, J.J., Ezenwa, V.O. & Seifert, A. W. (2016). Comparative analysis of ear-hole closure identifies epimorphic regeneration as a discrete trait in mammals. Nature Communications, 7.

Ge, S. X. (2017). Exploratory bioinformatics investigation reveals importance of “junk” DNA in early embryo development. BMC Genomics, 18(1), 200.

Goddard, M. J., & Pratt, H. P. (1983). Control of events during early cleavage of the mouse embryo: an analysis of the ‘2-cell block’. Development, 73(1), 111-133.

Graf, A., Krebs, S., Heininen-Brown, M., Zakhartchenko, V., Blum, H., & Wolf, E. (2014). Genome activation in bovine embryos: review of the literature and new insights from RNA sequencing experiments. Animal Reproduction Science, 149(1), 46-58.

Guinez, C., Morelle, W., Michalski, J. C., & Lefebvre, T. (2005). O-GlcNAc glycosylation: a signal for the nuclear transport of cytosolic proteins? The international journal of biochemistry & cell biology, 37(4), 765-774.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M. & MacManes, M. D. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols, 8(8), 1494.

Harrison, P. F. (2017). Varistran: Anscombe’s variance stabilizing transformation for RNA-seq gene expression data. The Journal of Open Source Software. DOI: http://doi.org/10.21105/joss.00257

Harrison, P. F. (2018). Topconfects R package (Version 1.0.1). Zenodo. DOI: http://doi.org/10.5281/zenodo.1165213

Heyn, P., Kircher, M., Dahl, A., Kelso, J., Tomancak, P., Kalinka, A. T., & Neugebauer, K. M. (2014). The earliest transcribed zygotic genes are short, newly evolved, and different across species. Cell reports, 6(2), 285-292.

Huson, D. H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.J., & Tappu, R. (2016). MEGAN community edition-interactive exploration and analysis of large-scale microbiome sequencing data. PLoS Computational Biology, 12(6), e1004957.

Ivanova, I., Much, C., Di Giacomo, M., Azzi, C., Morgan, M., Moreira, P. N., Monahan, J., Carrieri, C., Enright, A.J. & O’Carroll, D. (2017). The RNA m6A reader YTHDF2 is essential for the post-transcriptional regulation of the maternal transcriptome and oocyte competence. Molecular Cell, 67(6), 1059-1067.

Kanitz, A., Gypas, F., Gruber, A. J., Gruber, A. R., Martin, G., & Zavolan, M. (2015). Comparative assessment of methods for the computational inference of transcript isoform abundance from RNA-seq data. Genome Biology, 16(1), 150.

Langmead, B. & Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods, 9(4), 357.

Lelièvre, J. M., Peynot, N., Ruffini, S., Laffont, L., Le Bourhis, D., Girard, P. M., & Duranthon, V. (2017). Regulation of heat-inducible HSPA1A gene expression during maternal-to-embryo transition and in response to heat in in vitro-produced bovine embryos. Reproduction, Fertility and Development, 29(9), 1868-1881.

Leshkowitz, D., Feldmesser, E., Friedlander, G., Jona, G., Ainbinder, E., Parmet, Y., & Horn-Saban, S. (2016). Using synthetic mouse spike-in transcripts to evaluate RNA-Seq analysis tools. PloS One, 11(4), e0153782.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics, 25(16), 2078-2079.

Lindeberg, M., Hovatta, O., & Ährlund-Richter, L. (2004). Real-time reverse transcription-polymerase chain reaction analysis of translation initiation factor 1A (eIF-1A) in human and mouse preimplantation embryos. Reproductive Biomedicine Online, 8(3), 338-343.

Mamrot, J., Legaie, R., Ellery, S. J., Wilson, T., Seemann, T., Powell, D. R., Gardner, D.K., Walker, D.W., Temple-Smith, P., Papenfuss, A.T. & Dickinson, H. (2017). De novo transcriptome assembly for the spiny mouse (Acomys cahirinus). Scientific Reports, 7(1), 8996.

Mantikou, E., Jonker, M. J., Wong, K. M., van Montfoort, A. P., De Jong, M., Breit, T. M., Repping, S. & Mastenbroek, S. (2015). Factors affecting the gene expression of in vitro cultured human preimplantation embryos. Human Reproduction, 31(2), 298-311.

McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Research, 40(10), 4288-4297.

Memili, E., & First, N. L. (2000). Zygotic and embryonic gene expression in cow: a review of timing and mechanisms of early gene expression as compared with other species. Zygote, 8(1), 87-96.

Nagaraj, R., Sharpley, M.S., Chi, F., Braas, D., Zhou, Y., Kim, R., Clark, A.T. and Banerjee, U. (2017). Nuclear localization of mitochondrial TCA cycle enzymes as a critical step in mammalian zygotic genome activation. Cell, 168(1-2), 210-223.

Niakan, K. K., Han, J., Pedersen, R. A., Simon, C., & Pera, R. A. R. (2012). Human pre-implantation embryo development. Development, 139(5), 829-841.

Nowak, R. M. (1999). Walker's mammals of the world (Vol. 1). JHU Press.

O'Neil, D., Glowatz, H., & Schlumpberger, M. (2013). Ribosomal RNA Depletion for Efficient Use of RNA-Seq Capacity. Current Protocols in Molecular Biology, 4-19.

Pasco, R., Gardner, D. K., Walker, D. W., & Dickinson, H. (2012). A superovulation protocol for the spiny mouse (Acomys cahirinus). Reproduction, Fertility and Development, 24(8), 1117-1122.

Piko, L., & Clegg, K. B. (1982). Quantitative changes in total RNA, total poly (A), and ribosomes in early mouse embryos. Developmental Biology, 89(2), 362-378.

Priyam, A., Woodcroft, B.J., Rai, V., Munagala, A., Moghul, I., Ter, F., Gibbins, M.A., Moon, H., Leonard, G., Rumpf, W. & Wurm, Y., 2015. Sequenceserver: a modern graphical user interface for custom BLAST databases. Biorxiv, 033142.

Quinn, T. A., Ratnayake, U., Dickinson, H., Castillo-Melendez, M., & Walker, D. W. (2016). The feto-placental unit, and potential roles of dehydroepiandrosterone (DHEA) in prenatal and postnatal brain development: a re-examination using the spiny mouse. The Journal of Steroid Biochemistry and Molecular Biology, 160, 204-213.

Quinn, T.A., Ratnayake, U., Dickinson, H., Nguyen, T.H., McIntosh, M., Castillo-Melendez, M., Conley, A.J. & Walker, D.W., (2013). Ontogeny of the adrenal gland in the spiny mouse, with particular reference to production of the steroids cortisol and dehydroepiandrosterone. Endocrinology, 154(3), 1190-1201.

Richter, J. D., & Sonenberg, N. (2005). Regulation of cap-dependent translation by eIF4E inhibitory proteins. Nature, 433(7025), 477.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 26(1), 139-140.

Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biology, 11(3), R25.

Schier, A. F. (2007). The maternal-zygotic transition: death and birth of RNAs. Science, 316(5823), 406-407.

Schramm, R. D., & Bavister, B. D. (1999). A macaque model for studying mechanisms controlling oocyte development and maturation in human and non-human primates. Human Reproduction, 14(10), 2544-2555.

Schuierer, S., Carbone, W., Knehr, J., Petitjean, V., Fernandez, A., Sultan, M., & Roma, G. (2017). A comprehensive assessment of RNA-seq protocols for degraded and low-quantity samples. BMC Genomics, 18(1), 442.

Seifert, A. W., Kiama, S. G., Seifert, M. G., Goheen, J. R., Palmer, T. M., & Maden, M. (2012). Skin shedding and tissue regeneration in African spiny mice (Acomys). Nature, 489(7417), 561-565.

Simkin, J., Gawriluk, T. R., Gensel, J. C., & Seifert, A. W. (2017). Macrophages are necessary for epimorphic regeneration in African spiny mice. eLife, 6, e24623.

Song, L., & Florea, L. (2015). Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads. GigaScience. 4:48.

Svoboda, P. (2017). Mammalian zygotic genome activation. In Seminars in Cell & Developmental Biology. Academic Press.

Tadros, W., & Lipshitz, H. D. (2009). The maternal-to-zygotic transition: a play in two acts. Development, 136(18), 3033-3042.

Taft, R. A. (2008). Virtues and limitations of the preimplantation mouse embryo as a model system. Theriogenology, 69(1), 10-16.

Taylor, D. M., Ray, P. F., Ao, A., Winston, R. M., & Handyside, A. H. (1997). Paternal transcripts for glucose-6-phosphate dehydrogenase and adenosine deaminase are first detectable in the human preimplantation embryo at the Three-to Four-Cell stage. Molecular reproduction and development, 48(4), 442-448.

Telford, N.A., Watson, A.J. & Schultz, G.A. (1990). Transition from maternal to embryonic control in early mammalian development: a comparison of several species. Molecular Reproduction and Development, 26(1), 90-100.

Tesařík, J., Kopečný, V., Plachot, M., & Mandelbaum, J. (1987). Ultrastructural and autoradiographic observations on multinucleated blastomeres of

human cleaving embryos obtained by in-vitro fertilization. Human reproduction, 2(2), 127-136.

Vassena, R., Boué, S., González-Roca, E., Aran, B., Auer, H., Veiga, A., & Belmonte, J. C. I. (2011). Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. Development, 138(17), 3699-3709.

Wang, H., & Dey, S. K. (2006). Roadmap to embryo implantation: clues from mouse models. Nature Reviews Genetics, 7(3), 185.

Wei, T., & Simko, V. (2017). R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). Available from https://github.com/taiyun/corrplot

Whitby, S., Salamonsen, L. A., & Evans, J. (2017). The Endometrial Polarity Paradox: Differential Regulation of Polarity Within Secretory-Phase Human Endometrium. Endocrinology, 159(1), 506-518.

Williams, C. R., Baccarella, A., Parrish, J. Z., & Kim, C. C. (2017). Empirical assessment of analysis workflows for differential expression analysis of human samples using RNA-Seq. BMC Bioinformatics, 18(1), 38.

Wilson, D.E. & Reeder, D.M. eds., (2005). Mammal species of the world: a taxonomic and geographic reference. JHU Press.

Xie, D., Chen, C.C., Ptaszek, L.M., Xiao, S., Cao, X., Fang, F., Ng, H.H., Lewin, H.A., Cowan, C. & Zhong, S. (2010). Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. Genome research, 20(6), 804-815.

Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C. Y., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y.E. & Liu, J. Y. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. Nature, 500(7464), 593.

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J. & Huang, J., 2013. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nature Structural and Molecular Biology, 20(9), 1131.

Yartseva, V., & Giraldez, A. J. (2015). The maternal-to-zygotic transition during vertebrate development: a model for reprogramming. In Current Topics in Developmental Biology, Vol. 113, 191-232. Academic Press.

## 4.2 Supplementary figures

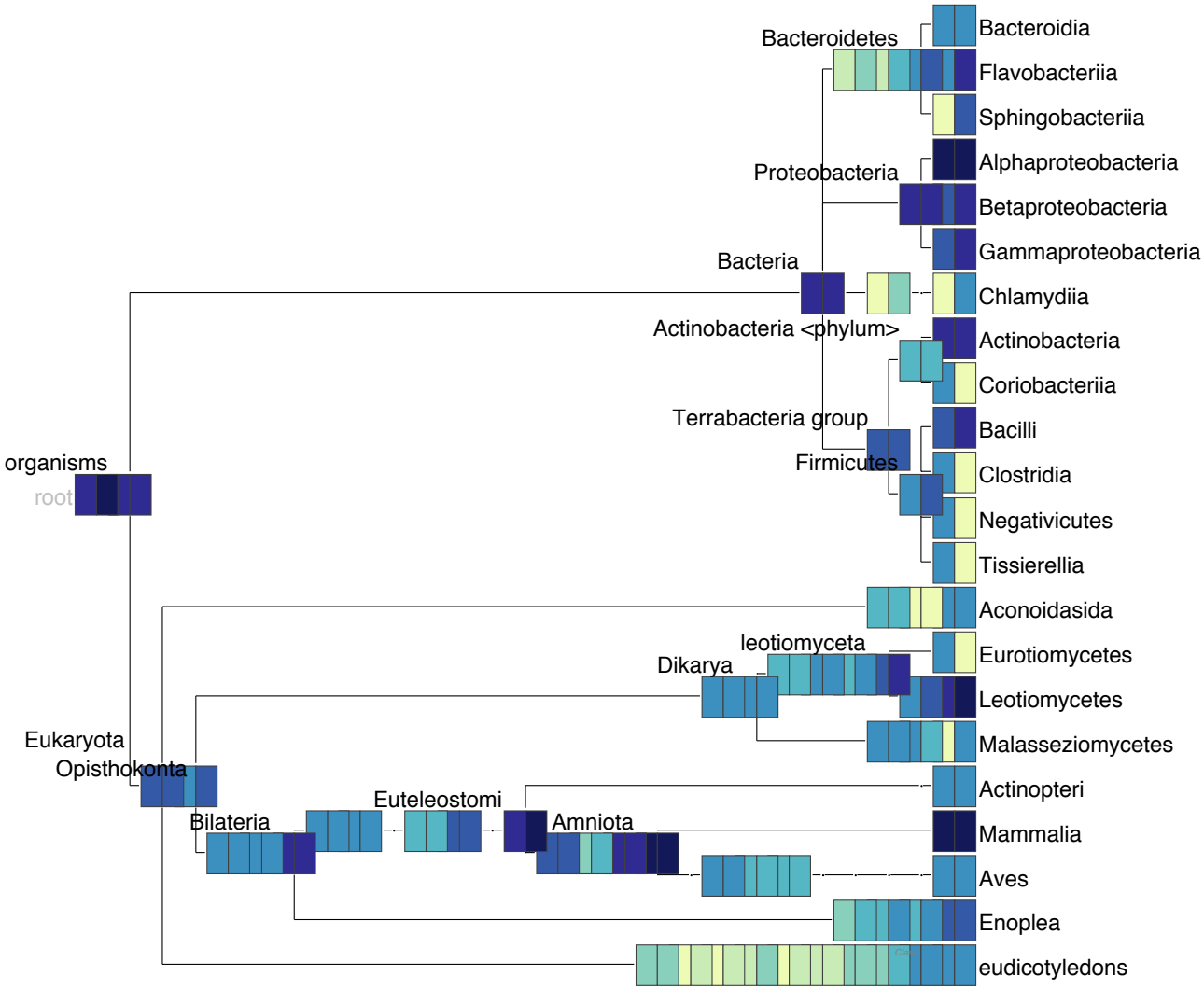



Figure S1: Metatranscriptomic analysis of Trinity-normalized reads in samples 2-cell_C and 8-cell_A illustrating read alignment to eukaryotic and prokaryotic taxa within the NCBI nr protein database ("non-redundant" proteins; n=4,348,972). Read alignments were summarised at the Class level using MEGAN6 implementing the 'Blues' colour scale: a higher proportion of aligned reads is represented by a darker colour (highest number of reads aligned per taxonomic group = "Midnight Blue"). Within these two contaminated samples ~30% of total reads aligned to Mammalia, ~30% aligned to Alphaproteobacteria, and ~40% were spread across other Classes as indicated.

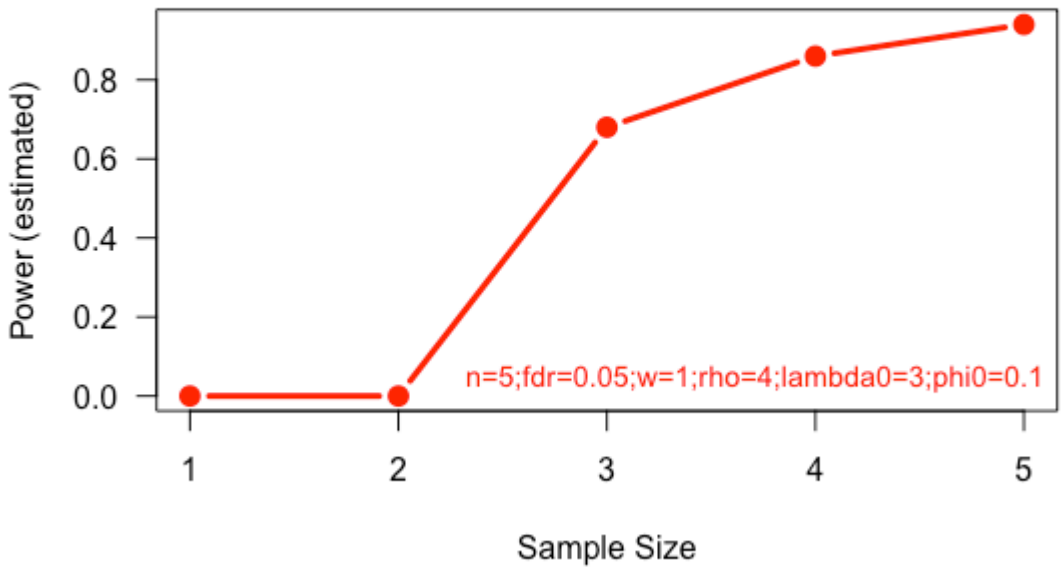



Figure S2: Power estimates for various sample sizes. Parameters represent expected values based on past / similar experiments. "w": expected normalization factor for sample groups (a value of 1 representing approximately equal read counts across sample groups). "rho": fold change required for significance, "FC=4" => $\log_2(FC)=2$. "lambda0": anticipated average read count per sample (actual values were higher than predicted: Figure 2); "phi0": average dispersion across samples (actual dispersion value was slightly lower than expected). With the parameters specified, n=4 in each group is recommended to achieve power >0.8.

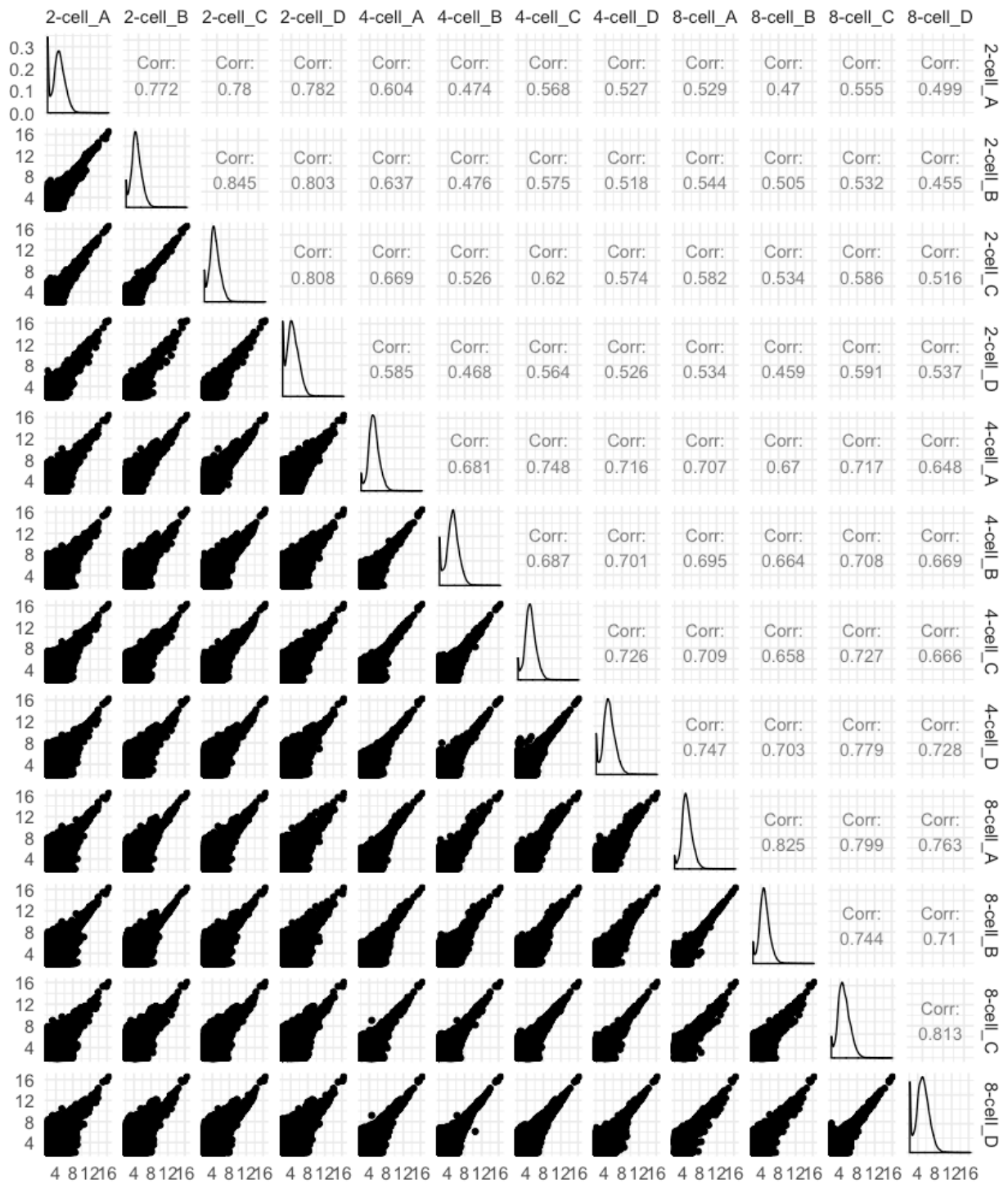



Figure S3: Correlation matrix for all samples (n=12). Spearman correlation values (upper right), distribution (diagonal) and concordance (lower left) of gene cluster (iso-group) abundance are illustrated.

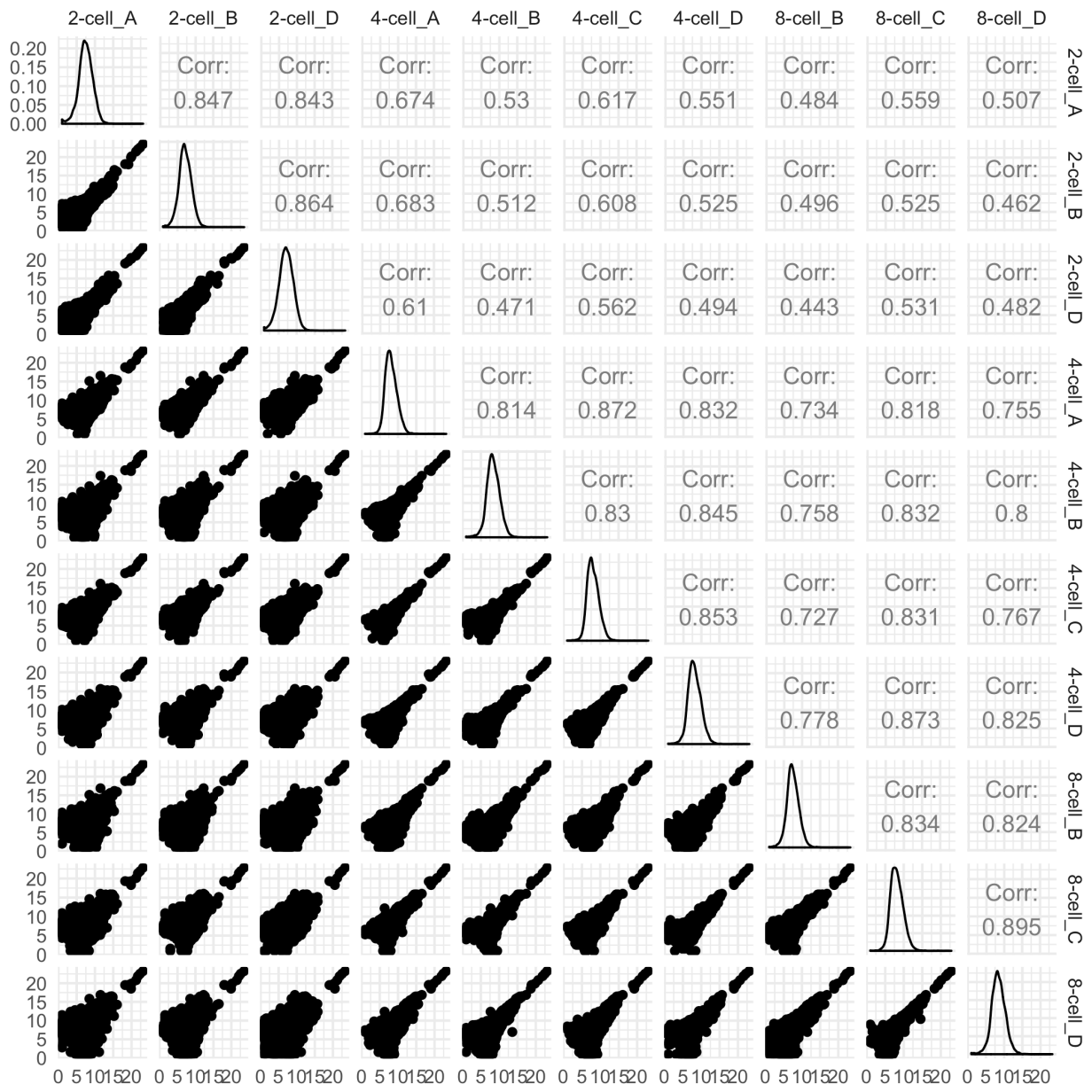



Figure S4: Correlation matrix for uncontaminated samples (n=10). Spearman correlation values (upper right), distribution (diagonal) and concordance (lower left) of gene cluster (iso-group) abundance are illustrated.

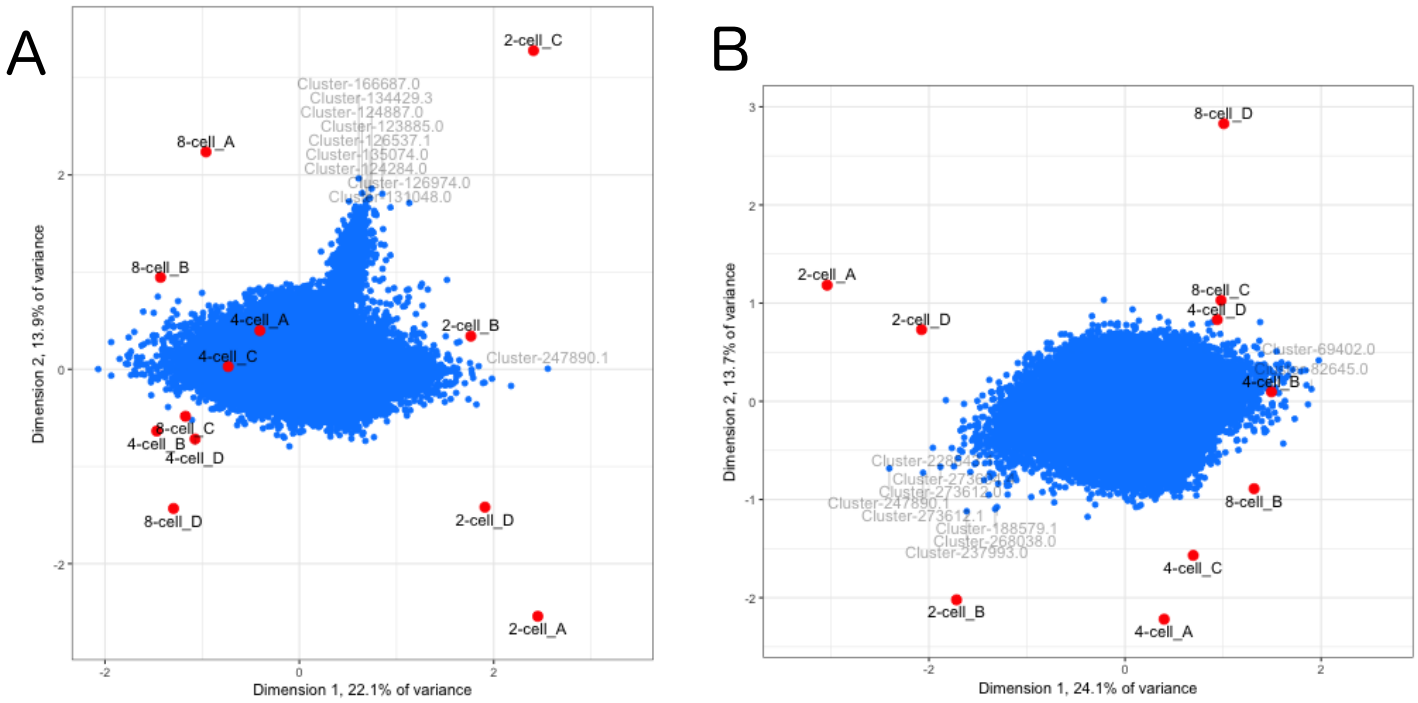



Figure S5: Singular Value Decomposition (SVD) biplots of iso-group expression per sample. Blue dots represent iso-group expression values and red dots represent samples, with (A) contaminated samples included, and (B) with contaminated samples excluded. The top 10 differentially expressed iso-groups are labelled in each plot. Many of the top DE iso-groups in (A) correspond to prokaryotic taxa (9/10). In comparison, after the contaminated samples were excluded in (B) the top DE iso-groups all correspond to mammalian taxa.

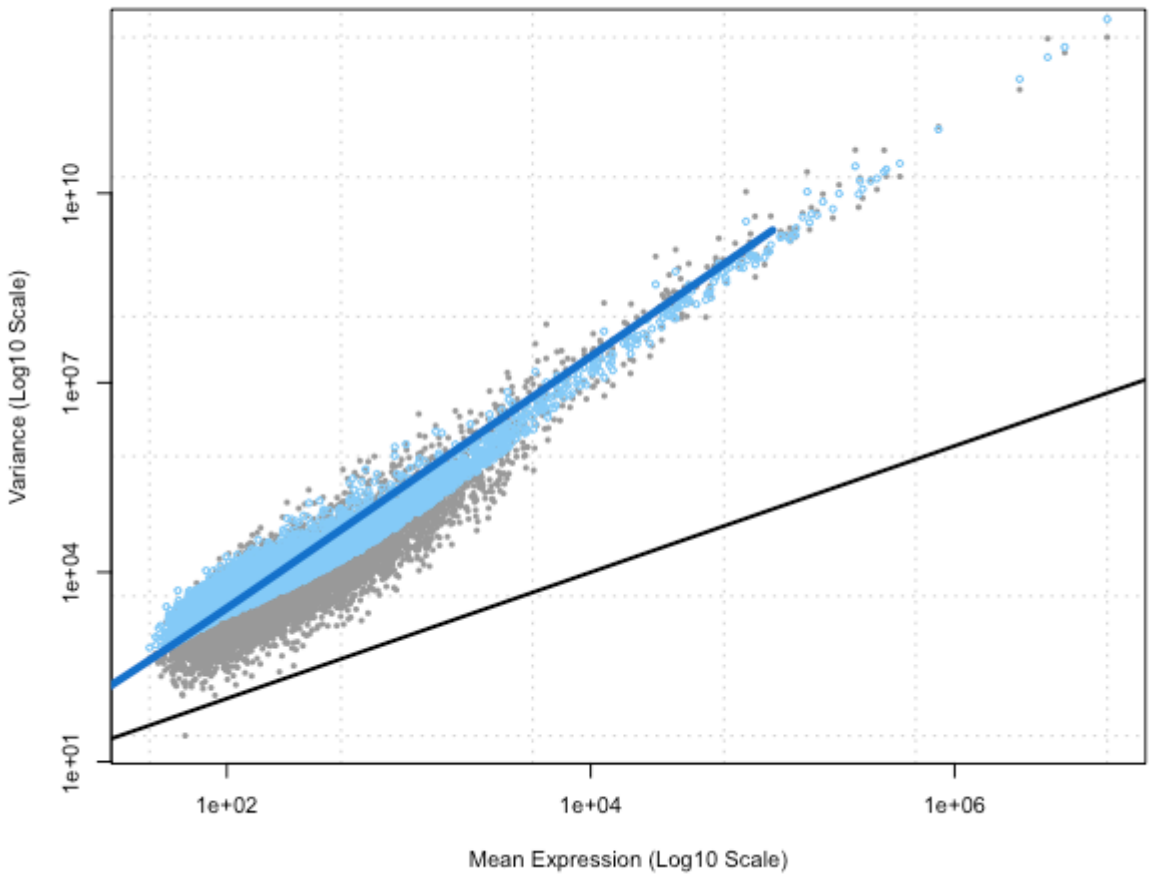



Figure S6: Fit of the edgeR negative binomial distribution to iso-group counts.

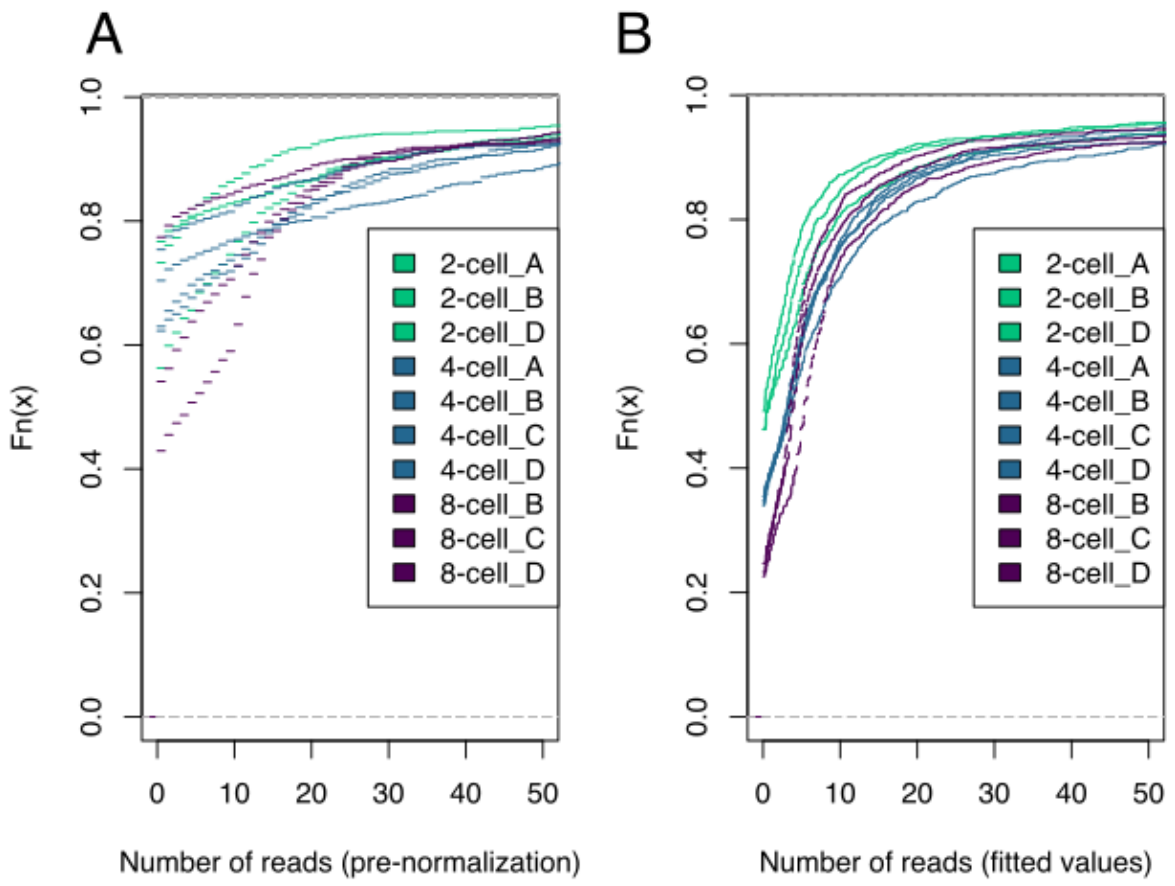
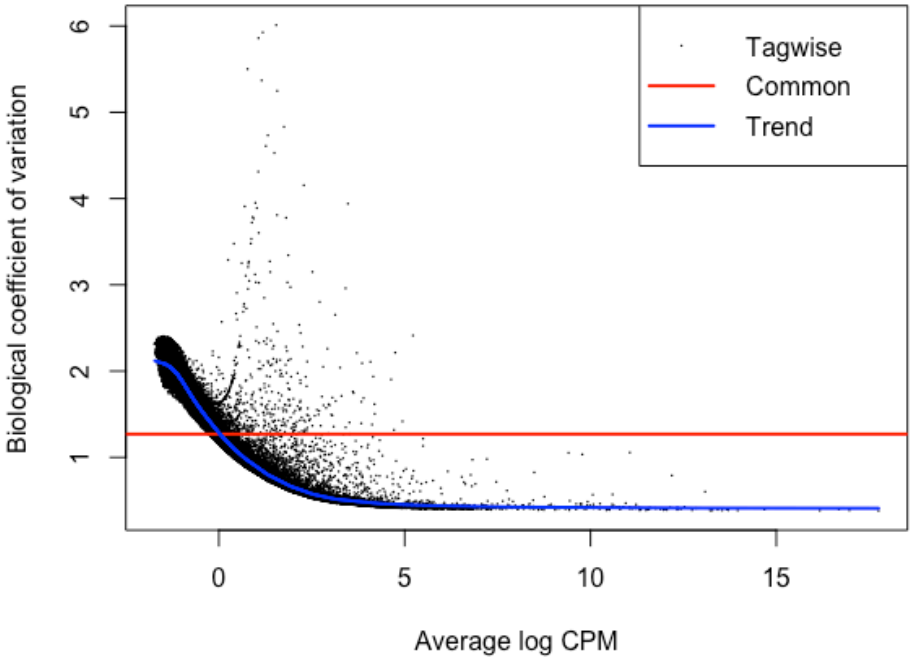



Figure S7: Multiple empirical cumulative distribution of reads for each sample. (A) Read counts for all iso-groups prior to normalization. (B) Normalization and fitting using the negative binomial model improved grouping by developmental stage, especially for below-average read counts.

A




B

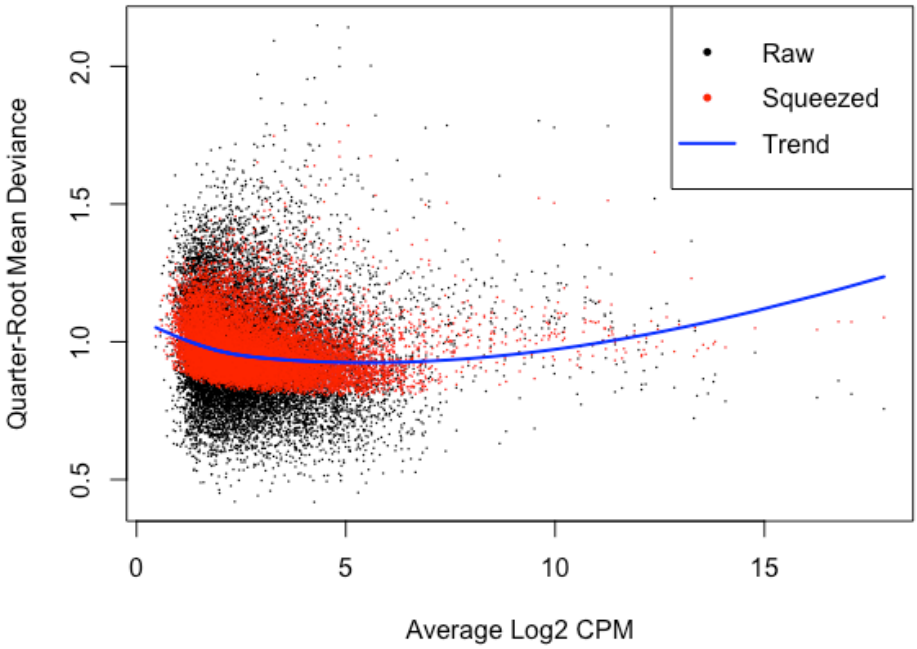



Figure S8: Scatterplots illustrating (A) the biological coefficient of variation and (B) the quarter-root of the quasi-likelihood dispersions for all iso-groups. cpm=counts-per-million.

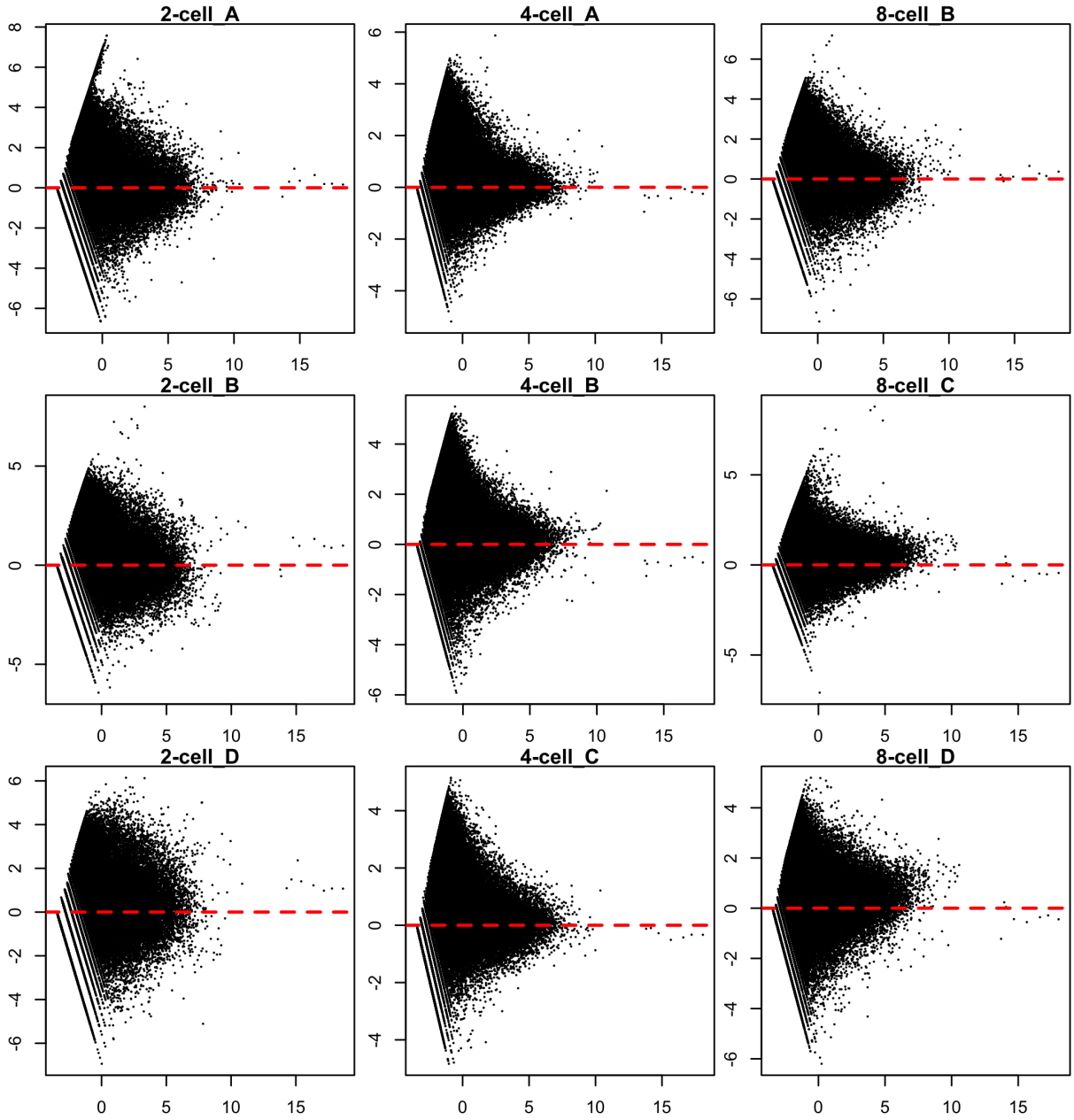



Figure S9: Mean-Difference (MD) plots comparing each uncontaminated sample to an artificial reference library constructed from the average of all other samples. Sample 4-cell_D was included in the analysis but excluded from this figure for readability; full figure with all samples: https://doi.org/10.4225/03/5a9531283d103. Positive skew in samples (eg 8-cell_C and 8-cell_D) corresponds to greater variation in TMM normalization factors (Table S2).

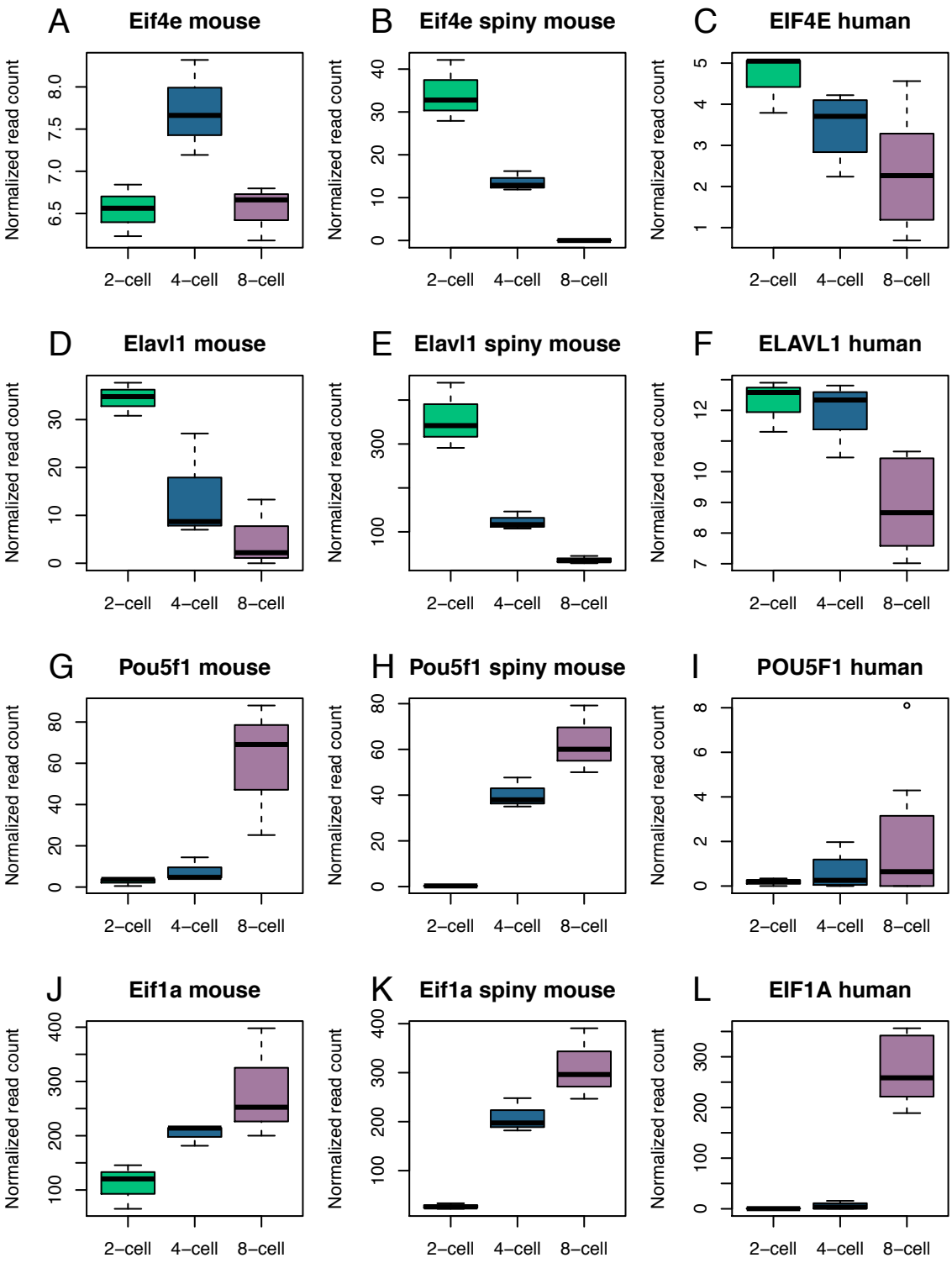



Figure S10: Expression of select iso-groups in common mouse, spiny mouse and human embryos during the 2-cell, 4-cell and 8-cell stages of development. Eukaryotic translation initiation factor 4E (EIF4E) is a key component of the translation machinery and a known driver of genome activation in mammals. ELAV like RNA binding protein 1 (ELAVL1) is an RNA stabilizer involved in maternally-inherited transcript clearance. POU Class 5 Homeobox 1 (POU5F1), also known as OCT3/4, is a key regulator of pluripotency with highest expression at the morula and blastocyst stages. Eukaryotic translation initiation factor 1A (EIF1A) is required for protein biosynthesis and an increase in expression occurs during the EGA.

Supplementary Table 1: Statistics for Trinity transcriptome assembly (output from TrinityStats.pl).

```
################################
## Counts of transcripts, etc.
################################
Total trinity 'genes':    353118
Total trinity transcripts:        595435
Percent GC: 46.63

########################################
Stats based on ALL transcript contigs:
########################################

        Contig N10: 1585
        Contig N20: 1072
        Contig N30: 784
        Contig N40: 599
        Contig N50: 471

        Median contig length: 300
        Average contig: 429.27
        Total assembled bases: 255601661


####################################################
## Stats based on ONLY LONGEST ISOFORM per 'GENE':
####################################################

        Contig N10: 1354
        Contig N20: 853
        Contig N30: 610
        Contig N40: 469
        Contig N50: 379

        Median contig length: 277
        Average contig: 378.65
        Total assembled bases: 133709809
```

Supplementary Table 2: TMM-normalized library sizes

| Sample | Library size | Adjusted library size |
|---|---|---|
| 2cell_A | 30584677 | 30523713.55 |
| 2cell_B | 32974138 | 24386500.70 |
| 2cell_C | 37049595 | 22227903.12 |
| 2cell_D | 35508389 | 20709266.29 |
| 4cell_A | 31371210 | 38551172.77 |
| 4cell_B | 32900876 | 48943239.95 |
| 4cell_C | 28156913 | 36219884.07 |
| 4cell_D | 29500819 | 39510735.95 |
| 8cell_A | 34693598 | 31935445.77 |
| 8cell_B | 29556455 | 28958706.22 |
| 8cell_C | 26386705 | 29998846.68 |
| 8cell_D | 34873989 | 41815650.37 |

Supplementary Table 5: Identification of Benchmarking Universal Single-Copy Orthologs (BUSCOs) from the OrthoDBv9 'Mammalia' dataset set in each assembly.

| Read supported Trinity embryo-specific contigs: | |
|---|---|
| Complete BUSCOs (C) | 712 |
| Complete and single-copy BUSCOs (S) | 337 |
| Complete and duplicated BUSCOs (D) | 375 |
| Fragmented BUSCOs (F) | 1384 |
| Missing BUSCOs (M) | 2008 |
| Total BUSCO groups searched | 4104 |

| Trinity embryo-specific assembly: | |
|---|---|
| Complete BUSCOs (C) | 989 |
| Complete and single-copy BUSCOs (S) | 380 |
| Complete and duplicated BUSCOs (D) | 609 |
| Fragmented BUSCOs (F) | 1705 |
| Missing BUSCOs (M) | 1410 |
| Total BUSCO groups searched | 4104 |
| TransRate score | .1255 |

| Read supported Trinity embryo-specific contigs combined with the Trinity_v2.3.2 assembly: | |
|---|---|
| Complete BUSCOs (C) | 3668 |
| Complete and single-copy BUSCOs (S) | 853 |
| Complete and duplicated BUSCOs (D) | 2815 |
| Fragmented BUSCOs (F) | 358 |
| Missing BUSCOs (M) | 78 |
| Total BUSCO groups searched | 4104 |

# General Discussion

Early development in mammals is dynamic. Gamete maturation, fertilization, joining of haploid genomes, cell division, implantation and development *in utero* are highly complex events, and the mechanisms that control the transition from a single cell to a whole organism are not completely understood. With regards to human infertility, assisted reproduction is our only clinical recourse when these processes go awry and often the underlying causes of developmental failure are not known. Hence, improving our understanding of the processes driving early development will improve our ability to treat or even prevent infertility and improve clinical outcomes. Historically, much of the work in this field has been conducted using laboratory animals such as the mouse, however differences between mice and humans can limit translation of findings. This thesis outlines investigation of early development in the spiny mouse (*Acomys cahirinus*), characterizing embryonic gene expression and evaluating the utility of this species as a model of early human developmental biology. Here, I describe the progress made developing methods to produce interspecies and intersex chimeric embryos, development of methods to increase the efficacy of embryo transfer with a view to implement this technique in the spiny mouse, assembly of a transcriptome for the spiny mouse, and characterization of gene transcription in the spiny mouse embryo during the embryonic genome activation. These studies have contributed to our knowledge of early developmental biology in this species.

## Embryology

Prior to the studies described here, little was known about spiny mouse embryology. There is a single publication by Pasco et al. (2012) in which the authors describe methods for inducing superovulation in spiny mice, however the majority of work in this species has

focused on development later in gestation and in the neonate, for example Dickinson et al. (2005; 2007; 2011; 2013; 2014; 2017), Ellery et al. (2013; 2015; 2016; 2017), Fleiss et al. (2012), Ireland et al. (2008; 2011), La Rosa et al. (2016), O'Connell et al. (2011; 2013a; 2013b), Quinn et al. (2013; 2014; 2016) and Ratnayake et al. (2012; 2014). The rationale for exploring embryo development in the spiny mouse was built on the findings described in the above publications. Specific characteristics of interest include the human-like hormonal profile of the spiny mouse (cortisol and DHEA), their precocial offspring with *in utero* development largely complete by the time of birth, their placentation that more closely represents human placentation than other rodents, and sex-specific differences in fetal and placental physiology identified in normal and pathological pregnancies. Combined, these findings and the outcomes described in this thesis suggest the spiny mouse is well-suited to specific applications in developmental biology.

Development of methods to produce chimeric embryos in the spiny mouse posed many challenges. The majority of these were successfully overcome: embryos were successfully obtained from mice, cultured *in vitro*, manipulated at the blastocyst stage ("reconstructed"), and embryo transfer into the uterus of pseudopregnant recipient mice was optimised (Chapter 2). Despite the progress made, two challenges impeded translation of developed methods to the spiny mouse: the 4-cell block in spiny mouse embryos cultured *in vitro* was not overcome and pseudopregnancy was not successfully induced in female spiny mouse embryo recipients. Unsuccessful attempts to overcome these problems included *in vitro* culture of spiny mouse embryos in 72 different formulations of existing mouse/human embryo culture media and application of traditional methods of pseudopregnancy induction, such as administration of exogenous progesterone, cervical stimulation and pairing with vasectomised males. Results from these experiments have been presented at

conferences (e.g. https://doi.org/10.4225/03/5a9d418d2acb5 / https://doi.org/10.4225/03/5a9d4329d9f85) and research continues in these areas for current post-graduate students (Nadia Bellofiore and Jarrod McKenna, personal communication), however I was unable to continue pursuing the original aims of my PhD project without effective solutions to these technical challenges. The discovery of fundamental differences in reproductive physiology between the spiny mouse and other rodents that occurred 3.5 years into my PhD provides new perspective to our inability to induce pseudopregnancy in this species using traditional methods.

The discovery of menstruation in the spiny mouse (Bellofiore et al., 2017) provides a logical explanation for poor response to rodent-specific assisted reproduction techniques (ARTs). Physiological response to ART treatments can be unpredictable when applied to novel species, however the absence of characteristics common in rodents such as the Whitten effect (1956, 1966, 1968) was surprising given the close phylogenetic relationship of the spiny mouse to other members within the rodent family (Agulnik & Silver, 1996; Chevret & Hänni, 1994). We now know this is due to their primate-like hormonal milieu and their atypical reproductive physiology, which is unique amongst rodents. This discovery was unexpected and these extraordinary physiological differences were not anticipated when designing these experiments. As a consequence of these atypical characteristics the application of methods developed for non-rodent species will likely improve the response of spiny mice to ARTs, such as described by Pasco et al. (2012) where larger doses of extraneous hormones with increased frequency of administration were required to successfully stimulate follicle recruitment and ovulation in spiny mice, relative to their size (similar to protocols used in primates e.g. Liu et al., 2018). Consequently, future efforts to

induce pseudopregnancy and to culture embryos *in vitro* past the 4-cell block may benefit from incorporating ART methods developed for primates, rather than for other rodents.

Overcoming the 4-cell block in spiny mouse embryos cultured *in vitro* proved to be another significant challenge. Historically, the first major steps to overcome the cell-block phenomenon in mammals occurred in the 1950s with three landmark papers: Whitten (1956) reported development of eight-cell mouse embryos to blastocyst when cultured in a basic growth medium (in this thesis I report the same finding for spiny mouse embryos obtained at the 8-cell stage and cultured *in vitro*: Methods Development chapter), McLaren and Biggers (1958) reported that these blastocysts could be transferred into the uterus of a recipient female to implant and develop, and Chang (1959) successfully fertilized and cultured rabbit oocytes *in vitro* and transferred them to recipient females to produce live young (Biggers, 1958). These discoveries led to extensive efforts from several research groups to address the *in vitro* cell block, and since the 1960s the conditions required to circumvent cell blocks *in vitro* have been established for almost every species investigated (Gardner & Lane, 1996; Lawitts & Biggers, 1991).

Species with published methods to overcome the embryonic cell blocks in culture include primates (Goodeaux et al., 1990), cows (Camous et al., 1984; Thibault et al., 1966), pigs (Beckmann et al., 1990), sheep (Gandolfi & Moor, 1987, Walker et al., 1992), hamsters (Schini & Bavister, 1988), rabbits (Cole & Paul, 1965), rats (Miyoshi, 2016), gerbils (Baltz & Tartia, 2009; Obata & Tsujii, 2010) and various strains of mice (Gardner & Lane, 1996). The failure of spiny mouse embryos to respond to methods used successfully in these mammalian species was not anticipated and, combined with our inability to induce pseudopregnancy in the spiny mouse, further investigation was not conducted. Recent advances in this field provide further insight into the underlying mechanisms of cell blocks

during *in vitro* embryo culture (Chen et al., 2017; Nagaraj et al., 2017; Zhang et al., 2018), therefore an important future direction for this work will be to thoroughly investigate the cause of the 4-cell block in the spiny mouse. Overcoming the 4-cell block will likely require sophisticated methods, such as profiling gene transcription during this period of development in response to different culture conditions and comparing findings with other mammalian species such as primates.

## Genetics

As with spiny mouse embryology, very little was known about spiny mouse genetics prior to the studies conducted in this PhD project. The NCBI contained a single dataset, the mitochondrial genome, which was sequenced for phylogenetic analyses (Barome et al., 1998). In order to obtain genetic information for the spiny mouse, my primary supervisor (Hayley Dickinson) coordinated a project to build a reference transcriptome to simplify techniques such as qPCR primer design, *in situ* hybridisation probe design, and phylogenetic comparisons (for example, my unpublished analysis of the spiny mouse CRH gene conducted with Sherwyn Dutt: https://doi.org/10.4225/03/5aa2376b18b3f). Assembly and annotation of this spiny mouse transcriptome is described in Chapter 3, however, given the rapid development of software, protocols and literature in the field of *de novo* transcriptome assembly, many of the methods used are now deprecated or superseded. A future direction for this work is to download all available RNA-Seq data for the spiny mouse and assemble a new, more accurate and complete transcriptome using modern methods. Despite limitations in my approach the successful outcomes described in Chapter 3 enabled further investigation and characterization of gene expression during early development in the spiny mouse.

Our hypothesis that the 4-cell block in spiny mouse embryos correlates with the embryonic genome activation (EGA) was based on the observation that the 2-cell block in mouse embryos and 4-cell block in human embryos both coincide with the EGA (Latham, 1999). The timing of the EGA and preceding zygotic genome activation (ZGA) is well-characterized in the literature for many mammalian species (Kojima et al., 2014), however much of our knowledge of this developmental event is derived from studies conducted in mice and species-specific differences between mouse and human embryos have been identified (Latham et al., 1999; Lee et al., 2014). The specific genes activated during the EGA have been shown to differ significantly between mice and humans, with only ~40% concurrence between the two species (Heyn et al., 2014; Xie et al., 2010). Differences in the timing of gene transcription initiation and the 'burst' in expression are hypothesised to represent differences in the underlying mechanisms of the EGA, although this has not been fully explored as we currently lack an accurate model of the EGA in humans. The spiny mouse model may enable us to better understand the underlying mechanisms involved in the EGA and in the regulation of early development.

A primary goal of this project was to identify the timing of the EGA in the spiny mouse. This was successfully achieved, with the EGA shown to begin at the 2-cell to 4-cell stage. This is similar to the initiation of the EGA in mice, however the spiny mouse has a greater number of genes that are transcribed later in development (at the 4- to 8-cell stage) compared to the mouse. This pattern of transcription is a closer representation of the pattern of expression seen in human embryos than the common mouse, however further investigation into other aspects of development is required to confirm this initial finding and determine whether the spiny mouse is a superior model of embryonic gene transcription

compared to the mouse. Despite success in pursuit of my overall aims, there were a number of limitations.

## Limitations

Various technical limitations affected each stage of this PhD project. These were primarily related to the use of a non-model species. A major limitation in genomic analyses is the absence of a complete genome. A draft genome for the spiny mouse was assembled in 2016 (Jessica Alföldi; personal communication), however it is relatively incomplete and is not suitable for many applications, such as a reference for RNA-Seq studies (Ashley Seifert, Jeremy Johnson, Joerg Mueller, personal communication). Other limitations included sample contamination and imperfect rRNA depletion when evaluating gene expression during the EGA, as well as difficulty procuring antibodies specific to the spiny mouse required for various scientific techniques such as staining cells for specific markers (immunohistochemistry), protein quantification (Western blotting), cell sorting (FACS), and antigen detection (RIA / ELISA). In addition to these specific limitations, there is a general scarcity of information regarding the basic biology of the spiny mouse. Many of these limitations will likely be overcome in the near future as the scientific literature on the spiny mouse evolves, increasing the utility of the spiny mouse for applications in medical research. Although there are many limitations of this work, overall I was successful in characterizing the earliest period of development in this species. The experiments described in this thesis provide strong evidence of untapped potential for future scientific discoveries in this species.

## Future directions

Future directions for this work include characterization of other stages of embryo development, such as the oocyte, zygote, 16-cell, morula and blastocyst stages, and to apply the knowledge gained from this thesis to overcome the 4-cell block in the spiny mouse embryo when cultured *in vitro*, enabling use of spiny mouse embryos clinically for quality control (i.e. as a more sensitive 1-cell mouse embryo assay) and as a more sensitive assay for newly developed human embryo culture media (Morbeck, 2017). Based on the results described here, further characterization of other aspects of the EGA is another future direction for this work. DNA and histone methylation, imprinting, transcription factors such as Paf, Jun, Jund, Junb, Egr1, Fosl2, Cebpb, Sp4, Sp1, Atf3, Btg2, Fosb, and Klf5, non-coding RNA, genome conformation, alternative RNA splicing, RNA editing, proteomics, recruitment and clearance of maternally-inherited factors are all potential targets for future experiments. Finally, characterization of reproduction and development in the spiny mouse is an ongoing future direction for the work described in this thesis (for example, we are currently investigating anatomy and physiology of the reproductive tract in the male spiny mouse: https://doi.org/10.4225/03/5aa5fc4f50df9).

## Conclusion

The findings described in this thesis show for the first time that the spiny mouse is a promising candidate for modelling human embryology and genetics during early development. This species provides advantages over other model species currently used for this purpose displaying physiological characteristics such as a menstrual cycle and human-like endocrine profile. This research provides tools to facilitate further investigation into this species, including a reference transcriptome and profile of gene expression

activation during early embryonic development. Each of the studies outlined in this thesis have significantly contributed to our understanding of early developmental biology in the spiny mouse and provide a foundation for evaluating the utility of this species as an animal model of early human development.

## References

Agulnik, S. I., & Silver, L. M. (1996). The Cairo spiny mouse Acomys cahirinus shows a strong affinity to the Mongolian gerbil Meriones unguiculatus. Molecular biology and evolution, 13(1), 3-6.

Baltz, J. M., & Tartia, A. P. (2009). Cell volume regulation in oocytes and early embryos: connecting physiology to successful culture media. Human reproduction update, 16(2), 166-176.

Barome, P. O., Monnerot, M., & Gautun, J. C. (1998). Intrageneric Phylogeny of Acomys (Rodentia, Muridae) Using Mitochondrial Gene Cytochrome b. Molecular phylogenetics and evolution, 9(3), 560-566.

Beckmann, L. S., Cantley, T. C., Rieke, A. R., & Day, B. N. (1990). Development and viability of one-and two-cell porcine embryos cultured through the “four-cell block”. Theriogenology, 33(1), 193.

Bellofiore, N., Ellery, S. J., Mamrot, J., Walker, D. W., Temple-Smith, P., & Dickinson, H. (2017). First evidence of a menstruating rodent: the spiny mouse (Acomys cahirinus). American Journal of Obstetrics & Gynecology, 216(1), 40-e1.

Biggers, J. D. (1998). Reflections on the culture of the preimplantation embryo. The International journal of developmental biology, 42(7), 879-884.

Camous, S., Heyman, Y., Méziou, W., & Menezo, Y. (1984). Cleavage beyond the block stage and survival after transfer of early bovine embryos cultured with trophoblastic vesicles. Journal of reproduction and fertility, 72(2), 479-485.

Chen, F., Lin, J., Sun, X., Xiao, B., Ning, S.F., Zhu, S., Wang, H.L. and Tan, J.H. (2017). Mechanisms by which in vitro meiotic arrest and sexual maturity improve developmental potential of mouse oocytes. Scientific reports, 7(1), 15763.

Chevret, P., & Hänni, C. (1994). Systematics of the spiny mouse (Acomys: Muroidea): Molecular and biochemical evidence. Israel Journal of Zoology, 40(2), 247-254.

Cole, R. J., & Paul, J. (1965). Properties of cultured preimplantation mouse and rabbitembryos, and cell strains derived from them. In Ciba Foundation Symposium-Preimplantation Stages of Pregnancy (pp. 82-122). John Wiley & Sons, Ltd.

Dickinson, H., Ellery, S., Ireland, Z., LaRosa, D., Snow, R., & Walker, D. W. (2014). Creatine supplementation during pregnancy: summary of experimental studies suggesting a treatment to improve fetal and neonatal morbidity and reduce mortality in high-risk human pregnancy. BMC pregnancy and childbirth, 14(1), 150.

Dickinson, H., Ellery, S., Davies-Tuck, M., Tolcos, M., Nitsos, I., Walker, D. W., & Miller, S. L. (2017). Description of a method for inducing fetal growth restriction in the spiny mouse. Journal of developmental origins of health and disease, 8(5), 550-555.

Dickinson, H., Ireland, Z. J., LaRosa, D. A., O'Connell, B. A., Ellery, S., Snow, R., & Walker, D. W. (2013). Maternal dietary creatine supplementation does not alter the capacity for creatine synthesis in the newborn spiny mouse. Reproductive Sciences, 20(9), 1096-1102.

Dickinson, H., O'Connell, B. A., Walker, D. W., & Moritz, K. M. (2012). Sex-specific effects of prenatal glucocorticoids on placental development. In Glucocorticoids-New Recognition of Our Familiar Friend. InTech.

Dickinson, H., Walker, D. W., Cullen-McEwen, L., Wintour, E. M., & Moritz, K. (2005). The spiny mouse (Acomys cahirinus) completes nephrogenesis before birth. American Journal of Physiology-Renal Physiology, 289(2), F273-F279.

Dickinson, H., Walker, D. W., Wintour, E. M., & Moritz, K. (2007). Maternal dexamethasone treatment at midgestation reduces nephron number and alters renal gene expression in the fetal

spiny mouse. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 292(1), R453-R461.

Ellery, S. J., Ireland, Z., Kett, M. M., Snow, R., Walker, D. W., & Dickinson, H. (2013). Creatine pretreatment prevents birth asphyxia–induced injury of the newborn spiny mouse kidney. Pediatric research, 73(2), 201.

Ellery, S. J., LaRosa, D. A., Kett, M. M., Della Gatta, P. A., Snow, R. J., Walker, D. W., & Dickinson, H. (2015). Maternal creatine homeostasis is altered during gestation in the spiny mouse: is this a metabolic adaptation to pregnancy? BMC pregnancy and childbirth, 15(1), 92.

Ellery, S. J., LaRosa, D. A., Kett, M. M., Della Gatta, P. A., Snow, R. J., Walker, D. W., & Dickinson, H. (2016). Dietary creatine supplementation during pregnancy: a study on the effects of creatine supplementation on creatine homeostasis and renal excretory function in spiny mice. Amino acids, 48(8), 1819-1830.

Ellery, S.J., LaRosa, D.A., Cullen-McEwen, L.A., Brown, R.D., Snow, R.J., Walker, D.W., Kett, M.M. and Dickinson, H. (2017). Renal dysfunction in early adulthood following birth asphyxia in male spiny mice, and its amelioration by maternal creatine supplementation during pregnancy. Pediatric research, 81(4), 646.

Fleiss, B., Parkington, H.C., Coleman, H.A., Dickinson, H., Yawno, T., Castillo-Melendez, M., Hirst, J.J. and Walker, D.W. (2012). Effect of maternal administration of allopregnanolone before birth asphyxia on neonatal hippocampal function in the spiny mouse. Brain research, 1433, 9-19.

Fushan, A.A., Turanov, A.A., Lee, S.G., Kim, E.B., Lobanov, A.V., Yim, S.H., Buffenstein, R., Lee, S.R., Chang, K.T., Rhee, H. and Kim, J.S. (2015). Gene expression defines natural changes in mammalian lifespan. Aging cell, 14(3), 352-365.

Gandolfi, F., & Moor, R. M. (1987). Stimulation of early embryonic development in the sheep by co-culture with oviduct epithelial cells. Journal of reproduction and fertility, 81(1), 23-28.

Gardner, D. K., & Lane, M. (1996). Fertilization and early embryology: alleviation of the ‘2-cell block’and development to the blastocyst of CF1 mouse embryos: role of amino acids, EDTA and physical parameters. Human Reproduction, 11(12), 2703-2712.

Goodeaux, L. L., Anzalone, C. A., Voelkel, S. A., Menezo, Y., Thibodeaux, J. K., & Roussel, J. D. (1990). Nonsurgical collection and transfer of rhesus monkey embryos. Theriogenology, 33(1), 234.

Heyn, P., Kircher, M., Dahl, A., Kelso, J., Tomancak, P., Kalinka, A. T., & Neugebauer, K. M. (2014). The earliest transcribed zygotic genes are short, newly evolved, and different across species. Cell reports, 6(2), 285-292.

Ireland, Z., Castillo-Melendez, M., Dickinson, H., Snow, R., & Walker, D. W. (2011). A maternal diet supplemented with creatine from mid-pregnancy protects the newborn spiny mouse brain from birth hypoxia. Neuroscience, 194, 372-379.

Ireland, Z., Dickinson, H., Snow, R., & Walker, D. W. (2008). Maternal creatine: does it reach the fetus and improve survival after an acute hypoxic episode in the spiny mouse (Acomys cahirinus)?. American Journal of Obstetrics & Gynecology, 198(4), 431-e1.

Kojima, Y., Tam, O. H., & Tam, P. P. (2014). Timing of developmental events in the early mouse embryo. In Seminars in cell & developmental biology (Vol. 34, pp. 65-75). Academic Press.

LaRosa, D. A., Ellery, S. J., Parkington, H. C., Snow, R. J., Walker, D. W., & Dickinson, H. (2016). Maternal creatine supplementation during pregnancy prevents long-term changes in diaphragm muscle structure and function after birth asphyxia. PloS one, 11(3), e0149840.

Latham, K. E. (1999). Mechanisms and control of embryonic genome activation in mammalian embryos. In International review of cytology (Vol. 193, pp. 71-124). Academic Press.

Lawitts, J. A., & Biggers, J. D. (1991). Overcoming the 2-cell block by modifying standard components in a mouse embryo culture medium. Biology of reproduction, 45(2), 245-251.

Lee, M. T., Bonneau, A. R., & Giraldez, A. J. (2014). Zygotic genome activation during the maternal-to-zygotic transition. Annual review of cell and developmental biology, 30, 581-613.

Liu, Z., Cai, Y., Wang, Y., Nie, Y., Zhang, C., Xu, Y., Zhang, X., Lu, Y., Wang, Z., Poo, M. and Sun, Q. (2018). Cloning of macaque monkeys by somatic cell nuclear transfer. Cell, 172(4), 881-887.

McLaren, A., & Biggers, J. D. (1958). Successful development and birth of mice cultivated in vitro as early embryos. Nature, 182(4639), 877.

Miyoshi, K. (2016). Development of a culture medium for rat 1-cell embryos. Journal of Mammalian Ova Research, 33(1), 11-16.

Morbeck, D. E. (2017). Mouse Embryo Assay for Quality Control in the IVF Laboratory. Principles of IVF Laboratory Practice: Optimizing Performance and Outcomes, 69.

Nagaraj, R., Sharpley, M.S., Chi, F., Braas, D., Zhou, Y., Kim, R., Clark, A.T. and Banerjee, U. (2017). Nuclear localization of mitochondrial TCA cycle enzymes as a critical step in mammalian zygotic genome activation. Cell, 168(1-2), 210-223.

Obata, R., & Tsujii, H. (2010). Valine Metabolism of Mongolian Gerbil (Meriones unguiculatus) Embryos. Journal of Mammalian Ova Research, 27(3), 144-149.

O'Connell, B. A., Moritz, K. M., Roberts, C. T., Walker, D. W., & Dickinson, H. (2011). The placental response to excess maternal glucocorticoid exposure differs between the male and female conceptus in spiny mice. Biology of reproduction, 85(5), 1040-1047.

O'Connell, B. A., Moritz, K. M., Walker, D. W., & Dickinson, H. (2013a). Sexually dimorphic placental development throughout gestation in the spiny mouse (Acomys cahirinus). Placenta, 34(2), 119-126.

O'Connell, B. A., Moritz, K. M., Walker, D. W., & Dickinson, H. (2013b). Treatment of pregnant spiny mice at mid gestation with a synthetic glucocorticoid has sex-dependent effects on placental glycogen stores. Placenta, 34(10), 932-940.

Pasco, R., Gardner, D. K., Walker, D. W., & Dickinson, H. (2012). A superovulation protocol for the spiny mouse (Acomys cahirinus). Reproduction, Fertility and Development, 24(8), 1117-1122.

Quinn, T. A., Ratnayake, U., Castillo-Melendez, M., Moritz, K. M., Dickinson, H., & Walker, D. W. (2014). Adrenal steroidogenesis following prenatal dexamethasone exposure in the spiny mouse. Journal of Endocrinology, 221(2), 347-362.

Quinn, T. A., Ratnayake, U., Dickinson, H., Castillo-Melendez, M., & Walker, D. W. (2016). The feto-placental unit, and potential roles of dehydroepiandrosterone (DHEA) in prenatal and postnatal brain development: a re-examination using the spiny mouse. The Journal of steroid biochemistry and molecular biology, 160, 204-213.

Quinn, T.A., Ratnayake, U., Dickinson, H., Nguyen, T.H., McIntosh, M., Castillo-Melendez, M., Conley, A.J. and Walker, D.W. (2013). Ontogeny of the adrenal gland in the spiny mouse, with particular reference to production of the steroids cortisol and dehydroepiandrosterone. Endocrinology, 154(3), 1190-1201.

Ratnayake, U., Quinn, T. A., Castillo-Melendez, M., Dickinson, H., & Walker, D. W. (2012). Behaviour and hippocampus-specific changes in spiny mouse neonates after treatment of the mother with the viral-mimetic Poly I: C at mid-pregnancy. Brain, behavior, and immunity, 26(8), 1288-1299.

Ratnayake, U., Quinn, T., Daruwalla, K., Dickinson, H., & Walker, D. W. (2014). Understanding the behavioural phenotype of the precocial spiny mouse. Behavioural brain research, 275, 62-71.

Schini, S. A., & Bavister, B. D. (1988). Two-cell block to development of cultured hamster embryos is caused by phosphate and glucose. Biology of reproduction, 39(5), 1183-1192.

Walker, S. K., Heard, T. M., & Seamark, R. F. (1992). In vitro culture of sheep embryos without co-culture: successes and perspectives. Theriogenology, 37(1), 111-126.

Whitten, W. K. (1956). Modification of the oestrous cycle of the mouse by external stimuli associated with the male. Journal of Endocrinology, 13(4), 399-404.

Whitten, M. K. (1957). Effect of exteroceptive factors on the oestrous cycle of mice. Nature, 180(4599), 1436.

Whitten, W. K., & Biggers, J. D. (1968). Complete development in vitro of the pre-implantation stages of the mouse in a simple chemically defined medium. Journal of reproduction and fertility, 17(2), 399-401.

Whitten, W. K., Bronson, F. H., & Greenstein, J. A. (1968). Estrus-inducing pheromone of male mice: transport by movement of air. Science, 161(3841), 584-585.

Xie, D., Chen, C.C., Ptaszek, L.M., Xiao, S., Cao, X., Fang, F., Ng, H.H., Lewin, H.A., Cowan, C. and Zhong, S. (2010). Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. Genome research, 20(6), 804-815.

Zhang, J., Zhao, J., Dahan, P., Lu, V., Zhang, C., Li, H., & Teitell, M. A. (2018). Metabolism in Pluripotent Stem Cells and Early M ammalian Development. Cell metabolism, 27(2), 332-338.