Emergence and Evolution of Placentation in Mammals

Alysha Shirley Taylor

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

School of Biology

June 2020

The candidate confirms that the work submitted is his/her/their own and that appropriate credit has been given where reference has been made to the work of others. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement"

Acknowledgements

I could not have gotten to this point without the support of so many wonderful people over the course of my PhD.

To my parents, Linda and Noel, who always supported me, encouraged my love of learning and kept me calm throughout, thank you. Your love and wisdom brought me to this point and kept me going. You are the source of my strength. Thank you 697211 times for making me who I am today. To my Nana Phil, who has always been an inspiration, thank you for all your love. You taught me what a strong woman was.

To my wonderful supervisors, Dr Niamh Forde and Dr Mary O'Connell: thank you for being such wonderful examples of scientific brilliance. They have each been central to my development as a scientist. Thank you, Mary, for introducing me to computational biology and the wonderful world of molecular evolution, for encouraging my growth as a researcher and helping me think outside the box. Thank you, Niamh, for teaching me to keep pushing on through failed experiments, for patiently improving my critical thinking skills and giving me the experience of bringing my computational work to the bench. Each of you pushed me to my full potential and I am so grateful to have had the opportunity to work with you.

To my colleagues in the Computational and Molecular Evolutionary Biology Group, thank you for the chats, the coffee and the laughs. Thank you for allowing me to bounce ideas off you and share frustration with misbehaving programs. To my colleagues in the Forde group, thank you for the cakes, for the laughs, for the commiserations over failed PCRs and the troubleshooting advice.

Sincere thanks to Dr Lynn McKeown for her kind gift of murine tissue. Sincere thanks as well to all those in LICAMM for their support and friendship.

Abstract

The placenta emerged on the stem Eutherian mammal lineage ~93 million years ago and defines the Eutherian clade. While placentation has arisen multiple times in *Animalia,* the evolution of mammal placenta is a unique and unreversed event. Unlike other organs, the mammal placenta varies in morphology, influenced by the life history and reproductive methods of the organism. Due to the highly dynamic nature of its development, the evolution of placental tissue was likely facilitated by changes to both protein coding and regulatory regions of the genome. This thesis examines (i) large scale genomic analyses, and (ii), expression analyses in *in vitro* models of early pregnancy events, to explore the contribution of positive selective pressure and novel miRNAs to the evolution of the placenta.

Using a comparative genomics approach, single gene orthologous families were analysed for heterogeneity in selective pressure. In total 115 conserved gene families were identified as having undergone positive selection on the stem eutherian lineage. Through the use of curated database sources and high-quality vertebrate genomes, we identified 112 miRNA families that emerged in therian mammals with various retention patterns throughout extant placental mammals. Of these, there were 14 that were gained on the therian or eutherian node and were subsequently never lost in any descendant lineage - two on the therian stem lineage, and a further 12 on the eutherian stem lineage. Targets of these 14 miRNAs were predicted using TargetScan and are implicated in uterine remodelling, cell proliferation and angiogenesis. These 14 miRNAs have targets in 84/115 positively selected genes. An examination of the binding sites for these 14 miRNAs in genes that underwent positive selection on the stem eutherian lineage found a greater number of binding sites for stem lineage miRNAs than would be expected by chance. This indicates that these elements are part of a novel regulatory network that emerged in eutherian mammals.

Expression of these miRNAs was detected in reproductive tissues from 5 different species. Treatment of bovine or human endometrial cells *in vitro* with molecules important for early pregnancy (progesterone, Interferon Tau, macrophage capping protein and protein disulfide-

isomerase) altered the expression of members of 11 of the conserved miRNA families. Two conserved genes (mannosyl-oligosaccharide glucosidase and NSE1 homolog, SMC5-SMC6 complex component) that underwent positive selection on the stem eutherian lineage were significantly downregulated in response to progesterone treatment.

Taken together, these analyses yield important insights into innovation in protein coding and regulatory elements that may have driven the evolution of mammal placenta. Using a comparative genomics approach, conserved miRNAs and genes that underwent positive selection on the stem eutherian lineage were identified and found to be regulated by key molecules necessary for early pregnancy *in vitro.*

TABLE OF CONTENTS

CHAPTER 1

Abbreviations

°C	degrees Celcius
3β-HSD	3β-hydroxysteroid dehydrogenase
AA	Amino acids
ABAM	Antibiotic Antimycotic Solution
ACD	ACD shelterin complex subunit and telomerase recruitment factor
ADAR	adenosine deaminase acting on RNA
Adh	Alcohol dehydrogenase
Ago2	Argonaute-2
AIC	Akaike Information Criterion
ALK2	activin receptor-like kinase-2
AR	acrosomal reaction
ARL13B	ADP ribosylation factor like GTPase 13B
ASPM	Microcephaly-associated gene
ATP	Adenosine triphosphate
BCL-2	B-cell lymphoma 2
BEB	Bayes empirical Bayes
BIC	B-cell Integration Cluster
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
BS-REL	Branch-Site Random Effects Likelihood
BSA	Bovine serum albumins
C17orf75	chromosome 17 open reading frame 75
χ^2	Chi-squared
C5orf51	chromosome 5 open reading frame 51
C7orf26	chromosome 7 open reading frame 26
cAMP	Cyclic adenosine monophosphate
CAPG	macrophage-capping protein
CCDC22	Coiled-coil domain containing 22
CCR4	carbon catabolite repressor protein 4
CDK	cyclin-dependent kinases
CDK5RAP2	CDK5 Regulatory Subunit Associated Protein 2
CDKN1A	Cycline-dependant kinase inhibitor 1A
cDNA	complementary DNA
CEA	carcinoembryonic antigen
CEBPB	CCAAT Enhancer Binding Protein Beta
CGB7	hCGbeta6 gene
cGMP	Cyclic guanosine monophosphate
CL	corpus luteum

CLSPN	Claspin
CNE	conserved non-coding elements
CO_2	Carbon dioxide
COMMD5	COMM domain containing 5
COX2	cyclooxygenase 2
CREB	cAMP response element binding protein
CRISPs	Cysteine-Rich Secretory Proteins
CRL4B	Cullin4B-RING E3 ligase complex
CRMP	Collapsin response mediator protein 1
Ct	cycle number where the fluorescence of the reaction crosses the threshold
CUL4B	Cullin 4B
CX3CL1	Fractalkine
DCAF11	DDB1 and CUL4 associated factor 11
⊗· ⎮	Ct value of miRNA less the average Ct of normaliser genes
⊗⊗· ⎮	⊗· ⎮ ⎮⎲ ⎴⎡∇⊆ℑ⎣∑ ∫ ∫ ⎮⎮∑ ⟨ϖ∑⟩⟨©∑ · ⎮ ⎮⎲ ⎮⎮∑ ϖ∑⎮⎡ ⅄∑ ⎮⎮⎨ ⎮⟩⎮⎣
⊗⎣	Difference between lnL scores of two models
DMEM/F-12	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12
dN	Nonsynonymous nucleotide substitutions per nonsynonymous site
DNA	Deoxyribonucleic acid
DPH5	diphthamide biosynthesis 5
DPP4	dipeptidyl peptidase 4 tumor suppressor gene
dS	Synonymous nucleotide substitutions per synonymous site
E-value	Expect value
E2F1	E2F transcription factor 1
ECSIT	ECSIT signalling interactor
EEPD1	Endonuclease/Exonuclease/Phosphatase Family Domain Containing 1
EGF	epidermal growth factor
EGFR	Endothelial growth factor receptor
EIF2AK4	Eukaryotic translation initiation factor 2-alpha kinase 4
EJC	Exon junction Complex
ERBB2	erb-B2 receptor tyrosine kinase 2
ERK	extracellular-signal-regulated kinase
ERVs	Endogeneous Retroviruses
ESR	Estrogen receptor alpha
EV	extracellular vesicle
EXPH5	exophilin 5
FBS	Fetal Bovine Serum
FGF2	fibroblast growth factor 2
FOXO1A	forkhead Box Protein O1
FOXP3	forkhead box protein 3

FSH	follicle-stimulating hormone
g	Gram
GAB1	GRB2-associated-binding protein 1
GC	Content of guanine and cytosine
GDF9	Growth differentiation factor 9
GFP	Green fluorescent protein
GLUT1	glucose transporter 1
GnRH	gonadotropin-releasing hormone
GO	Gene Ontology
GRB2	Growth factor receptor-bound protein 2
GRB7	Growth factor receptor-bound protein 7
GSP	Penicillin-Steptomycin-Glutamine
GTP	Guanosine-5'-triphosphate
GW182	glycine-tryptophan protein of 182 kDA
HBSS	Hanks Balanced Salt Solution
hCG	Human chorionic gonadotrophin
HDL	high density lipoprotein
HIRIP3	HIRA interacting protein 3
HLA-G	Histocompatability antigen
HNF1α	hepatocyte nuclear factor 1 alpha
HNF1β	hepatocyte nuclear factor 1 beta
HOXA11	homeobox A11
HOXA13	homeobox A13
HOXB8	Homeobox b8
HUVEC	Human umbilical vein endothelial cells
HyPhy	Hypothesis Testing Using Phylogenies
IFNA	Interferon alpha
IFNAR1/2	Interferon alpha/beta receptor
IFNB	Interferon beta
IFND	Interferon delta
IFNT	Interferon Tau
IFNW1	Interferon omega
IGF2	Insulin-like growth factor type-2
IGF2BP/IMP	Insulin-like growth factor 2 mRNA-binding proteins
IGF2R	Insulin-like growth factor type-2 receptor
IL-8	Interleukin 8
INPP5E	Inositol polyphosphate-5-phosphatase E
INSL4	insulin-like 4
ISG	Interferon stimulated gene
ISGF3	Interferon-stimulated gene factor 3
IVF	*In vitro* fertilization

JAK-STAT	Janus activated kinase-signal transducer and activator of transcription
JCAD	junctional cadherin 5 associated
kDA	kilo daltons
KEGG	Kyoto Encyclopedia of Genes and Genomes
L	Litre
L1	Larval stage 1
L1CAM	L1 cell adhesion molecule
L2	Larval stage 2
LBD	ligand-binding domain
LDL	low density lipoprotein
LGALS15	Lectin, Galactoside-binding, soluble, 15
LGI-ADAM	Leucine-rich glioma inactivated-a disintegrin and metalloprotease
LH	luteinizing hormone
LIN-14	abnormal cell LINeage 14
lin-4	abnormal cell LINeage 4
LNA	Locked nucleic acids
lnL	Log likelihood
LNX3	ligand of number-protein x 3
LRT	Likelihood-ratio test
LTR	Long terminal repeat
MAFFT	Multiple Alignment using Fast Fourier Transform
MAPK	mitogen-activated protein kinase
MASTL	micro tubule associated serine/threonine kinase like
MBTPS2	membrane bound transcription factor peptidase site 2
MDM1	Mdm1 nuclear protein
MECP2	methyl CpG binding protein 2
MECP2	methyl CpG binding protein 2
MET	hepatocyte growth factor receptor
⌠g	Microgram
mg	Milligram
miRISC	minimal miRNA-induced silencing complex
miRNA	microRNA
miRNAa	miRNA-induced RNA activation
miRQC	miRNA quality control study
mL	Millilitre
⌠l⌡	Microlitre
⌠m⊂	micro meters
MMP	matrix metalloprotease
MMR	Mismatch repair
MOGS	mannosyl-oligosaccharide glucosidase

MOGS-CDG	mannosyl-oligosaccharide glucosidase congenital disorder of glycosylation
MREs	miRNA response elements
mRNA	messenger RNA
MRP	maternal recognition of pregnancy
MSA	multiple sequence alignment
MUC1	mucine 1, cell surface associated promotor
MUSCLE	Multiple Sequence Comparison by Log-Expectation
MYA	Million years ago
MYDGF	myeloid derived growth factor
N4BP3	NEDD4 binding protein 3
NDNL2	NSE3 homolog, SMC5-SMC6 Complex Component
N_e	Effective Population Size
NF-κB	Nuclear Factor kappa-light-chain-enhancer of activated B cells
NFRKB	Nuclear factor related to kappa-B-binding protein
NK	Natural Killer cells
NMD	Nonsense mediated decay
NMDA	N-methyl-D-aspartate receptor
Nmol	Nanomolar
NOP9	NOP9 nucleolar protein
NOXA	Phorbol-12-myristate-13-acetate-induced protein 1
NrCAM	Neuronal cell adhesion molecule
NSE1	non-structural maintenance of chromosomes element 1
NSMCE1	NSE1 homolog, SMC5-SMC6 Complex Component
NSMCE2	NSE2 (MMS21) homolog, SMC5-SMC6 Complex Component
NSMCE4	non-structural maintenance of chromosomes element 4
nt	Nucleotide
NUP88	nucleoporin 88
O_2	Oxygen
P4	progesterone
P450scc	cytochrome P450 side-chain cleavage enzyme
P4HB	protein-disulphide isomerase
PABPC	polyadenylate-binding protein
PACT	protein activator of protein kinase RNA
PAML	Phylogenetic Analysis by Maximum Likelihood
PAN2	poly(A)-nuclease deadenylation complex subunit 2
PAN3	poly(A)-nuclease deadenylation complex subunit 3
Pax6	paired box proteina
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PG	prostaglandin

PGF	prostaglandin F
PGF2α	prostaglandin F2α
PGRA	progesterone receptor A
PGRB	progesterone receptor B
Ph	polyhomeotic
PI3K	phosphatidylinositol-3-kinase
PLAG1	zinc finger pleomorphic adenoma gene
PRL	Prolactin
PRX1	Peroxiredoxin
PTK6	Protein-tyrosine kinase 6
RAB	Ras-associated binding protein
RABAC1	Rab acceptor 1
RB1	retinoblastoma protein
RHO	Ras homologous protein
RIF	Recurrent implantation failure
RIPK1	Receptor (TNFRSF)-Interacting Serine-Threonine Kinase 1
RISC	RNA-Induced Silencing Complexes
RLC	RISC loading complex
RLT1	Retrotransposon-like 1
RNA	Ribonucleic acid
RP2	RP2 activator of ARL3 GTPase
RPL	Recurrent pregnancy loss
rRNA	Ribosomal ribonucleic acid
RT	Reverse transcription
RT-qPCR	Quantitative reverse transcription PCR
RUNX1	runt-related transcription factor 1
RUNX3	runt-related transcription factor 3
Sema3A	Semaphorin 3A
Sema4D	Semaphorin 4D
SGK1	serum/glucocorticoid regulated kinase 1
SGO	Single Gene Orthologue
SHC1	Beta-ketoacyl synthase
Siglecs	Sialic acid-binding immunoglobulin-type lectins
siRNA	Small interfering RNA
SLR	Sitewise likelihood ratio
SMC5	Structural maintenance of chromosome protein 5
SMC6	Structural maintenance of chromosome protein 6
STAT	signal transducer and activator of transcription
STAT3	signal transducer and activator of transcription 3
SUMO	small ubiquitin-like modifiers
TCFL5	transcription factor like 5

TERT	telomerase reverse transcriptase
TEs	Transposable Elements
TFAP2	Transcription factor AP-2 alpha
TGF-β	transforming growth factor beta
TGFBR1	Transforming Growth Factor Beta Receptor 1
TICAM1	Toll Like Receptor Adaptor Molecule 1
T_m	Melting temperature
TMEM186	Transmembrance protein 186
TNF	tumour necrosis factor
TNFR1	tumor necrosis factor receptor 1
TNS	Tensin
TNT	Tree Analysis Using New Technology
TP53	Tumor protein p53
TRADD	Tumor Necrosis Factor Receptor Type 1-Associated DEATH Domain Protein
TRBP	Tar RNA binding protein
TSR	Thrombospondin type 1 repeat
TYK2	tyrosine kinase 2
UK	United Kingdom
USA	United States of America
USPL1	ubiquitin specific peptidase like 1
UTR	untranslated region
v/v	Volume for volume
VEGF	vascular endothelial growth factor
VEGFR	vascular endothelial growth factor receptor
VESPA	Very large-scale Evolutionary and Selective Pressure Analyses
⎤	Rate of non-synonymous substitution per non-synonymous site to Synonymous substitution per synonymous site
WNT5A	wnt Family Member 5A
WOI	window of implantation
WRAP53	WD repeat containing antisense to TP53
xg	Times gravity
Xist	X-inactive specific transcript
ZP	zona pellucida sperm-binding protein

List of Figures

List of Tables

Chapter 1: Introduction

Evolution of mammal placenta: Current hypotheses

Placentation in mammals evolved c. 93 MYA with the divergence of *Eutheria* from the subclass *Theria* (Tarver *et al.*, 2016). Eutherian placenta is an evolutionary novelty, i.e. not homologous to any organ in the ancestral lineage (Wagner and Lynch, 2010). It is distinct from the short-lived yolk-sac placenta present in *Theria* in that eutherian placenta is highly vascularised and facilitates physiological exchange for the entire pregnancy through its attachment to the maternal endometrium (Freyer and Renfree, 2009). Mammal placenta can be defined as 'an approximation or combination of an embryo's tissues with those of its natural or surrogate parent for physiological interchange' (Mossman, 1987). The function of the placenta is to facilitate the exchange of nutrients, gases and waste between mother and foetus to enable foetal development (Allen C. Enders and Blankenship, 1999) along with the provision of maternal antibodies; oxygen; glucose, amino acids and water; and hormones (Griffiths and Campbell, 2015).

1.1.1 Placental mammal phylogeny

The placental mammal clade (*Eutheria*), diverged from *Marsupialia* c. 93 MYA (million years ago) (Tarver *et al.*, 2016). After the mass extinction at the end of the Cretaceous period, *Eutheria* underwent adaptive radiation, adjusting to novel ecological niches (Lillegraven *et al.*, 1987). This period of diversification has led to morphological variation within eutherian mammals, with extant members including *Craseonycteris thonglongyai* (Kitti's hog-nosed bat), the smallest living mammal weighing approximately 1.5-2.0g; and the largest living mammal - *Balaenoptera musculus* (the blue whale), which weighs approximately 150 tonnes (Behringer, Eakin and Renfree, 2005). Diversity in *Eutheria* includes metabolic rate, inversely correlated with body mass, which has been reported to vary between eutherians up to a factor of 3:1-10:1 (McNab, 2008). This variability may be explained by factors other than body mass which is thought to influence metabolic rate to the greatest extent. These other factors include food source and climate, factors which fluctuate across *Eutheria*. Further life history trait variation within the eutherian clade includes germ-line generation time, with ages of reproductive maturity and

length of gestation positively correlating with body mass. For example in mice reproductive maturity occurs at 1-2 months, and has a gestation length of ~20, humans reach reproductive maturity at 13-14 years, with ~37-42 weeks gestation, while elephants reach reproductive maturity at ~14 years, and have a gestation period of ~22-24 months (Behringer, Eakin and Renfree, 2005; Rasier *et al.*, 2006; Garratt *et al.*, 2013).

The diversity of these traits across *Eutheria* have caused complications in the analysis of eutherian evolution. It has been observed that smaller body size (and by extension, higher metabolic rate and faster generation time) correlates with a higher synonymous substitution rate in both nuclear and mitochondrial DNA and a higher nonsynonymous substitution rate in nuclear DNA (Welch, Bininda-Emonds and Bromham, 2008). It is thought that this may be due to errors accumulated as the genome is copied each generation, with smaller mammals producing more generations per unit time; or induced by increased levels of DNA damage due to the production of metabolic by-products in species with higher metabolic rates (Bromham, 2011). This higher rate of mutation has also been linked to increased GC content in the genome as a result of DNA damage repair activities and the well known mammal-linked phenomenon of biased gene conversion, resulting in variable sequence composition within *Eutheria* (Romiguier *et al.*, 2010). The variability of mutation rate and sequence composition therefore complicates phylogenetic studies of the eutherian clade, requiring that models of evolution can accommodate variation of substitution rate and composition across both sites and lineages.

This is exemplified by the efforts to resolve the position of the root of the placental mammal phylogeny (Teeling and Hedges, 2013). Three major competing hypotheses proposed for the position of the placental mammal root are shown in Figure 1.1. It is uncontested that placental mammals (*Eutheria*) as an order diverged from *Marsupiala* c.93 million years ago (MYA) (Tarver *et al.*, 2016). However, it is the identity of the first diverging eutherian branch which has proved contentious. The Afrotherian position places elephants and aardvarks as the earliest diverging branch of placental mammals (Murphy *et al.*, 2001) (Figure 1.1A). The Xenarthran position placed armadillos and sloths as the earliest diverging branch in the placental mammals (Nishihara, Okada and Hasegawa, 2007) (Figure 1.1B). Finally, the Atlantogenata hypothesis proposed that the

group containing Xenarthra and Afrotheria is the sister to all other placental mammals (Hallström *et al.*, 2007; Morgan *et al.*, 2013; Tarver *et al.*, 2016) (Figure 1.1C). A number of factors contributed to this conflict, including the application of homogeneous models to heterogeneous data and poor data quality. The homogeneous models cannot accommodate differences in rates of mutation across the data and in separate branches of the phylogenetic tree. Considering the variation in mutation rate and observed diversity across the eutherian clade discussed above, homogeneous models are therefore not biologically realistic (Morgan *et al.*, 2013). Using a data driven approach and heterogeneous models (Morgan *et al.*, 2013; Moran, Morgan and O'Connell, 2015) – progress towards a resolved placental mammal tree has been made and the position of the earliest diverging placental mammal lineages has been resolved (Morgan *et al.*, 2013; Moran, Morgan and O'Connell, 2015; Tarver *et al.*, 2016). Given current data and models, the Atlantogenata position is the favoured position for the root of the placental mammal phylogeny and is therefore the phylogeny applied throughout this thesis.

While the root of the placental mammal phylogeny is potentially resolved, complete resolution of the full eutherian phylogeny is ongoing. The position of taxa within *Laurasiatheria*, for example, has proven challenging due to insufficient taxon sampling and the rapid divergence of the clade (Hallström and Janke, 2008; Chen, Liang and Zhang, 2017). With amount of available data increasing, along with the continual improvements in sequencing depth and quality, progress towards the resolution of the major orders within *Laurasiatheria* has been made (Upham, Esselstyn and Jetz, 2019), though the position of single species within these orders is contentious (Hawkins *et al.*, 2019). The phylogeny used throughout this thesis is based on the Atlantogenata position for the placental mammal root as is supported by current data and heterogeneous models of evolution, and we use species representative of each of the eutherian clades avoiding the inclusion of species whose placement is contentious.

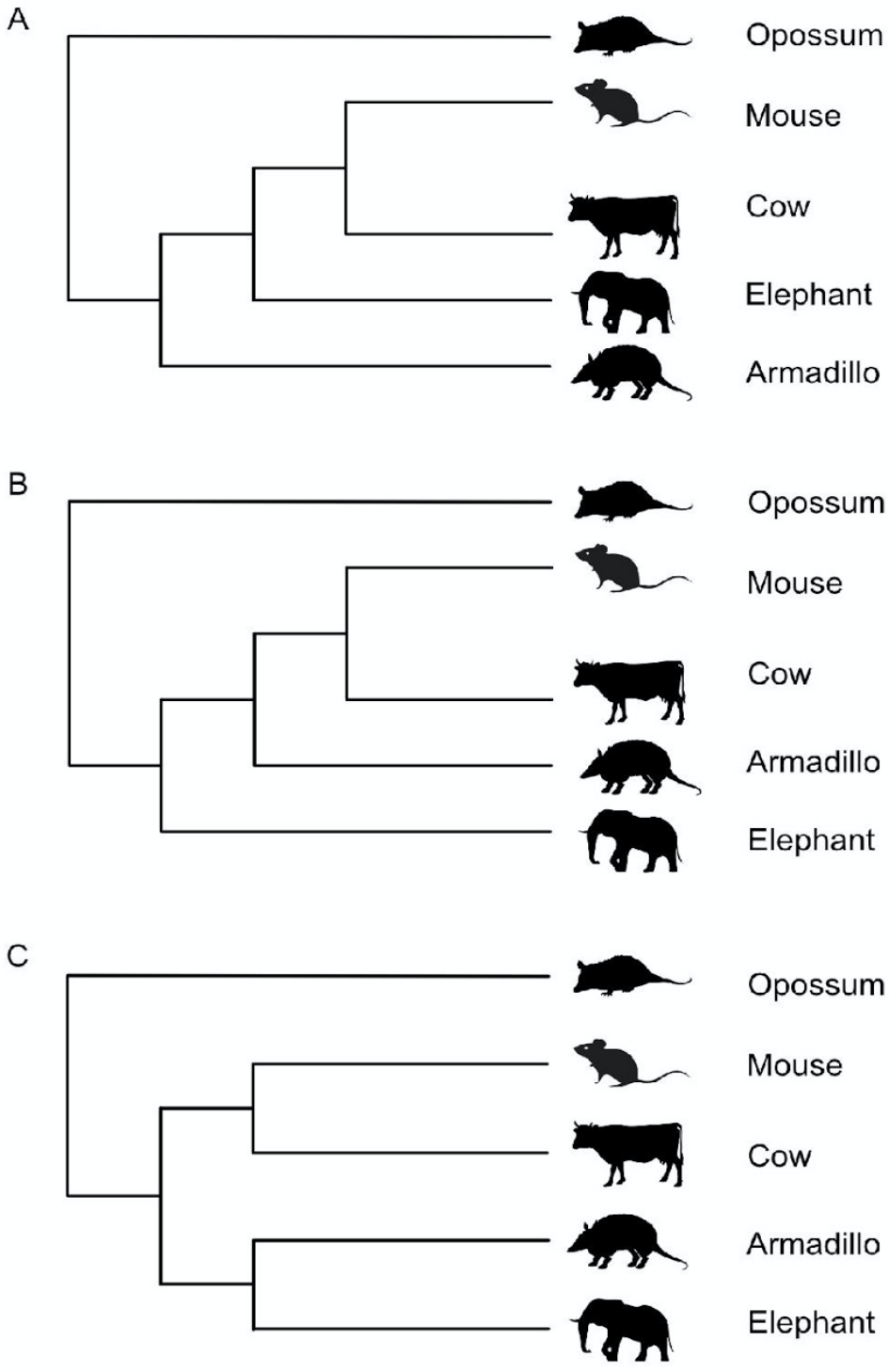


Figure 1.1 A depiction of the three competing hypotheses for the position of the root of the placental mammal tree*. In each case Opossum is the uncontested outgroup species.* ***(A)*** *represents the Xenarthra hypothesis,* ***(B)*** *shows the Afrotheria hypothesis and* ***(C)*** *depicts the Atlantogenata hypothesis used in this thesis.*

1.1.2 Current theories on the emergence of novel traits

The placenta is an individualised tissue with a specialised function. Studies that take into account the evolution and development of novel tissues/organs have made significant progress in recent years for understanding organs and tissues such as vision, flight and mammary glands (Wagner and Lynch, 2010). In the context of this thesis when discussing novel traits, also called evolutionary novelties, we are referring to the definition from (Wagner and Lynch, 2010):

> *"A novel body part that is neither homologous to any body part in the ancestral lineage nor serially homologous to any other body part of the same organism."*

However, many questions remain about the molecular mechanisms and genomic changes that are necessary for new tissues to arise. Genes related to phenotypic innovation in animal clades may have existed prior to the emergence of the novel trait and may have been co-opted into a new role; such as paired box protein *(Pax6),* a bilaterian gene that arose in Metazoa and is involved in body patterning in *Bilateria* and in eye development in vertebrates (Shubin, Tabin and Carroll, 2009; Eddy, 2012). Therefore, protein coding alterations (such as selective pressure variation, co-option and gene loss), along with innovations in regulatory networks drive the development of novel traits (Wagner, 2015).

These networks have been implicated in heart development, with a set of core transcription factors found to regulate gene expression during heart formation in both vertebrates and *Drosophila* (Davidson and Erwin, 2006). Transcription factor networks have also been found to regulate gene expression during endoderm formation in echinoderm embryonic development, with a shared set of regulatory elements found across the echinoderm phylogeny (Hinman *et al.*, 2003). Progress has been made on the role of regulatory innovation in the emergence of eutherian pregnancy. Decidualisation is an uncommon process of morphological changes to the endometrial stromal cells to prepare for invasive placentation (Okada, Tsuzuki and Murata, 2018). Prolactin (*PRL*), a marker of decidualisation, was found to be regulated by 3 transcription factors – forkhead box protein O1 *(FOXO1A),* homeobox A11 *(HOXA11)* and CCAAT enhancer

binding protein beta *(CEBPB)*. The cooperative interaction between *FOXO1A* and *HOXA11,* and, *FOXO1A* and *CEBPB,* evolved on the stem eutherian lineage, with their interaction promoting *PRL* expression and facilitating decidualisation prior to the formation of the invasive placenta (Kin *et al.*, 2016). Further investigation into transcriptional rewiring in mammalian pregnancy discovered 2218 genes which evolved endometrial expression on the stem mammalian lineage, with 835 genes found to have been co-opted into the endometrium on the eutherian stem lineage. This alteration in gene expression was found to be mediated by transposable elements (TEs) found in the *cis*-regulatory elements of these 'recruited' genes. In addition, these TEs were found to be incorporated into progesterone receptor binding sites, indicating that TE-mediated alterations of gene regulatory networks may have enabled the evolution of the uterine progesterone response during early pregnancy (Lynch *et al.*, 2015).

Therefore, alterations in gene expression induced by changes in regulatory networks can influence development of novel morphologies in animals and in particular have been shown to have a role in the emergence of eutherian pregnancy. Considering this, in order to gain a better understanding of alterations in molecular elements that contributed to the evolution of mammal placenta, it is necessary to investigate adaptations in both protein coding and regulatory elements.

1.1.3 Role of protein coding elements in the evolution of novel traits

Comparative genomics has taught us that morphological diversity does not reflect similarly diverse protein coding regions of the genome (Sanetra *et al.*, 2005). Furthermore, larger genome size does not correlate with greater morphological complexity (Eddy, 2012). Therefore it has been proposed that morphological complexity emerges via a number of mechanisms/processes, including, but not limited to novel gene emergence. Gene/regulatory element co-option, adaptation of genes and regulatory elements to novel functions, gene/regulatory element loss all contribute to the genomic landscape underpinning phenotypic change. Some examples of protein coding innovation and phenotypic evolution are discussed below.

Multiple mechanisms of protein coding innovation have contributed to the evolution of the radula, the primary feeding organ of *Mollusca*. Using a phylostratigraphic approach, a significantly higher number of genes involved in radula development and function arose on the stem *Mollusca* lineage than in parent lineages. Furthermore, genes involved in mollusc foregut and shell development, such as chitin synthase and hephaestin, were found to be co-opted, developing radula-specific expression (Hilgers *et al.*, 2018). Gene co-option has also been implicated in morphological diversity in *Mollusca*, with evidence for species-specific co-option of genes that assist in the development of shell structure (Aguilera, McDougall and Degnan, 2017).

The insect chemosensory system evolved through multiple protein coding alterations. Expansion and contraction of genes within the chemosensory system has been observed in multiple ant species (Zhou *et al.*, 2012), *Drosophila* (Robertson, Warr and Carlson, 2003) and the waterflea *Daphnia pulex,* (Peñalva-Arana, Lynch and Robertson, 2009) via gene birth and death. These processes of gene birth and death have led to varying populations of chemosensory genes within various *Arthropoda*, indicating a dynamic evolutionary history (Benton, 2015). Further study of these chemosensory genes in *Drosophila* has also found evidence of positive selection on recently duplicated genes (Almeida *et al.*, 2014), suggesting that this family of sensory receptors evolved through multiple alterations to protein coding genes.

The vertebrate eye is another novel organ which required multiple mechanisms of protein coding alterations. Two co-option events have been discovered, involving the α- and β-crystallins, proteins that are expressed in the lens and assist in light refraction. Using sequence homology searches, α-crystallin was found to have high homology to vertebrate heat shock proteins, while β-crystallin was highly homologous to microbial stress proteins. These two co-option events of ancestral stress-related proteins into the vertebrate eye may indicate a new role for these proteins in protecting the lens of the vertebrate eye from excessive light-related damage (True and Carroll, 2002). There is support for the role of positive selection in the emergence of the vertebrate eye. Morphological modelling of eye evolution which incorporated gradual beneficial adaptations to flat, light-sensitive epithelial cells (modelling the most primitive ‘eye’) that increased the level of spatial information processed have shown that positive selective pressure

induced the evolution of the vertebrate eye. When using the most minimal selection intensity, morphological models indicate that the vertebrate eye may have emerged via positive selective pressure in c. 400,000 years (Nilsson and Pelger, 1994).

This link between genes displaying signatures of positive selection and morphological development was further established in a transcriptomic study of gene expression alterations during the development of seven organs (brain, cerebellum, heart, kidney, liver, ovary and testis) across *Eutheria.* During the development of each organ, the proportion of expressed genes with signatures of positive selection increased as morphology became more complex, leading the authors to hypothesize that positively selected genes may drive morphological divergence during organ development and thus may have contributed to the evolution of these complex organs themselves (Cardoso-Moreira *et al.*, 2019).

Protein coding innovations have been investigated in relation to the emergence of mammalian pregnancy and the eutherian placenta. Insulin-like 4 *(INSL4),* also known as placentin, emerged via successive gene duplications of relaxin-like genes early in eutherian evolution, and it is expressed in cytotrophoblast and syncytiotrophoblast cells, which are the inner and outer layers of the trophoblast, respectively. *INSL4* orthologues have been found throughout placental mammals, with all *INSL4* genes undergoing inactivation in all placental mammals except for catarrhine primates, indicating a specific role in placentation and reproduction in Old World monkeys (Arroyo *et al.*, 2012). Innovation in protein coding elements has also led to the emergence of non-coding elements. X-inactive specific transcript *(Xist),* a long non-coding RNA gene involved in X chromosome inactivation in placental mammals, was found to have partially evolved from the duplication of the vertebrate ligand of numb-protein x 3 (*LNX3*) gene, with *LNX3* later losing its function in *Eutheria*. However, only 3 *Xist* exons showed homology to the *LNX3* gene, indicating that the *Xist* gene may have arisen from multiple mechanisms of duplication and shuffling of existing gene loci to form the new gene (Duret *et al.*, 2006). The emergence of novel genes and their role in the evolution of placental mammals also has been examined on a whole-genome scale (Dunwell, Paps and Holland, 2017). Using a dataset of 20 genomes from *Eutheria*, *Marsupiala, Monotremata, Aves, Amphibia, Reptilia* and Fish, 357 gene families were found to

have arisen on the stem eutherian lineage, with 87 present in at least 9 of the 10 eutherian mammals analysed. Using Gene Ontology (GO) annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, these novel genes were found to function in immune response, spermatogenesis and keratinization. In order to further analyse the possible roles of these genes, gene expression profiling was used to identify distinct groups of genes expressed in pre- and post-implantation embryos, testes, fallopian tubes and mammary gland tissue, indicating roles for these novel genes in multiple traits specific to *Eutheria*.

Another example of protein coding innovation specific to *Mammalia* is genomic imprinting, a process by which maternal and paternal alleles of a gene are expressed at different levels. This phenomenon was first discovered in murine insulin-like growth factor type-2 receptor (*Igf2r),* which was observed to be entirely expressed by the maternal gene (Barlow *et al.*, 1991). Imprinted genes appear to function mainly during gestation, with the expression of many imprinted genes being downregulated post-natally (Bartolomei and Ferguson-Smith, 2011). *Igf2* itself is implicated in placental function, with a placenta-specific *Igf2* found to modulate nutrient transfer across the murine placenta (Sibley *et al.*, 2004). *Peg10* and *Peg11/Rtl1* are paternally expressed genes derived from long terminal repeat (LTR) retrotransposons that are necessary for placental development and maintenance, with knockout mice failing to form spongiotrophoblasts or a labyrinthine structure, affecting the placenta and feto-maternal communication (Kaneko-Ishino and Ishino, 2012). Paternally expressed genes 10 and 11 (*PEG10* and *PEG11*)*,* being derived from retrotransposons, are also examples of protein coding innovation via domestication from transposons, by which transposons are recruited as 'neogenes', enabling novel functions (Brandt *et al.*, 2005) Further evidence of domestication in placental evolution is the MER20-mediated recruitment of cAMP-signalling and progesterone-regulated genes into the stromal cell (Lynch *et al.*, 2011).

Endogenous Retroviruses (ERVs) are sequences composed of remnants of retroviruses (Johnson, 2019). ERVs have been shown to contribute novel genetic material in placental mammals and have been co-opted to function in cell fusion and the suppression of the maternal immune system

to allow placental invasion and are highly expressed in the placenta (Rote, Chakrabarti and Stetzer, 2004). In addition ERVs have been shown to function in initiation of the placental trophoblast cell (Chuong *et al.*, 2013), promoting trophoblast fusion during placental development. Well known examples of ERVs with placental roles in mammals include syncytin-1 and syncytin-2 in monkeys and apes (Esnault *et al.*, 2013), and syncytin-A and syncytin-B in rodents (Emera and Wagner, 2012). Though it is not known if ERV's only express in mammalian placenta, or are common to all animals with a primitive placenta, they are another crucial element in the analysis of how the eutherian placenta developed, invaded and was maintained.

Within mammalian reproductive organs, evidence of positive selection was found in *HOXA11* and Homeobox A13 (*HOXA13)* (Lynch *et al.*, 2004). *HOXA11* and *HOXA13* have previously been implicated in the development of the mammal reproductive tracts (Taylor, Vanden Heuvel and Igarashi, 1997). In addition, *HOXA13* has been found to be essential in the formation of umbilical arteries in mice (Stadler, Higgins and Capecchi, 2001). Evidence of strong positive selection in these genes, on the stem mammal lineage, suggests a role in the emergence of the mammalian reproductive tract and mammal pregnancy.

It is clear that major innovations in morphology require adaptations in the protein coding component of the genome - our thesis is that positive selection on the stem eutherian lineage was crucial in the evolution of the mammal placenta.

1.1.4 Role of regulatory elements in novel organ evolution

Alterations to the regulation of a gene have been shown to facilitate temporal and spatial variation in gene expression. As discussed above, TEs act to 'recruit' genes into endometrial expression during the evolution of eutherian mammals (Lynch *et al.*, 2015). These TEs were later domesticated into *cis*-regulatory elements in the endometrium. An example of a regulatory element that evolved in this way is MER20, a progesterone-responsive *cis*-regulatory element involved in cyclic adenosine monophosphate (cAMP) signalling during decidualisation (Lynch *et al.*, 2011, 2015). Fusion of three *cis*-regulatory elements (AmnSINE1, X6b_DNA and MER117) into one single conserved non-coding element (CNE) at the AS3_9 locus serves as a distal enhancer of

the signal transducer wnt family member 5A (*WNT5A*) during secondary palate formation in mammals (Nishihara *et al.*, 2016), this was a crucial innovation that enables concurrent breathing and eating. In addition to domestication of TEs and fusion of existing cis-regulatory elements, changes to the sequence of regulatory elements may also induce phenotypic alterations. Alterations in the sequence of peroxiredoxin 1 (*Prx1*) were found to have contributed to the elongation of the forelimb during bat development, facilitating the evolution of the bat wing. Examining the sequence variation in *Prx1* between bat and mouse, species with similar early limb development patterns, variations in the *cis*-regulatory region of the bat *Prx1* were found to promote forelimb elongation during development (Cretekos *et al.*, 2008)

Alterations in microRNAs and their targets may also impact on phenotype, as discussed in detail in section 1.3.

1.2 Selective Pressure and Molecular Evolution

Molecular evolution refers to the process of change in the sequence and structure of cellular molecules such as DNA, RNA and proteins over time. Mutation induces changes in the sequence. These changes may then be fixed or removed from the population depending on a number of related factors including selective pressure (either positive or negative) and the effective population size (N_e) (Lynch, 2007). Theoretically, in a finite and sexually reproducing population where no gene confers an advantage, allele frequencies may vary depending on which gametes are transmitted. The degree of fluctuation in allele frequencies may be due to decreasing N_e, an unequal sex ratio within the population or due to natural disasters disrupting the ecosystem and affecting mating (Nei and Tajima, 1981). In the same theoretical population, substitutions which do not confer an advantage/disadvantage are considered 'neutral' (Kimura, 1968). The neutral theory of evolution hypothesizes that the majority of genetic variance is due to the accumulation of neutral or nearly neutral mutations. Substitutions that confer a selective advantage are considered a rare occurrence (Nei, Suzuki and Nozawa, 2010) and become fixed in the population by impacting reproductive success (Mitchell-Olds, Willis and Goldstein, 2007), these are referred to as positively selected sites throughout this thesis. The neutral theory of evolution was later expanded to include 'nearly neutral' mutations, mutations that are slightly deleterious or slightly advantageous or nearly neutral (Kimura and Ohta, 1974; Ohta and Gillespie, 1996). Neutral and nearly neutral substitutions can become fixed in the population due to random genetic drift, while nearly neutral mutations are acted upon by selection to be fixed or removed from the genome (Kimura and Ohta, 1974).

1.2.1 Neutral theory of evolution

Under the neutral model, sequence variance within a population is the result of random genetic drift, rather than directional selective pressure. This variance fluctuates due to arbitrary mutation and extinction events (Kimura, 1968). The probability that these random substitutions become fixed within a population is described in the following equation (Equation 1.1).

$$P_x = \frac{1}{2N_e}$$

Equation 1.1: *The probability of the fixation of a random mutation in a population by genetic drift.*

As seen in the equation, the probability of a substitution becoming fixed within a population (P_x) is inversely proportional to the size of the population (N_e). Under the neutral model of evolution, the lack of selective pressure on alleles within a population allows the accumulation of heterozygosity, which is directly proportional to N_e. The expected heterozygosity of a finite population at equilibrium is calculated using Equation 1.2, also known as the infinite allele model, where *N* refers to the effective population size and *u* is the neutral mutation rate (Kimura and Crow, 1964).

$$Expected\ heterozygosity = \frac{4N_u}{1 + 4N_u}$$

Equation 1.2: *The expected heterozygosity in a finite population in equilibrium between random drift and mutation.*

An upper limit of protein heterozygosity was established, with the median heterozygotic loci in an individual estimated to be 10.6% (Lewontin, 1974). Using this estimate of heterozygosity, the N_e required for the neutral theory to be true for a given population is between 10^4-10^5. Issues with this narrow population range were raised when examining enzyme heterozygosity in *Drosophila*, where the same alleles were observed at similar frequencies across populations (Ayala, Powell and Tracey, 1972). With an estimated N_e of 10^9, the observed heterozygosity was far lower than expected if the population was evolving neutrally. In this case, the 'nearly neutral' theory is posited to account for populations of a greater effective size, with slightly deleterious alleles being selectively removed (Ohta and Gillespie, 1996).

These 'nearly neutral' substitutions may become fixed within a population by random genetic drift (Nei, Suzuki and Nozawa, 2010). This hypothesis is built upon the observation of varying rates of evolution over a sequence, along with the King and Jukes' observation that functionally unimportant sites accumulate substitutions at a higher rate than functionally important sites (King and Jukes, 1969). Under this theory that functionally important sites are under constraint, with functionally unimportant sites accumulating substitutions randomly, that later become fixed via random genetic drift (Kimura and Ohta, 1974).

At that time, support for the neutral theory was built upon observations of high levels of substitutions occurring in a sequence over time without altering the central function of the gene, such as cytochrome c, which has been observed to be functionally interchangeable within mammalian cells (Jacobs and Sanadi, 1960). Further evidence for the neutral theory was found in early studies of the amino acid sequence of α- and β-hemoglobins, with the number of substitutions directly proportional to the amount of time since the divergence of lineages found between human α- and β-hemoglobin amino acid sequences and that of the Port Jackson shark, (Kimura, 1991). While the neutral theory predicts a level of substitution proportional to divergence time, levels of substitution beyond that predicted by the neutral theory may indicate directional evolution.

1.2.2 Natural selection

Natural selection can be described as the mechanism by which genetic diversity is maintained or removed from a population based upon its impact on reproductive success. Natural selection may occur via a combination of purifying selection, random genetic drift/neutral processes and positive selection.

Purifying selection is the process by which deleterious mutations, substitutions with deleterious effects, are removed from a population, preserving functional capabilities of a molecule and reducing the transmission of potentially harmful mutations (Cvijović, Good and Desai, 2018). In contrast, positive selection allows for molecular innovation when a substitution confers an advantage, allowing the advantageous allele to spread through a population and become fixed (Mitchell-Olds, Willis and Goldstein, 2007). Nonsynonymous substitutions refer to nucleotide changes which alter the resulting amino acid. Synonymous substitutions refer to nucleotide changes which, due to the degeneracy of the genetic code, the resulting amino acid is not changed (Zhang, Nielsen and Yang, 2005; Yang, 2007). Selective pressure can be assessed in protein coding regions by comparing the amount of nonsynonymous nucleotide substitutions per nonsynonymous site (dN) to synonymous nucleotide substitutions per synonymous site (dS) and quantifying the dN/dS ratio (ω); where a $\omega > 1$ provides evidence of positive selection, a $\omega = 1$ suggests neutral evolution and a $\omega < 1$ signifies purifying evolution.

Positive selection and protein functional shift have been considered synonymous (Lan, Wang and Zeng, 2013). The link between positive selection and protein functional shift has been formally tested by a small number of studies where rational mutagenesis and biochemical assays have been combined to determine impact of specific amino acid identities at positively selected sites. For example, positive selection at three amino acid sites in the mammal heme-peroxidase myeloperoxidase, has been shown to confer the unique chlorination activity (Loughran *et al.*, 2012). Positive selection is also implicated in the evolution of sweet taste perception in hummingbirds (Baldwin *et al.*, 2014), repurposing the ancestral T1R1-T1R3 umami receptor to detect nectar. Furthermore, positive selection acting on sites in DNA damage related genes senataxin (SETX) and ATM serine/threonine kinase (ATM) was found to contribute to telomere maintenance in longer-lived *Myotis* species (Foley *et al.*, 2018). The documented role of positive selective pressure in protein functional shift has led to the development of multiple methods to assess selective pressure in sequences of interest.

1.2.3 Assessment of selective pressure

1.2.3.1 Distance-based methods of assessing selective pressure variation

One of the first distance-based methods to detect selective pressure variation was proposed in 1985 (Li, Wu and Luo, 1985), using a nucleotide classification method based on the degeneracy of the genetic code. Nucleotide sites were divided into three categories; nondegenerate sites, with all possible changes to the site being nonsynonymous; twofold degenerate sites, where one of the three possible changes synonymous, with transitions considered synonymous and transversion considered nonsynonymous; and fourfold degenerate sites, where all possible changes are synonymous. This method, however, failed to account for the fact that transitions occur at a higher rate than transitions, resulting in an over-estimation of the rate of synonymous substitutions (Li, 1993). In an attempt to correct for this, the Kimura two-parameter (K2P) model (Kimura, 1980) was introduced. The K2PModel takes the weighted average of synonymous transitions at fourfold and twofold degenerate sites. This is then compared to the weighted average of nonsynonymous transversions at nondegenerate and twofold degenerate sites to obtain the ω value (Li, 1993). However, this method lacks statistical power to detect positive selection when it affects only a few sites within the sequence of interest. It relies on the weighted

average of synonymous transitions at fourfold and twofold degenerate sites and nonsynonymous transversions at nondegenerate and twofold degenerate sites over a sequence, rather than examining a sequence on a site by site basis. (Kosakovsky Pond and Frost, 2005).

1.2.3.2 Population-based methods of assessing selective pressure variation

Population based methods for detecting variation in selective pressure involve investigating variance in substitution rates within a population. One of the earliest methods is Tajima's D test statistic (Tajima, 1989) which sets out to test the hypothesis that a given multiple sequence alignment is evolving neutrally and may be used to detect selective sweeps. Using two values, (i) the observed nucleotide diversity of the sequences of interest ($\hat{\theta}_T$) within the population (Equation 1.3) and (ii) the expected heterozygosity ($\hat{\theta}_W$) of said population (Equation 1.4).

$$\hat{\theta}_T = \frac{\sum_{i<j} \quad d_{ij}}{n(n-1)/2}$$

Equation 1.3: *The average number of pairwise differences between two sequences (given by i and j). The number of differences between the two sequences is denoted by d_{ij}. The number of sequences is denoted by n.*

$$\hat{\theta}_W = \frac{S}{\sum_{i=1}^{n-1} \quad 1/i}$$

Equation 1.4: *The expected heterozygosity ($\hat{\theta}_T$) of the sample population, where S is the number of variable sites in the sample. The number of individuals in the population is n and the index of summation is i.*

The difference between the two values $\hat{\theta}_T$ and $\hat{\theta}_W$ is computed. In the event that the population of interest is evolving neutrally, the two values should be equal and the value returned is approximately zero. In the presence of a selective pressure acting on the population, a positive result indicates positive or purifying selection. A negative result indicates balancing selection. In cases where a positive value is returned, nucleotide diversity is greater than the expected population heterozygosity, possibly due to a decrease in population size or the maintenance of a number of different alleles of a gene within the population by purifying selection. A negative value is returned in cases where nucleotide diversity is lower than the expected population

heterozygosity, indicating an increase in population size or the removal of alleles by purifying selection.

Another population-based method to analyse selective pressure variation is the McDonald-Kreitman test (McDonald and Kreitman, 1991). This test, first applied on the *alcohol dehydrogenase* (*Adh*) gene in *Drosophila melanogaster*, uses an input multiple sequence alignment and a phylogeny of the species of interest to distinguish between variation within a species and substitutions that occur between species. The McDonald-Kreitman test uses the supplied phylogeny to detect polymorphisms. In this case, polymorphisms are defined as nonsynonymous substitutions at nonsynonymous sites that occur on the branch leading to the species of interest (a.k.a the 'within-species' branch). Nonsynonymous substitutions that occur on the branch leading to the species of interest are shared between all members on that branch. These are then compared to fixed substitutions - nonsynonymous substitutions at nonsynonymous sites on 'between-species' branches, i.e. branches that connect this group of species to the outgroup lineage (in this case the outgroup must be a closely related species to that silent substitutions are measurable). Under this framework, alterations due to selective pressure that occur on a within species branch (polymorphisms) occur only on the branch of interest, but not in other species, whereas alterations that occur on a between species branch are shared among all resulting species (fixed substitutions). Positive selective pressure is observed when the ratio of nonsynonymous to synonymous polymorphisms on the branch of interest is less than the ratio of nonsynonymous to synonymous fixed substitutions in the outgroup, as alterations in the nucleotide sequence due to positive selective pressure fix in the population more quickly than those due to genetic drift. Using the null hypothesis that the ratio of nonsynonymous to synonymous polymorphisms is equal to the ratio of nonsynonymous to synonymous fixed substitutions, a maximum-likelihood test of statistical significance called a G-test is conducted to determine if there is variance from neutrality. Positive selection has been reported to occur at varying levels in protein coding regions across *Metazoa* using the McDonald-Kreitman test. For example, studies of protein coding loci in *Drosophila* have reported signatures of positive selection in 41 - 45% of protein coding loci. A comparison of 43 orthologous gene families shared between *D. simulans* and *D. yakuba* reported 45% of examined loci were under

positive selection using the McDonald-Kreitman test (discussed below) (Smith and Eyre-Walker, 2002). A larger dataset of 115 protein coding gene families shared between *D. melanogaster, D. simulans* and *D. yakuba* reported 41% of examined loci were under positive selection across each examined species using the McDonald-Kreitman test (Welch, 2006).

The McDonald-Kreitman test, however, has been documented to induce Type II errors when slightly deleterious polymorphisms are included in the analysis, an issue which persisted even when attempts were made to remove slightly deleterious polymorphisms from the analysis by identifying low-frequency polymorphisms or estimating the distribution of fitness effects of novel polymorphisms (Messer and Petrov, 2013).

1.2.3.3 Likelihood-based methods of assessing selective pressure variation

Likelihood-based methods aim to assess the probability of observing the data under said model of evolution (Anisimova, Bielawski and Yang, 2001). Likelihood-based methods of assessing selective pressure variation aim to minimise the required number of parameters necessary to describe the underlying patterns of sequence change by comparing nested models using a likelihood-ratio test (LRT). LRTs may be applied with models of evolution that detect non-neutral evolution on a site-specific basis, also known as the sitewise likelihood-ratio (SLR) method (Massingham and Goldman, 2005), or models of evolution that evaluate positive selection on a whole-sequence basis, using information from all sites within the sequence (Nielsen and Yang, 1998).

These likelihood models may be integrated with codon-based models of evolution (Table 1.1) and *post-hoc* Bayesian analyses to determine first, the likelihood of positive selection within a sequence and second, the location of sites under positive selection in said sequence (Yang, 2007). These codon-based models of evolution were originally developed to use both the depth of coding sequence information available at the nucleotide level and the ability of amino acid sequences to remove background noise induced by the degeneracy of the genetic code in the analysis of distantly related species (Goldman and Yang, 1994). By modelling sequence evolution at the codon level, it is possible to analyse variation in selective pressure with sufficient sequence

information to quickly find substitutions. Codon sequences also can distinguish between substitutions which change the resulting amino acid sequence by altering the codon (non-synonymous substitutions at non-synonymous sites) and those that do not (synonymous substitutions at synonymous sites).

Table 1.1 Codon models of site-specific substitution implemented in the CodeML program within the PAML package.

CodeML Model	Free parameters	Fixed parameters
M0	ω	
M1neutral	$p_0, p_1 (p_1 = 1 - p_0)$	$\omega_0 < 1, \omega_1 = 1$
M2a	$p_0, p_1, p_2 (p_2 = 1 - p_0 - p_1)$	$\omega_0 < 1,\ \omega_1 = 1, \omega_2 > 1$
M3	$p_0, p_1, p_2 (p_2 = 1 - p_0 - p_1)$	$\omega_0 < 1,\ \omega_1 = 1, \omega_2 > 1$
M7(beta)	None	p, q
M8(beta & *ω)*	$p_0, p_1 (p_1 = 1 - p_0), p, q$	$\omega_s > 1$

Models of rates of evolution on a site-specific basis used in CodeML. ω refers to a rate of evolution over the entire multiple sequence alignment. ω_0 refers to purifying selection, ω_1 refers to neutral evolution and ω_2 refers to positive selection.. The proportion of sites evolving under ω_0 is denoted by p_0, the proportion of sites evolving under ω_1 is denoted by p_1 and the proportion of sites under ω_2 are denoted by p_2.

The simplest codon model of evolution in the CodeML suite of models from PAML (Yang, 2007) is M0, which assumes all sites within the alignment are evolving at the same rate (ω), which is calculated over the whole alignment. The M1neutral model attempts to model neutrality and estimates the proportion of sites that evolve under one of two fixed ω values; $\omega_0 = 0$ (purifying selection); and $\omega_1 = 1$ (neutrality). M2a models purifying and neutral selection (ω_0 and ω_1) and introduces and third ω value which is estimated from the data and can be >1. M3 an extension of M0, where ω is allowed to vary freely with a specified number of site classes: M3(k=2) discrete allows 2 site classes (ω_0 and ω_1, where $\omega_0 < 1$ and $\omega_1 = 1$), and M3(k=3) allows 3 site classes (ω_0, ω_1 and ω_2). M3 is used to model a variable ω over a sequence which is then compared to M0 to find the best fitting model of sequence variability. M7 is a beta distribution model of neutral evolution, where ω is allowed to vary between *p* and *q,* which are equal to 0 and 1 in this case. M7 may be compared against M8beta&ω. Under the M8beta&ω model, ω_0 varies on a beta distribution between *p* and *q,* where *p*=0 and *q*=1. A second ω value, ω_s, allows a proportion of

the sites in the sequence to be evolving under positive selection, where $\omega_s > 1$ (Goldman and Yang, 1994; Nielsen and Yang, 1998; Yang *et al.*, 2000; Wong and Nielsen, 2004; Yang, 2007). In order to test the fit of each model to the data, a likelihood ratio test (LRT) is performed. The LRT evaluates the statistical significance of adding in additional parameter by comparing a parameter rich model to a less parameter rich model (i.e. M1neutral is compared to M0). The LRT computes twice the difference of the log likelihood of both models (2Δl), followed by a chi-squared (χ^2). If the 2Δl value is greater than the χ^2 critical value, the more parameter rich model is determined to be significant. An example LRT for site-specific codon models of evolution is shown in Figure 1.2

	LRT	Null Model lnL	Alternative Model lnL	D	Critical Value	Null Rejected?
PTHR10059	m0 vs m3Discrtk2	-2822.9716	-2751.9354	142.072544	5.99	Yes
	m1Neutral vs m2Selection	-2752.786	-2743.1498	19.272464	5.99	Yes
	m3Discrtk2 vs m3Discrtk3	-2751.9354	-2741.653	10.282368	1	Yes
	m7 vs m8	-2753.2937	-2741.9358	22.715862	5.99	Yes
	m8a vs m8	-2751.0676	-2741.9358	18.263632	2.71	Yes

Figure 1.2 LRT of CodeML site specific models. *A gene family, PTHR10059, was analysed using site specific models in CodeML. M0 is compared to the M3Discrtk2 model to assess the probability of multiple ω values over a sequence. M1Neutral is a model of neutral evolution and is compared to a model that allows positive selection (m2Selection). M3Discrtk2 is compared to M3Discrtk3 in order to determine the likelihood of a sequence having 2 or 3 different ω values. The m7 beta distribution model of purifying selection and neutral evolution is compared to the m8 beta distribution model that allows positive selection. The m8 beta distribution model is then compared to a null m8 beta distribution model. The log likelihood of the null model and alternative model are denoted under "Null Model lnL" and "Alternative Model lnL". "D" refers to the difference between the two log likelihoods and the χ^2 critical value is given under "Critical Value". If the null model is rejected, it is stated under "Null Rejected"* (Yang, 2007).

If the model is determined to be statistically significant using LRTs, the posterior probability of a specific site in the sequence of interest being under positive selection is calculated using Empirical Bayes methods (Yang, Wong and Nielsen, 2005). Naïve Empirical Bayes, a previously applied method based on the estimated of maximum-likelihood parameters is subject to Type I errors in small datasets (Anisimova, Bielawski and Yang, 2002). The Bayes empirical Bayes (BEB) method is an improvement on the Naïve Empirical Bayes method, where a prior probability of observing positive selection is assigned to the model parameters and the prior probability is

averaged over, reducing the effect of small dataset size on the outcome of an Empirical Bayes calculation (Yang, Wong and Nielsen, 2005).

1.2.3.4 Lineage-based methods of assessing selective pressure variation

Early studies of molecular evolution led to the observation that different lineages can have different underlying rates of evolution (Langley and Fitch, 1974; Wu and Li, 1985; Britten, 1986). For example, by comparing amino acid sequences of 12 mammalian genes across 5 phylogenetic groups (*Primates, Rodentia, Lagomorphia, Artiodactyla, Carnivora)*, an increased substitution rate was observed on the branch leading to *Rodentia,* when compared to the other 4 phylogenetic groups (Li *et al.*, 1990). Further studies of variable evolutionary rates using primate insulin nucleotide sequences provided evidence of a slower rate of evolution on the branch leading to *Catarrhini* (Old World Monkeys, including *Homo sapiens)* when compared to *Platyrrhini* (New World Monkeys, such as the *Cebus* genus of capuchin monkeys) (Seino, Bell and Li, 1992). Using models that allow ω to vary among lineages, it is possible to investigate selective pressure variation within a species or, in the case of this thesis, at the root of *Eutheria*.

The program HyPhy (Pond, Frost and Muse, 2004) was designed to implement multiple analyses of molecular evolution, such as likelihood-based model fitting, molecular clock analyses, phylogenetic reconstruction using maximum likelihood methods and selective pressure analyses on sites and lineages. HyPhy was designed for multi-gene datasets and may be implemented as either a graphical user interface or as a command line application. Originally, the HyPhy branch-site package (BS-REL) assumed three site classes (ω_0, ω_1 and ω_2) and allowed each site to evolve under any of the three site classes (Kosakovsky Pond *et al.*, 2011). Later implementations of branch-site models in HyPhy estimates the optimal number of site classes for each branch in the dataset using the Akaike Information Criterion (AIC), shown in Equation 1.5 (below) (Smith *et al.*, 2015; Kosakovsky Pond *et al.*, 2019).

$$AIC = 2k - 2ln\,(\hat{L})$$

Equation 1.5: *Akaike Information Criterion. This estimates the quality of statistical models for a set of data, where k is the number of parameters, $\hat{L}$ is the maximum likelihood of the model and lnL is the natural log of the likelihood.*

The motivation behind this branch-site model was to enable analysis of viral genomes, which are fast-evolving and may have variable rates of evolution between different lineages. By estimating

the optimal number of site classes for each branch, the HyPhy branch-site model aims to allow rapid assessment of selective pressure across branches with extreme variation in evolutionary rates (Smith *et al.*, 2015). In addition, this branch-site model allows variation on 'background' lineages (i.e. outgroups), along with the 'foreground' lineages (i.e. species or lineage of interest).

While the branch-site model offered in HyPhy is adaptable to datasets with extreme variation in evolutionary rates, it was not chosen for this analysis. CodeML was found to be more suitable in this case as it has been extensively validated (Gharib and Robinson-Rechavi, 2013) and is regularly updated (Yang and dos Reis, 2010); has been found to be robust in cases of variable GC content over a dataset (Gharib and Robinson-Rechavi, 2013), a factor which has complicated analyses of the eutherian phylogeny (Romiguier *et al.*, 2010). In addition, CodeML has been successfully scaled up by the O'Connell group in the VESPA analysis pipeline to analyse selective pressure variation on a multi-genome scale (Webb, Walsh and O'Connell, 2016a).

Early lineage-specific models of variable selective pressure (Goldman and Yang, 1994) allowed ω to vary either among lineages or among sites, calculating an average *dN* over all sampled sites. These models only detected positive selection if the average *dN* was greater than the average *dS.* It was found that these models lacked sufficient power when only a few sites were under positive selection (Yang and Nielsen, 2002). Later, models were developed that allowed ω to vary on a lineage-specific and site-specific basis (Yang and Nielsen, 2002).

Table 1.2 Codon models of lineage-specific substitution implemented in the CodeML program within the PAML package.

Model	Proportion of Sites	Background ω	Foreground ω
ModelA	$p_2 = \frac{(1-p_0-p_1)p_0}{p_0+p_1}, p_0, p_1$	$0 < \omega_0 < 1, \omega_1 = 1$	$0 < \omega_0 < 1, \omega_1 = 1, \omega_2 \geq 1$

Models of rates of evolution on a lineage specific basis used in CodeML. ModelA is the branch-site model implemented and attempts to model positive selection, where background branches may be under purifying selection, where ω_0 is allowed to vary between 0 and 1, or under neutral evolution with ω_1 equal to 1. ModelA2 allows positive selection to occur in foreground branches ($\omega_2 \geq 1$). The proportion of sites under ω_0 and ω_1 are estimated from the data with the proportion of sites under p_2 calculated from the equation given under "Proportion of Sites".

The original implementation of branch-site models included two models, modelA and modelB. Both models used a phylogeny separated into foreground and background lineages (Figure 1.3), where purifying selection ($\omega_0 = 0$), neutral evolution ($\omega_1 = 1$) and positive selection were accommodated on foreground lineages ($\omega_2 \geq 1$). Under modelA, background branches were under purifying selection ($\omega_0 = 0$) and neutral evolution ($\omega_1 = 1$), with the parameters of ω_0 and ω_1 set prior to the analysis. Under modelB, background branches were under purifying selection (ω_0) or evolving neutrally (ω_1). However, the parameters of ω_0 and ω_1 were estimated from the data (Yang and Nielsen, 2002). Updates were made to these models after simulation studies discovered Type I (false positive) error rates of 20%-70% occurred when the rate of ω_0 was relaxed in foreground branches (Zhang, 2004).

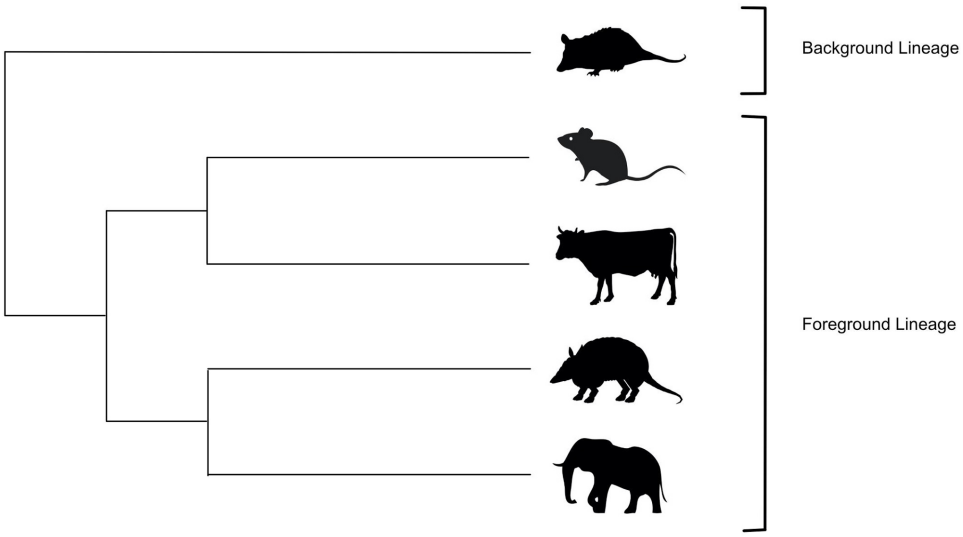


Figure 1.3 Sample phylogeny showing foreground and background lineages. *In this sample mammalian phylogeny, the outgroup which is Monodelphis domestica (the opossum) is the background lineage. The lineage of interest, which contains Mus musculus (mouse), Bos taurus (Cow), Dasypus novemcinctus (Armadillo) and Loxodonta Africana (Elephant), is the foreground lineage* (Yang, 2007).

An updated version of modelA is the currently used branch-site model. Under the updated modelA, background branches may be evolving under purifying selection, where ω_0 varies between 0 and 1; or be evolving neutrally, where $\omega_1 = 1$. The proportion of sites under ω_0 and ω_1 (p_0 and p_1) are estimated from the data. ModelA can accommodate positive selection in the foreground branches, where $\omega_2 \geq 1$, along with purifying selection (ω_0) and neutral evolution

(ω_1). In order to test for significance, likelihood ratio tests (LRTs) are performed, similarly to the site-specific models (detailed above). The updated modelA is compared first to M1neutral. If modelA is determined to be significant against M1neutral, it is then compared to modelAnull. Under modelAnull, $\omega_2 = 1$ in foreground lineages. By comparing model likelihood against two separate models of neutral evolution, the LRT is considered to be a stricter test of model likelihood . These models are tested for statistical significance in the same manner as the site-specific models, with the degrees of freedom being M1neutral (df = 1), modelA (df = 10) and modelAnull (df = 9) (Zhang, Nielsen and Yang, 2005) a sample LRT is detailed below in Figure 1.4.

	LRT	Null Model lnL	Alternative Model lnL	D	Critical Value	Null Rejected?
PTHR10088	m1Neutral vs modelA	-8869.7143	-8851.9075	35.613564	5.99	Yes
	modelAnull vs modelA	-8870.3955	-8851.9075	36.976142	3.84	Yes

Figure 1.4 LRT of CodeML lineage specific models. *A gene family, PTHR10088, was analysed using lineage specific models in CodeML. M1Neutral is a model of neutral evolution and is compared to a model that allows positive selection in the foreground lineage (modelA). modelA is then compared to the null modelA which sets ω = 1 in foreground lineages . The log likelihood of the null model and alternative model are denoted under "Null Model lnL" and "Alternative Model lnL". "D" refers to the difference between the two log likelihoods and the χ^2 critical value is given under "Critical Value". If the null model is rejected, it is stated under "Null Rejected". In this instance, PTHR10088 rejected both the m1Neutral model and the modelAnull. modelA, which allows positive selection in foreground lineages was determined to be the best fitting model for the data* (Yang, 2007).

These statistical tests of model likelihood have been found to positively influence the accuracy of a selective pressure variation analysis, with the false positive rate of LRT's implemented in CodeML estimated to be <5% (Zhai *et al.*, 2012) and the branch site model found to perform well under conditions of extreme sequence divergence and high GC content (Gharib and Robinson-Rechavi, 2013).

In contrast, only 3% of protein coding genes in 9 analysed whole genome simian primate assemblies had signatures of positive selection when analysed in a likelihood framework using the codon based models of evolution (CodeML) in the PAML software package (van der Lee *et al.*, 2017). The CodeML method was also used to assess 18 transcriptome and genome assemblies of bat species within *Chiroptera,* detecting signatures of positive selection in 16% of protein

coding genes (Hawkins *et al.*, 2019). The CodeML branch site models were used to investigate selective pressure variation in 4 marine mammal species: the walrus (*Odobenus rosmarus*), the bottlenose dolphin (*Tursiops truncatus*), the killer whale (*Orcinus orca*) and the West Indian manatee (*Trichechus manatus latirostric*). Of the protein coding loci shared between the 4 species, 1.13% were found to display signatures of positive selection (Foote *et al.*, 2015). A lack of resolution in sequencing data used may be a contributing factor in the variation in detected levels of positive selection, where variation may be clarified as sequencing quality improves. However, it has previously been established that different lineages may have different underlying rates of evolution (Langley and Fitch, 1974; Wu and Li, 1985; Britten, 1986), which may explain this variation.

1.3 miRNAs and Molecular Evolution

1.3.1 Brief synopsis of relevant biology of miRNAs

1.3.1.1 miRNA biogenesis and processing

MicroRNAs are short RNA sequences (~22nt in length) which regulate mRNA expression post-transcriptionally (Bartel, 2004). This post-transcriptional regulation acts via mRNA cleavage, or by translational repression (Bartel, 2009). MicroRNAs were first discovered in 1993 in *C. elegans* (Lee, Feinbaum and Ambros, 1993), with *lin-4*, a short RNA sequence, found to have imperfect complementarity to the 3' untranslated region (UTR) of the LIN-14 protein. *Lin-4* was found in three other *Caenorhabditis* species and is involved in post-embryonic development by negatively regulating LIN-14 (Lee, Feinbaum and Ambros, 1993). miRNAs have since been described in animals and plants and they have been shown to be dynamic regulators of many different pathways and processes (Zhang, Wang and Pan, 2006). miRNAs have been implicated in placental development, with miR-376c and miR-378a found to promote trophoblast proliferation by downregulating *Nodal* (Luo *et al.*, 2012; Guodong *et al.*, 2013). MicroRNAs have also been implicated in development, for example, miR-196, which acts as an inhibitor in vertebrate forelimb development by helping to restrict homeobox b8 (*HOXB8)* expression to the forelimbs, by downregulating *Hoxb8* in hindlimbs (Hornstein *et al.*, 2005). Singular miRNAs have been implicated in cell differentiation (Xiao *et al.*, 2007), trophoblast differentiation (Luo *et al.*, 2012) and forelimb development (Hornstein *et al.*, 2005). However, miRNAs have been found to work in 'clusters' or 'modules', forming regulatory groups to fine-tune the expression of target genes involved in common processes (Oliveira *et al.*, 2019). Evidence for this modular action of miRNAs includes a group of 28 miRNAs which work co-operatively to downregulate cycline-dependent kinase inhibitor 1A (*CDKN1A)*, an effector molecule regulating tumour suppressors (Wu *et al.*, 2010), with many of these miRNAs upregulated in tumour cells. miRNA modules are also implicated in colon cancer progression, with a group of 7 miRNAs significantly upregulated in tumour stromal cells and their expression positively correlated with serum carcinoembryonic antigen (CEA) levels, a prognostic biomarker for colorectal cancer (Zhou *et al.*, 2018). Studies of miRNA 'modules' have highlighted the ability of miRNAs to work in regulatory groups to fine tune target gene expression and underline the dynamic nature of these molecules.

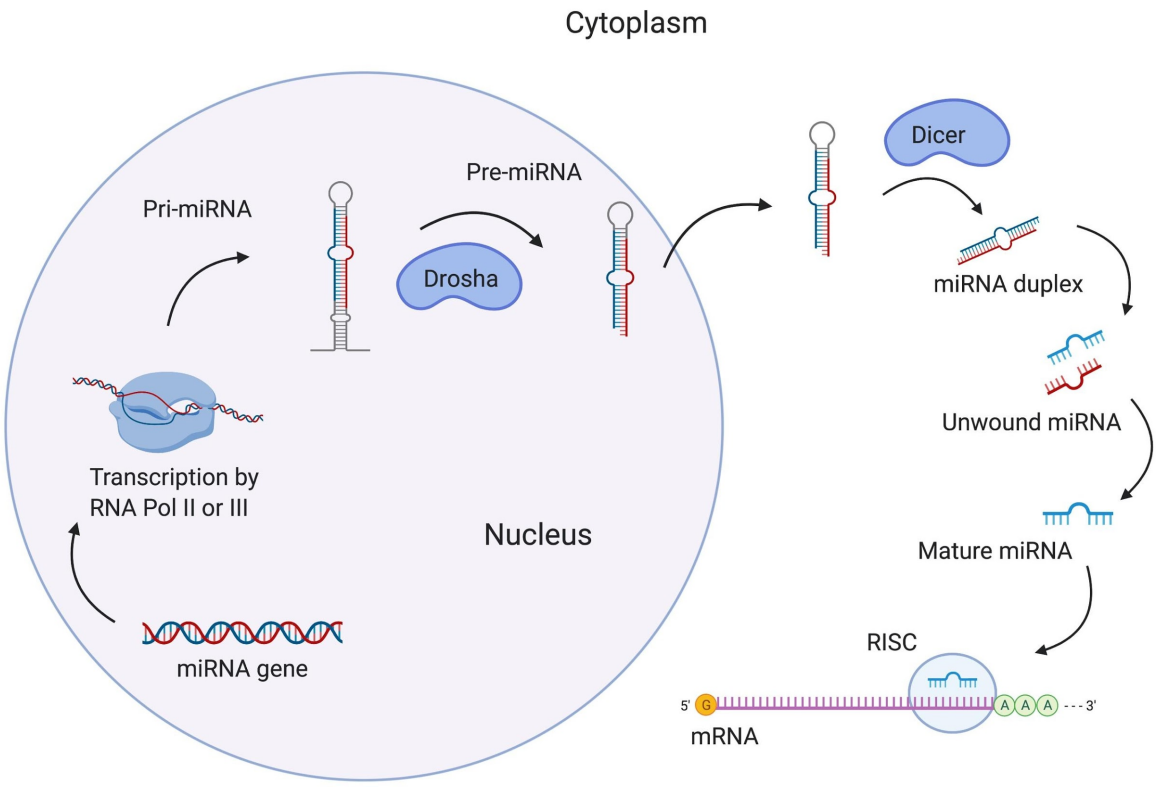


Figure 1.5 miRNA biogenesis and action. *miRNAs originate in the nucleus. The miRNA gene is transcribed by RNA Polymerase I or III. The resulting primary miRNA (pri-miRNA) is cleaved by Drosha in the nucleus to become a pre-miRNA. The pre-miRNA is transported from the nucleus into the cytoplasm. The pre-miRNA is processed by Dicer, a component of the RISC loading complex (RLC) to become a mature miRNA duplex. This duplex is unwound, with one strand loaded into the RISC as the guide strand and the other strand (the passenger strand) is degraded. The mature miRNA:RISC complex is then transported to the target gene, where the miRNA binds to the target site sequence in the 3'UTR of the target gene and regulates expression post-transcriptionally* (O'Brien *et al.*, 2018).

MicroRNAs originate in the nucleus (Figure 1.5) as primary microRNA (pri-miRNA) sequences of up to varying length, up to ~1000nt (Lee *et al.*, 2002). Multiple methods have been theorised to regulate miRNA expression. It is thought that RNA editing mediated by adenosine deaminase acting on RNA (ADAR) proteins may affect miRNA biogenesis and efficacy of action (Nishikura, 2010). It is hypothesized that knockout of ADAR proteins results in upregulation of miRNAs. For example, miR-142 expression was found to be significantly increased in ADAR1/2-null mice (Yang *et al.*, 2006), with similar results observed using RNA-sequencing in ADAR/ADARB1-null mice (Vesely *et al.*, 2012). Changes in miRNA expression have also been correlated with changes in the expression of a host gene in which the miRNA is encoded (Ye *et al.*, 2014). For example, expression of miR-155-5p and miR-155-3p, located within the B-cell Integration Cluster (BIC) gene

was found to be induced alongside BIC expression during B-cell activation (Elton *et al.*, 2013). This mechanism of miRNA stimulation is also implicated in a negative feedback loop – miR-127, which is encoded in the antisense retrotransposon-like 1 *(RTL1)* transcript, works to downregulate the sense *RTL1* transcript during placental formation and to prevent placentomegaly (Ito *et al.*, 2015a). Transcription factors have also been observed to regulate miRNA expression levels. For instance, 11 miRNAs were found to be regulated by hepatocyte nuclear factor 1 alpha (HNF1α) and beta (HNF1β), with these miRNAs found to be significantly downregulated in monogenic diabetes patients, who present with a mutation in the HNF1α or HNF1β gene (Fendler *et al.*, 2016). Transcription factor mediated miRNA regulation has also been observed during cell cycle progression. A member of the E2F transcription factor family of cell cycle regulators, E2F transcription factor 7 (E2F7), was found to antagonise 5 miRNAs (miR-25, miR-26a, miR-27b, miR-92a and miR-7) involved in cell proliferation, thereby restraining progression of the cell cycle (Mitxelena *et al.*, 2016).

Physiological factors can also alter miRNA expression. Oestrogen expression has been found to alter the expression of multiple miRNAs, including miR-29 (Zhang *et al.*, 2012); miR-26a and miR-181a (Maillot *et al.*, 2009); and miR-879 (Wetzel-Strong *et al.*, 2016). Corticosterone has been identified as a mediator of miRNA expression, inducing differential expression of 26 miRNAs in a model of corticosterone-mediated depression (Dwivedi *et al.*, 2015). Lastly, miR-21, a miRNA highly expressed in human tumours and cancer cell lines, is induced by the binding of Type I Interferon induced STAT3 to a promotor region upstream of miR-21 (Yang *et al.*, 2015). It is clear that miRNAs may be stimulated by multiple processes: via the action of transcription factors, host gene expression or physiological stimulation. Understanding the mechanisms regulating miRNA gene expression and how miRNAs go on to regulate target genes are important factors in the study of miRNA-mediated phenotypic alterations.

Once miRNA expression is induced, pri-miRNA sequences are transcribed by RNA polymerase II or III (Starega-Roslan *et al.*, 2011). The pri-miRNA are cleaved by Drosha, a microprocessing protein composed of an RNase III domain and a double stranded RNA binding domain (Han *et al.*, 2004). Post-cleavage, the microRNAs are known as pre-miRNAs, ~70nt hairpin sequences which

are transported into the cytoplasm by the Ran GTP-dependant factor exportin-5 (Shivdasani, 2006). In the cytoplasm, the pre-miRNAs are processed by the RISC (RNA-induced Silencing Complex) loading complex (RLC), which is composed of Dicer, TRBP (Tar RNA binding protein), PACT (protein activator of protein kinase RNA) and Argonaute-2 (Ago2) (Winter *et al.*, 2009). Dicer is responsible for the cleavage of pre-miRNA, with TRBP stabilising Dicer and PACT being associated with higher numbers of newly processed mature microRNAs, perhaps contributing to the efficiency of the RLC (O'Brien *et al.*, 2018). Mature microRNAs are produced from the cleavage of one arm of the stem-loop pre-miRNA, with one stem-loop capable of producing two mature microRNAs. Cleavage of the arm at the 5' end results in the 5p mature miRNA strand (e.g. miR-28-5p), which cleavage of the arm at the 3' end results in the 3p mature miRNA strand (e.g. miR-28-3p) (Figure 1.5). They are short RNA sequences of ~22nt with a 5' phosphate and a 2nt overhang on the 3' end (Wahid *et al.*, 2010).

1.3.1.2 miRNA regulation of gene expression

Post-miRNA maturation, the RLC detaches from the miRNA. RISC can vary in molecular composition, from a single Argonaute (AGO)-protein complex to more complicated molecules with multiple bound proteins (Pratt and MacRae, 2009). The mature miRNA is a duplex composed of a guide strand, complementary to the target UTR region, and a passenger strand. The strand with the lowest thermodynamic stability at the 5' end, or a uracil at position 1 on the 5' end, is selected as the guide strand. The second strand in the duplex becomes the passenger strand (O'Brien *et al.*, 2018). The miRNA duplex is unwound by a non-specific helicase, with the guide strand loaded into the RISC and the passenger strand degraded (Winter *et al.*, 2009). The RISC locates the target gene via imperfect complementarity between the miRNA and the target 3' UTR where gene expression is regulated by (i) mRNA cleavage, or (ii) translational repression (Bartel, 2004). The method of expression regulation is determined by the level of complementarity between the miRNA and target.

miRNA regulation can result in gene silencing through the action of the minimal miRNA-induced silencing complex (miRISC), comprising the guide strand and an AGO protein (Kawamata and Tomari, 2010). The miRISC complex interacts with complementary target sites on the target

mRNA, known as miRNA response elements (MREs) (Bassett *et al.*, 2014). The mode of miRISC-mediated gene silencing is dependent on the level of complementarity between the miRNA and target site sequence. High complementarity results in mRNA cleavage, while a lower level of complementarity results in translational repression (Zeng, Yi and Cullen, 2003; Correia de Sousa *et al.*, 2019).

MicroRNA cleavage of mRNA is mediated by Ago2 proteins in the RISC and occurs between the 10th and 11th nucleotide, resulting in mRNA decay (Winter *et al.*, 2009). MicroRNAs can also induce mRNA decay by deadenylation. This process involves recruitment of a member of the glycine-tryptophan protein of 182 kDA (GW182) family. This GW182 protein facilitates adenylation of the target mRNA by interacting with polyadenylate-binding protein (PABPC), which in turn recruits the poly(A)-nuclease deadenylation complex subunit 2 (PAN2)- poly(A)-nuclease deadenylation complex subunit 3 (PAN3) complex and the carbon catabolite repressor protein 4 (CCR4)-NOT complexes (Gebert and MacRae, 2019).

However, miRNAs have also been found to positively regulate gene expression, although all mechanisms of this have not yet been fully elucidated (O'Brien *et al.*, 2018). It is hypothesized that upregulation of gene expression may be carried out by miRNAs either by direct binding to the seed sequence, similarly to the method of downregulation, or by engaging in competitive binding with a miRNA that is repressing the gene of interest (Vasudevan, 2012). Previous work investigating miRNA-mediated upregulation discovered that miR-145 positively regulates vascular smooth muscle development from progenitor cardiac cells, possibly by upregulating Kruppel-like factor 4 and Calmodulin kinase II-delta, both positive regulators of smooth muscle development (Cordes *et al.*, 2009). Further studies of the Kruppel-like factor 4 discovered a translational control element in the 3'UTR of Kruppel-like factor 4 which acts as a binding site for miR-206 to positively or negatively regulate gene expression during epithelial cell proliferation (Lin *et al.*, 2011). Promotor RNA elements have also been found to assist in the upregulation of cyclooxygenase-2 (COX2), with miR-589 found to target a promotor RNA element located in the COX2 3'UTR (Matsui *et al.*, 2013). Further support for this miRNA-promoter induced activation of transcription was found in regulatory T-cells, where miR-4281 was found to activate the

expression of human forkhead box protein 3 (FOXP3) transcription factor by binding its TATA-box motif and promoting regulatory T cell differentiation (Zhang *et al.*, 2018). This mechanism of activation is termed miRNA-induced RNA activation, or miRNAa, providing one candidate pathway to study miRNA-mediated upregulation of gene expression.

1.3.2 The role of miRNAs in morphological innovation

miRNAs have been found throughout *Metazoa*, plants and algae (Moran *et al.*, 2017) and are thought to have originated separately on the plant and animal lineages. In plants, miRNAs are produced wholly in the nucleus and exported to the cytoplasm, while animal miRNAs are first produced as pri-miRNAs in the nucleus and processed to become pre-miRNAs, then exported to the cytoplasm to form the mature miRNA (Bellato *et al.*, 2019). These differences, have led to the theory that miRNAs have evolved convergently in plant and animal ancestral lineages (Tarver, Donoghue and Peterson, 2012), with each lineage altering and expanding the miRNA repertoire.

Mature miRNA sequences are highly conserved, due to evolutionary limitations. For the miRNA to function, the mature miRNA sequence must remain the same, in order to be recognized by the target sequence(s). During the stem-loop formation step in the processing of miRNA, a change in one arm of the loop must be balanced by a change in the other arm, increasing the steps required for a mutated miRNA to remain functional (Tanzer and Stadler, 2006). MicroRNAs, once established, are rarely lost, with the *C. elegans* miRNA *let*-7 conserved across worms, flies and humans (Pasquinelli *et al.*, 2000) and only 12 documented miRNA losses in deuterostome evolution (Campo-Paysaa *et al.*, 2011). To this end, miRNAs are hypothesized to be continuously expanding (Berezikov, 2011).

Significant miRNA family expansions have been found at the base of many divergences in the animal tree of life. The miRNA repertoire was found to have expanded at the base of *Bilateria* (Hertel *et al.*, 2006; Prochnik, Rokhsar and Aboobaker, 2007), indicating a requirement for novel regulatory patterns associated with innovation in morphological complexity. Other examples include early radiation of *Lepidoptera* (Quah, Hui and Holland, 2015), 41 miRNA families are theorized to have contributed to the development of vertebrate complexity (Heimberg *et al.*,

2008) due to their expression in vertebrate-specific tissues, such as the thymus (Heimberg *et al.*, 2008). In *Eutheria* miRNA emergence was found to occur at a rate three times higher in marsupial and eutherian mammals compared to miRNA emergence in monotremes and birds (Soumillon *et al.*, 2013). This comparatively rapid expansion at a time of morphological innovation may indicate the importance of these new miRNAs in eutherian diversification (Berezikov, 2011).

Due to their phylogenetically distinct nature and observed functions in the regulation of tissue and organ development, miRNAs have been studied in the context of the evolution of eutherian pregnancy. Studies of miRNA sequences across *Vertebrata* have identified a number of miRNAs that are hypothesized to be specific to *Eutheria*. The miR-1302 miRNA gene family is derived from the MER53 transposable element, which is only found in *Eutheria*. The miR-1302 gene family is only found in the mammalian placenta, indicating that the miR-1302 gene family arose after the divergence of eutherian mammals and may have assisted in the development of the placenta (Yuan *et al.*, 2010). Another placental mammal specific group of microRNAs are the mir-379/mir-656 cluster, located in the DLK-DIO3 region of human chromosome 14 and also known as the C14MC miRNA cluster (Laddha *et al.*, 2013). miRNAs in this cluster have been found to be differentially expressed in blood plasma between Day 0 and Day 60 of bovine pregnancy (Ioannidis and Donadeu, 2017). Another member of the C14MC cluster, miR-411, is expressed in the human placenta and was found to be significantly downregulated in preeclamptic placentas (Yang *et al.*, 2019).

The phylogenetically distinct nature of miRNAs, along with their documented expansion in repertoire at the base of *Eutheria* (Soumillon *et al.*, 2013) indicate that they may be informative in the study of novel traits in *Eutheria*. It is our hypothesis that miRNAs that arose on the stem eutherian lineage induced alterations in gene regulation and contributed to the evolution of mammal placenta.

1.3.3 miRNA Target Prediction

As mentioned above miRNA expression may be induced by multiple mechanisms, including transcription factor regulation, physiological stimulation or RNA editing (Gulyaeva and Kushlinskiy, 2016). miRNAs work to regulate target gene expression by binding to a target site sequence, inducing either cleavage, degradation or upregulation of the target mRNA (O'Brien *et al.*, 2018). In order to investigate the effects of miRNA-mediated gene regulation, a number of computational approaches have been developed to identify miRNA targets on a large scale.

An important factor in determining the biological reality of a predicted target is examining the size and complementarity of the seed region. The four major seed types match differing positions in the seed region (Table 1.1). 7mer-m8 and 8mer seed types match the largest area in the 3'UTR binding region, with such matches indicative of a 'true' positive (Agarwal *et al.*, 2015).

Table 1.3 Features of the canonical microRNA seed types.

Type	Length	Region Match
6mer	6nt	Position 2-7 of the seed
7mer-A1	7nt	Position 2-7 of the seed followed by an 'A'
7mer-m8	7nt	Position 2-8 of the miRNA (seed region + position 8)
8mer	8nt	Position 2-8 of the miRNA (seed region + position 8, followed by an 'A')

miRNA target interactions depend on complementary binding of the mature miRNA and the 3'UTR of the target transcript. It is theorized that the greater number of complementary nucleotides in the mature miRNA/target sequence match are indicative of increased binding strength and also may serve to identify biologically 'real' targets when predicting miRNA targets using computational methods (Lewis, Burge and Bartel, 2005a; Agarwal *et al.*, 2015).

High-throughput computational prediction of microRNA targets is based on the complementary seed-pairing of a microRNA to the target mRNA. While many programs use this basic searching method (e.g. TargetScan (Agarwal *et al.*, 2015), miRanda (Enright *et al.*, 2004), miTarget (Kim *et al.*, 2006)), different predictors rely on varying searching and scoring methods to ensure the great possible accuracy of the target prediction.

Table 1.4 Features of widely used miRNA target prediction software.

Method	Searching Method	Conservation Score?	Estimate of miRNA Effect	Dataset	Prediction Ranking
TargetScan	Seed pairing	Yes	Yes	5'UTR, 3'UTR, exon, intron	8mer > 7mer-m8 > 7mer-a1 > 6mer
miRanda	Seed pairing	Yes	Yes	3'UTR	By calculated alignment score
RNA22	Seed pairing	No	No	3'UTR	8mer > 7mer-m8 > 7mer-a1 > 6mer, binding energy
Pita	Seed pairing	Yes	Yes	3'UTR	8mer > 7mer-m8 > 7mer-a1 > 6mer, accessibility of target site

Multiple methods of miRNA target prediction exist. All 4 methods presented predict miRNA targets based on complementary binding of the miRNA to the seed sequence in the target gene. Conservation score, or the degree to which the miRNA seed sequence is conserved across species/lineages is computed in TargetScan, miRanda and PITA, using the hypothesis that a putative seed sequence observed in multiple species is more likely to be biologically functional. Estimate of the miRNA effect are calculated in TargetScan and miRanda, using cumulative scores of the site type, site abundance, stability of the seed pairing, conservation and local AU content. Estimate of miRNA effect is calculated in PITA by estimating the binding energy of the miRNA:target interaction. miRanda, RNA22 and PITA predict miRNAs using a dataset of 3'UTR sequences, while TargetScan uses multiple genomic regions, as miRNA targeting has been observed throughout the genome. TargetScan, RNA22 and Pita provide information on the seed pairing between the miRNA and seed sequence, with 8mer seed pairing denoted as optimal. miRanda uses this information to compute an alignment score of the miRNA and target site sequence to determine the level of seed pairing. (Enright *et al.*, 2004; Miranda *et al.*, 2006; Kertesz *et al.*, 2007; Agarwal *et al.*, 2015).

When using any microRNA target prediction software, a trade-off must be made between specificity and sensitivity, or the number of false positives (the percentage of correct predictions from the total predicted targets) and false negatives (the percentage of predicted correct targets from the total number of correct targets), i.e. whether the software can detect all candidate targets and discard unlikely candidate targets. Sensitivity is a preferred parameter when examining all targets for pre-selected microRNAs, in order to maximise the number of targets,

while specificity is preferred when investigating all microRNAs targeting a pre-selected gene or genes (Witkos, Koscianska and Krzyzosiak, 2011).

There is currently no 'best' miRNA-target prediction software. Comparison of the highest quality prediction software found that, among TargetScan, miRanda-mirSVR, Pita and RNA22, specificity and sensitivity were variable. While miRanda-mirSVR showed the highest sensitivity (0.617; with TargetScan, Pita and RNA22 scoring 0.524, 0.336 and 0.336 respectively), it scored 0.954 in specificity, with RNA22 and TargetScan outperforming miRanda-mirSVR with respective scores of 0.992 and 0.984 (Oliveira *et al.*, 2017). These metrics agree with previous benchmark analyses (Walsh, 2013; Fan and Kurgan, 2015) with TargetScan among the highest quality programs available. The algorithm searches multiple genomic regions along with 3'UTRs for binding sites (Lewis, Burge and Bartel, 2005b) and was designed using benchmark datasets from human, mouse and Drosophila. In addition, TargetScan returns matches based only on consecutive seed region matches (i.e. no gaps in the matched sequences), providing the best compromise between specificity and sensitivity. This compromise was of particular importance in the analyses conducted in this thesis, with interactions between possible novel eutherian miRNAs and any genes that may display statistically significant signatures of positive selection on the eutherian lineage being of interest in the context of novel eutherian regulatory networks

1.4 Mammalian Reproduction

Mammal placenta can be defined as 'an approximation or combination of an embryo's tissues with those of its natural or surrogate parent for physiological interchange' (Mossman, 1987). The function of the placenta is to form an interface between the mother and foetus, specifically facilitating the exchange of nutrients, gases and waste necessary for foetal development (Allen C. Enders and Blankenship, 1999) along with transmission of maternal antibodies; oxygen; glucose, amino acids and water; and hormones (Griffiths and Campbell, 2015).

Placentation in mammals evolved c. 93 MYA with the divergence of *Eutheria* from the subclass *Theria* (Tarver *et al.*, 2016). Placenta exists in a variety of forms (Figure 1.2) in different mammal phylogenetic groups, but all forms carry out the same essential function (Mess, 2014).

Variation in placental morphology includes the interhaemal barrier (Figure 1.2A) between mother and foetus (i.e. invasiveness) which can be classified as follows:

(i) Haemochorial placentation – observed in humans and other Old World primates, with no barrier between the foetus and the maternal blood, which has the largest degree of invasiveness

(ii) Endotheliochorial placentation – observed in *Carnivora*, which has a single barrier of tissue between the foetus and the maternal blood

(iii) Epitheliochorial placentas – observed in *Ungulata* are the least invasive, with three tissue layers between the maternal blood supply and the embryo (Garratt *et al.*, 2013)

Placental variation also extends to the degree of interdigitation, or point of exchange (Figure 1.2B), between the foetus and mother, with the following possible arrangements of vessels observed in mammals.

(i) Villous interdigitation – observed in humans and other Old World primates; where groups of villi are attached only at the connection to the chorionic wall, remaining otherwise free floating.

(ii) Trabecular placenta – observed in New World Primates, where trophoblast cells form branching structures known as trabeculae

(iii) Lamellar placenta – observed in *Carnivora*, where the fetal and maternal tissues are intricately folded together.

(iv) Labrynthine placenta – observed in *Rodentia*, where the villi are heavily interconnected (Garratt *et al.*, 2013; Carter and Mess, 2014).

Finally, the eutherian placenta differs among placental mammals in its shape (Figure 1.2C), or the distribution of the area of fetomaternal exchange.

(i) Diffuse placenta – observed in New World Monkeys, where the area of maternofetal exchange covers the entire luminal epithelium.

(ii) Cotyledonary placenta – observed in ruminants, where the fetal villi are grouped cotyledons separated by the chorion. Maternofetal exchange occurs where cotyledons are in contact with maternal caruncles

(iii) Zonary placenta – observed in *Carnivora*, where the area of maternofetal exchange is wrapped around the chorion in a 'belt'-like structure.

(iv) Discoid placenta – observed in *Euarchontoglires,* with the area of exchange confined to a single (discoid) or double (bidiscoid) circular area (Furukawa, Kuroda and Sugiyama, 2014).

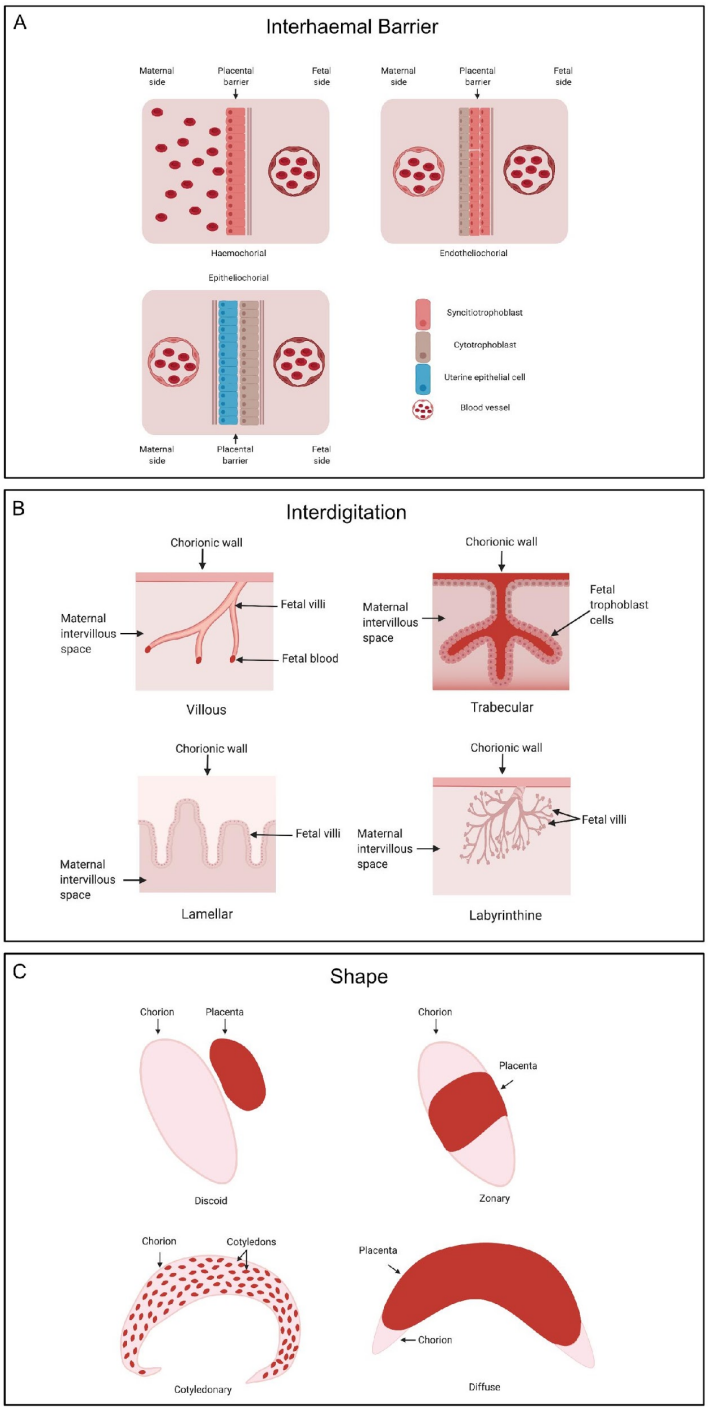


Figure 1.6 Variations in placental morphology across Eutheria. *The eutherian placenta serves as a point of physiological exchange between the mother and fetus throughout gestation. However, the placenta varies in morphology throughout Eutheria. Placental morphology may be defined by (A) the interhaemal barrier or level of separation between maternal and fetal tissues, ranging from the highly invasive haemochorial type, to the less invasive endotheliochorial type,to the epitheliochorial type, where three layers of tissue exist between the maternal and fetal blood; (B) the interdigitation or point of exchange between mother and fetus. The villous type of interdigitation consists of branching fetal villi attached at the chorionic wall and extend into the maternal intervillous space, with an increase in the degree of folding between maternal and fetal tissue observed in the trabecular interdigitation type, where fetal trophoblast cells increase the communication with maternal tissues. The lamellar interdigitation type increases the complexity of the folding between the fetal and maternal tissues, with the highly intricate labyrinthine type displaying the most complex interdigitation, where a web of fetal vessels interact with the maternal system; and (C) the shape of the placenta which can present as discoid, or a disc-like placenta attached to the chorion, zonary, where the placenta exists as a belt-like structure around the chorion, cotyledonary, where the physiological exchange is facilitated by placental cotyledonary structures over the whole chorion, or diffuse, where the placenta consists of villi densely scattered over the whole chorion.*

1.4.1 Mammalian reproductive cycles

Mammalian reproductive cycles vary across the eutherian phylogeny. Humans and Old World primates, such as gorillas and chimpanzees, undergo a menstrual cycle. In contrast to animals that undergo estrous, ovulation in menstruating mammals is not initiated by mating, but rather an increase in luteinizing hormone (LH) (Alvergne and Högqvist Tabor, 2018). Post-ovulation, the endometrium rapidly proliferates to establish uterine receptivity to implantation. If no implantation occurs, this thickened lining is lost via retrograde flow out of the vagina if no successful pregnancy occurs (Martin, 2007). Within New World monkeys, *Aotus nancymaae* (Nancy Ma's night monkey) and *Sapajus microcephalus* (the large headed capuchin), have shown histological evidence of a menstruation, whereby the thickened endometrial lining detaches and is lost (Mayor *et al.*, 2019). *Strepsirrhines*, which includes lemurs and galagos, do not menstruate (Emera, Romero and Wagner, 2012). Menstruation has also been observed in *Phyllostomidae,* the New World leaf-nosed bats (Rasweiler and Badwaik, 2000). In those mammals that don't menstruate, the cyclic changes that occur in the endometrium occur across what is termed the estrous cycle. This consists of (Radi, Marusak and Morris, 2009), proestrous, estrus (or sexual receptivity), metestrus (the end of sexual receptivity and the formation of the corpus luteum) and diestrus (increased progesterone secretion by the corpus luteum), followed either by a period of sexual dormancy known as anestrous (Lombardi, 1998). In polyestrous species, such as cattle, rats and mice, multiple estrous cycles occur throughout the year if conception does not occur (Crowe and Mullen, 2013; Ajayi and Akhigbe, 2020). In seasonally polyestrous species, such as cats and horses (Kutzler, 2015; Ball *et al.*, 2019), multiple estrous cycles occur over one breeding season; in contrast with monoestrous species such as the dog, which displays one estrous cycle during a breeding season, followed by a protracted period of anestrous (Concannon, 2011).

However, all eutherian mammals have the following cyclic phases in common: follicle maturation, ovulation and luteinisation (initiation of corpus luteum formation) (Bronson, 1989), with the hormonal regulation of the cycle subject to positive and negative feedback mechanisms (Crowe and Mullen, 2013; Messinis, Messini and Dafopoulos, 2014). While all follicles are laid down *in utero,* follicle maturation, governed by three major hormones: gonadotropin-releasing

hormone (GnRH), luteinizing hormone (LH) and follicle-stimulating hormone (FSH), occurs on a cyclic basis in animals that are spontaneous ovulators (such as humans, sheep and rats), in contrast to species in which follicle maturation is induced by mating (such as rabbits and cats) (Bakker and Baum, 2000). GnRH is synthesized in the hypothalmus at the beginning of the reproductive cycle and transported to the anterior pituitary gland, at which time GnRH initiates the secretion of the gonadotropins LH and FSH (Conn and Crowley, 1991). The release of FSH stimulates the recruitment of follicles from the ovarian primordial follicle pool, with one follicle developing more rapidly than others, becoming the dominant follicle. As the dominant follicle matures, serum estradiol concentration rises and FSH secretion is inhibited, suppressing further growth of subordinate follicles, which undergo programmed cell death via apoptosis (Silverthorn, 2010). This rise in estradiol concentration during the late follicular phase induces a surge in LH levels mediated by GnRH, which in turn induces oocyte maturation and the rupture of the follicle, releasing the egg (de Ziegler *et al.*, 2007). Ovulation may occur bilaterally (in both ovaries at once), such as in rodents, or unilaterally (in one ovary at a time), in canines, ruminants and primates (Sato, Nasu and Tsuchitani, 2016).

The LH surge also induces the migration of follicular thecal cells into the antral space, or the area of the follicle previously filled with follicular fluid, where they integrate into the population of granulosa cells which previously surrounded the oocyte and form the corpus luteum (CL), a gland which produces progesterone during early pregnancy. The granulosa cells then differentiate into large steroidogenic luteal cells, with the thecal cells differentiating into small steroidogenic luteal cells, enabling progesterone production and secretion (Stocco, Telleria and Gibori, 2007; Kumar and Magon, 2012). The pre-ovulatory LH surge also mediates the expression of factors involved in the post-ovulatory remodelling of the granulosa cell extracellular matrix via cleavage by extracellular proteases, allowing angiogenesis in the new corpus luteum (Niswender *et al.*, 1994). Post-ovulation, progesterone is produced by the corpus luteum in increasing concentrations. This induces transcriptional changes and remodels the endometrial landscape to facilitate successful pregnancy by suppressing the secretion of estradiol, promoting stromal cell proliferation and epithelial cell differentiation to thicken the endometrium in preparation for implantation (Radi, Marusak and Morris, 2009).

1.4.2 Progesterone production and mechanism of action

Progesterone is produced by the corpus luteum until the fully formed placenta takes over P4 production in some species. In others, P4 output is predominantly maintained by the CL (Young, 2013). Progesterone drives transcriptional changes by binding to either isoform of its nuclear receptor, progesterone receptor A (PGRA) and progesterone receptor B (PGRB) (Graham and Clarke, 1997). The expression of these receptor isoforms is positively regulated by estradiol and negatively regulated by progesterone, with PGRA and PGRB mRNA concentrations found to inversely correlate with serum progesterone concentrations during the early to mid secretory phase of the human corpus luteum (Misao *et al.*, 1998). This observation is also true in ruminants, where progesterone negatively regulates the expression of progesterone receptors during the mid luteal (or secretory) phase (Spencer and Bazer, 1995). In the absence of progesterone, PGRs are bound to heat shock proteins in the cytoplasm. Upon binding to progesterone, the PGRs undergo conformational changes, dimerizing and relocating to the nucleus where they initiate the transcription of PGR regulated genes by recruiting transcription complexes to the site of specific progesterone response elements in these PGR regulated genes (Lee *et al.*, 2006). While both progesterone receptor forms are equally abundant in humans, isoforms are expressed on a cell specific basis in humans. PGRA is highly expressed in the stroma and early decidua during the luteal phase, indicating that PGRA may have a role in epithelial-stromal communication during the decidualisation reaction in humans. PGRB is highly expressed in the glandular epithelium during the proliferative phase (Wang *et al.*, 1998). In the murine endometrium, the PGRA isoform is the primary transducer of progesterone action upon gene expression (Schneider *et al.*, 1991). Within the bovine endometrium, progesterone receptors are expressed in specific patterns during early pregnancy, with no PGR expression in the surface epithelium, low PGR expression in the uterine glands and high expression in the stromal cells (Boos *et al.*, 2006). During the bovine estrous cycle, *PGR* mRNA is expressed highly on day 5 of pregnancy, decreasing thereafter in response to increased progesterone secretion, with significant downregulation of *PGR* mRNA observed on day 13, during the late luteal phase coincident with bovine conceptus hatching and the beginning of elongation (Okumu *et al.*, 2010). The distinct expression pattern of PGR and its

isoforms indicates a cell or locale specific need for progesterone mediated transcriptional changes during early pregnancy. In the human uterine epithelium, PGRA and PGRB are involved in regulating the mucin 1, cell surface associated (*MUC1*) promoter, being upregulated by PGRB. The effects of PGRB are then antagonised by PGRA (Brayman *et al.*, 2006), downregulating the expression of *MUC1*, a transmembrane glycoprotein which has been found to inhibit adhesion during implantation (DeSouza *et al.*, 2000). PGRA was also found to directly antagonise MUC1 in murine endometrium, in contrast to its role in human where it downregulates MUC1 indirectly by antagonising PGRB action in human uterine epithelium (Brayman *et al.*, 2006). Within the human decidualization reaction, PGRA and PGRB were found to regulate transcription together and separately, with common pathways involved in angiogenesis and reaction to hypoxia. PGRA-specific pathways were found to be involved in the regulation of apoptosis and protein kinase activity, while PGRB-specific upregulated pathways were implicated in cell cycle regulation, chemokine secretion and cell adhesion, while PGRB-specific downregulated pathways were implicated in cell cycle regulation (Kaya *et al.*, 2015). The role of PGRB in downregulating genes related to the cell cycle may be due to the action of progesterone. During the window of receptivity, progesterone induces expression of the Hand2 transcription factor which downregulates fibroblast growth factors (Halasz and Szekeres-Bartho, 2013). Furthermore, there is support for a role of progesterone in supporting uterine receptivity during the human window of implantation (WOI), with 653 out of 3,052 genes found to be differentially expression during the WOI directly bound and regulated by an isoform of the progesterone receptor (Chi *et al.*, 2019). These genes regulated by the progesterone receptor during the human WOI were found to be implicated in functions related to the regulation of inflammatory response, estrogen response, cell death and interleukin/STAT signalling. Within the bovine endometrium, there is strong evidence for a role of progesterone in supporting conceptus elongation and development during the pre-implantation phase (Lonergan, 2011). These progesterone-induced alterations to the bovine maternal endometrium are most marked around Day 13 of pregnancy, with 4,421 genes found to be differentially expressed at Day 13 when compared to Day 7 of pregnancy. These genes were found to be implicated in pathways related to electron transport; cytoskeletal remodelling; and protein, nucleotide and energy source transport (Forde *et al.*, 2009). This final functional category may be related to a role of progesterone in histotroph secretion, the method

by which conceptus maturation is supported by secretions from the uterine lumen (Lonergan and Forde, 2014).

Progesterone itself is produced in the corpus luteum using cholesterol as a precursor for the steroidogenic pathway (Wiltbank *et al.*, 2012). *Rodentia* (such as mice and rats) and in *Ruminantia* (such as cattle) obtain cholesterol by the endocytosis of low-density lipoprotein (LDL). Primates (such as humans and the rhesus macaque) and suids acquire cholesterol for progesterone steroidogenesis from high-density lipoprotein (HDL) via the uptake of cholesterol esters (Christenson and Devoto, 2003). Once cholesterol transport to the corpus luteum has occurred, cholesterol is converted to pregnenolone by the cleavage of a 6-carbon side chain by the cytochrome P450 side-chain cleavage enzyme (P450scc), with the resulting pregnenolone subsequently converted to progesterone by 3β-hydroxysteroid dehydrogenase (3β-HSD) (Hu *et al.*, 2010). Progesterone concentrations in the blood increase after ovulation; until day 14 of the bovine estrous cycle; and day 6 of the human menstrual cycle, plateauing until day 13; if maternal recognition of pregnancy does not occur, the corpus luteum regresses (Sartori *et al.*, 2004; Baerwald, Adams and Pierson, 2005). This is caused by successive pulses of prostaglandin F2α (PGF2α) in *Ruminantia, Rodentia* and *Primates*; upregulating fibroblast growth factor 2 (FGF2) inhibitors, suppressing angiogenic action in the corpus luteum; and induce the expression of pro-inflammatory cytokines, such as the tumor necrosis factor (TNF) family to promote programmed cell death in the corpus luteum via apoptosis (Stocco, Telleria and Gibori, 2007; Skarzynski *et al.*, 2013), after which the menstrual or estrous cycle resumes in non-pregnant animals.

1.4.3 Events in early pregnancy

In order for a successful pregnancy to occur, a number of specialised events must take place After ovulation, the sperm binds to the oocyte, not by the apex but at the midsection of the sperm head, after which phagocytosis of the acrosomal domain occurs and binding to the zona pellucida is facilitated by the interaction of complementary proteins in the acrosome to those on the egg surface (Bedford, 2014). While the mechanism of invasion into the eutherian zona pellucida is yet not fully understood, degradation of the zona pellucida by a sperm borne protease, has been observed in domestic pigs (Yi *et al.*, 2007), humans (Morales *et al.*, 2004) and mice (Pasten,

Morales and Kong, 2005), indicating that proteases in the acrosome may assist the penetration of the zona pellucida during fertilization in *Eutheria*.

Once the sperm has entered the zona pellucida, the male pronucleus is formed from nuclear material within the sperm cell and enters the oocyte, at which point the female and male pronuclei undergo chromosome formation, fusion and duplication as the fertilised oocyte forms a zygote (Gilbert, 2000). The zygote then undergoes the first cleavage, producing two daughter cells (or blastomeres), with cleavage of these cells repeating to form the 16-cell morula. As the morula continues to develop, a blastocyst is formed from the blastomeres, with the inner cell mass eventually forming the embryo, and the outer cell mass forming the trophoblast, from which the placenta is derived (Martini *et al.*, 2017).

Maternal recognition of pregnancy (MRP) is the result of the molecular signalling interactions between the maternal endometrium and developing conceptus (the embryo and associated extra embryonic membranes). The signalling molecule and mechanism of action of MRP varies across eutherian mammals. Successful MRP prevents the regression of the corpus luteum, maintains progesterone production, and facilitates successful early pregnancy (Michael Roberts, Xie and Mathialagan, 1996). In humans and non-human primates, this function is carried out by chorionic gonadotropin, secreted on days 8-10 post-fertilisation (Fishel, Edwards and Evans, 1984). Maternal recognition of pregnancy in rodents is mediated by two separate hormone expression events: the secretion of prolactin at the point of mating, and lactogenic hormones which are released at Day 12 of pregnancy, maintaining progesterone production until it can be carried out by the placenta (Bazer *et al.*, 2011). Maternal recognition of pregnancy in ruminants is mediated by Interferon tau (IFNT) (Godkin *et al.*, 1982), a Type I interferon which must be secreted in sufficient quantities at Day 16 post-fertilization in cattle in order for MRP to occur. In ovine and caprine pregnancy, IFNT is secreted between days 9-21 and days 16-21 respectively. The porcine MRP process involves the secretion of estrogen on Day 11 of pregnancy, altering the secretory patterns of prostaglandin F (PGF) to prevent its luteolytic action, redirecting PGF secretion to the uterine lumen. In contrast, neither canine or feline pregnancies have a MRP signalling event, with the corpus luteum acting in concert with the placenta throughout the

pregnancy to maintain sufficient progesterone production (Bazer *et al.*, 2011; Raheem, 2017). While the method of MRP varies, the purpose of pregnancy recognition is conserved, by preventing the degradation of the corpus luteum the embryo maintains progesterone production, enabling successful implantation and establishment of pregnancy. Along with preventing the degradation of the corpus luteum, the MRP signal has also previously been noted to alter the expression of transcripts within the maternal endometrium to facilitate the establishment of pregnancy (Table 1.5).

Human chorionic gonadotrophin (hCG) has been found to assist in the decidualization of endometrial stromal cells by upregulating cyclooxygenase-2 (*COX-2*) and prostaglandin E_2 (*PGE2*) expression, promoting the formation of the decidua during early pregnancy (Han, Lei and Rao, 1999). As the trophoblast cells invade the maternal endometrium, hCG supports this process and enables adequate invasion by upregulating matrix metalloproteases (MMPs) such as MMP-9, endopeptidases which degrade the extracellular matrix of target cells and downregulating tissue inhibitors of matrix metalloproteases 1, 2 and 3 (MMP1, MMP2 and MMP3) (Licht, Russu and Wildt, 2001; Fluhr *et al.*, 2008). Once the trophoblast has invaded, hCG enhances trophoblast differentiation into syncytiotrophoblast cells and promotes the formation of gap junctions to support this process (Shi *et al.*, 1993), along with regulating the formation of new blood vessels and remodelling of the maternal blood supply by upregulating vascular endothelial growth factor (VEGF), thereby promoting correct blood vessel organisation to facilitate healthy placentation (Licht, Russu and Wildt, 2001). Studies of the effect of hCG on the proteome of human endometrial luminal epithelial cells have found 193 proteins significantly differentially expressed in response to hCG treatment, with functions associated with cell communication, proliferation, differentiation and metabolism (Greening *et al.*, 2016). In primary human endometrial epithelial cells, hCG treatment increased the expression of cytokines which promote uterine receptivity and facilitate blastocyst migration and adhesion by enhancing adhesion of endometrial epithelial cells to fibronectin and collagen, molecules that make up the extracellular matrix of the blastocyst (Paiva *et al.*, 2011).

The ruminant MRP signal, Interferon Tau (IFNT) has also been shown to induce transcriptional changes separately from the prevention of corpus luteum degradation. IFNT is a type I interferon, a family which includes Interferon alpha (IFNA), Interferon beta (IFNB), Interferon delta (IFND) and Interferon omega (IFNW1) (Piehler *et al.*, 2012). All type I interferons bind the same Interferon alpha/beta receptor (IFNAR1/2), which activates multiple signalling cascades, including the Janus activated kinase-signal transducer and activator of transcription (JAK-STAT) pathway, the phosphatidylinositol-3-kinase (PI3K)-signalling pathway, tyrosine kinase 2 (TYK2) pathway and the protein kinase p38 transcription pathway (Platanias, 2005). Type I interferons also induce the expression of IFN-stimulated genes (ISGs) by inducing the expression of interferon-stimulated gene factor 3 (ISGF3), or by stimulating the formation of gamma-activation factors, which bind to gamma-activated sites in the promotor region of some ISGs (Mamane *et al.*, 1999). The ISGs induced by IFNT secretion include Lectin, Galactoside-binding, soluble, 15 (LGALS15), which crosslinks glycoprotein and glycolipid receptors to promote cell differentiation, adhesion and proliferation, along with trophoblast cell migration; cathepsin proteases, which assist in extracellular matrix degradation during implantation; and glucose transporter 1 (GLUT1), to support the transition to a glucose energy substrate as the conceptus is formed (Bazer, Spencer and Johnson, 2009). At Day 16 of bovine pregnancy, IFNT secretion by the conceptus was found to induce large-scale transcriptional changes in the maternal endometrium, with 764 genes found to be significantly differentially expressed, 514 of which were upregulated. Among these upregulated genes were genes related to viral resistance, antigenic factors, translation factors and peptidases. Upon downstream pathway analysis of differentially expressed genes at the point of maternal recognition of pregnancy, pathways relating to IFN signalling, bacterial recognition, antigen presentation and apoptosis signalling were upregulated; while pathways related to notch signalling, WNT/β-catenin signalling, corticotrophin releasing hormone signalling and cAMP response element binding protein (CREB) signalling in neurons were downregulated (Forde *et al.*, 2011). IFNT mediates transcriptional changes in the endometrium, changing the maternal transcriptomic profile as the conceptus elongates and signals its presence. A study of the alterations to the endometrium at Day 16 (MRP) comparing cyclic and pregnant heifers discovered 459 differentially expressed genes at Day 16. At the beginning of conceptus elongation (Day 13), only 16 differentially expressed genes were found, indicating that, while

conceptus-mediated changes to the endometrial environment begin prior to MRP, IFNT induces larger changes, possibly to prepare the endometrium for pregnancy. A pathway analysis of the 459 differentially expressed genes observed on Day 16 of bovine pregnancy revealed that genes of interest were implicated in phosphoionositide 3-kinases(PI3K) signalling; lipid and mineral metabolism; DNA recombination, replication and cell cycle; and bacteria and viruses (Forde *et al.*, 2012). When analysing the changes induced by IFNA2 (another type I interferon which binds to the IFNA1/2 receptor) to understand the transcriptional alterations specifically induced by IFNT in comparison to other embryo-derived factors, IFNA2 treatment induced the expression of ISGs related to the activation of the complement immune system and the regulation of T cell development, including IDO1, a gene found to prevent rejection of the allograft-like conceptus in mice (Bauersachs *et al.*, 2012).

Along with the MRP signal, the bovine conceptus has been found to secrete proteins which induce transcriptional changes independently of the MRP signal (Bauersachs *et al.*, 2012), suggesting that prior to implantation, the conceptus secretes specific molecules to facilitate its own survival. This was further investigated in the bovine endometrium in the pre-implantation period discovered 30 proteins present in Day 16 conceptus-conditioned media that are also present in the uterine luminal fluid of heifers at Day 16 of pregnancy, with the transcripts encoding 12 of these proteins at least 2-fold more abundant in conceptus conditioned media compared to the uterine luminal fluid of pregnant heifers at Day 16, leading to the hypothesis that they are produced by the conceptus to facilitate implantation (Forde *et al.*, 2015). These proteins were found to be conserved across *Eutheria*, to the extent that a bovine recombinant form of macrophage-capping protein (bCAPG) and protein disulfide-isomerase (P4HB), proteins produced by the conceptus, elicited a transcriptomic response in human Ishikawa immortalised endometrial epithelial cells, suggesting a conserved role in eutherian implantation (Tinning *et al.*, 2020).

1.4.3.1 Implantation

Before the conceptus (the embryo and associated extraembryonic membranes) can implant, it must hatch from the zona pellucida, the glycoprotein coat surrounding the embryo. In humans,

non-human primates and rodents, implantation occurs shortly after hatching (Seshagiri *et al.*, 2009). However, the ruminant conceptus undergoes a protracted elongation period after hatching from the zona pellucida, during which time the extraembryonic membranes differentiate and the allantois forms. The bovine blastocyst hatches on days 8-10 post-fertilisation and begins to elongate on Day 12, reaching 6cm in length by Day 16, at which stage maternal recognition of pregnancy occurs, reaching 20cm in length at the point of implantation (Brooks, Burns and Spencer, 2014). The ovine conceptus undergoes elongation on day 12, reaching 15-19cm at the point of implantation (Brooks, Burns and Spencer, 2014).

Within *Eutheria*, there are two major implantation strategies: invasive and non-invasive. These strategies will be discussed here in the context of well-studied species displaying each implantation type: *Homo sapiens* (invasive) and *Bos Taurus* (non-invasive).

Implantation in humans consists broadly of the positioning of the conceptus close to the uterine wall, adhesion of the trophoblast, the penetration of the trophoblast into the endometrial epithelium, and invasion of the fetal trophoblast cells into the endometrium to modify the maternal vasculature (Schlafke and Enders, 1975). This process is facilitated by extensive communication between the conceptus and the receptive maternal endometrium, which may only occur during the window of receptivity, a short period of time between Days 16-22 of the human menstrual cycle, initiated by the MRP signal (Kim and Kim, 2017). Before implantation, the conceptus first comes into contact with the endometrium in a process known as apposition. During this process, trophoblast cells undergo repeated contractions and expansions as they communicate with the maternal endometrium, while conceptus microvilli interact with villi on the surface of the uterine lumen (Emiliani *et al.*, 2005). During apposition, L-selectins expressed on the trophectoderm interact with carbohydrate ligands on the uterine lumen, with higher L-selectin expression on human trophoblast cells positively correlated with increased rates of successful implantation (Wang *et al.*, 2008). In addition, chemokines, such as Interleukin 8 (IL-8), are secreted by the conceptus, inducing the expression of adhesion molecules in the receptive endometrium (Dominguez *et al.*, 2005). Once the conceptus has found an appropriate implantation site, a polarity is established, positioning the inner cell mass of the conceptus

towards the endometrium, facilitated by interactions between chemokine receptors expressed on the trophectoderm and chemokines, such as Fractalkine (CX3CL1) expressed on the maternal endometrium (van Mourik, Macklon and Heijnen, 2009). During this window, alterations in the permeability of the endometrial vasculature are carried out by Cyclooxygenase-derived prostaglandins, facilitating adequate adhesion and invasion of the blastocyst (Lim *et al.*, 1997). Blastocyst attachment is mediated by adhesion molecules; such as cadherins, found to be expressed in both blastocyst cells and in uterine epithelial cells at the time of blastocyst attachment (Kadokawa *et al.*, 1989);

Trophoblast cells undergo morphological changes to become cytotrophoblast and syncytiotrophoblast cells and invade the endometrium (Kim and Kim, 2017). The embryo is fully embedded in the endometrium in humans and other mammals with invasive placental morphologies, such as *Rodentia (e.g. Mus musculus)*, *Primates (Pongo pongo, Homo sapiens), Chiroptera (Myotis lucifugis)* and *Eulipotyphyla (e.g. Erinaceus europaeus)*, with the fetal trophoblast cells combining with the uterine epithelial cells (McGowen *et al.*, 2014). Upon invasion, trophoblast cells utilise proteinases, such as matrix metalloproteinases, remodelling the maternal spiral arteries to allow a slower blood flow for sufficient exchange of molecules during pregnancy. The efficacy of this process has a direct effect on the competence of the resulting placenta, and by extension, the pregnancy itself, with inadequate trophoblast invasion being linked to placental disorders such as preclampsia (Allen C Enders and Blankenship, 1999; Hunkapiller *et al.*, 2011).

In cattle, the process of implantation is non-invasive, whereby the conceptus attaches to the maternal endometrium without the deep invasion of trophoblast cells into the maternal endometrium observed in invasive implantation (McGowen *et al.*, 2014). Bovine implantation begins once the conceptus has fully elongated (Bauersachs *et al.*, 2006). During the implantation period trophectoderm cells proliferate, a process facilitated by molecules secreted by the uterine lumen, termed the histotroph (Lonergan and Forde, 2014). Apposition begins at the embryonic disc at the centre of the elongated bovine conceptus, similarly to the human apposition process.The bovine conceptus is adhered to the endometrial epithelium , assisted by the

formation of binucleate cells, which fuse with placentomes on the endometrium (Fair, 2016) around Day 16 of pregnancy, with adhesion of the bovine trophectoderm and uterine lumen occurring around Day 18 (Guillomot, 1995). Once the conceptus has adhered to the uterine lumen, binucleate cells migrate from the trophectoderm into the uterine luminal epithelium and flatten, fusing with the uterine luminal epithelium and differentiating to form syncytial cells. These syncytia later cover the thicken epithelium of the bovine uterus, differentiating to form the placentomes that make up the bovine cotyledonary placenta (Bazer, Song and Thatcher, 2012). In direct contrast with the invasive implantation type, the trophoblast cells do not invade the maternal epithelium.

It is clear that the successful establishment of pregnancy in eutherian mammals requires extensive alterations to the maternal uterine environment, mediated by successive signals in early eutherian pregnancy. As discussed above, these signals induce transcriptional changes in order to prepare the maternal endometrium, the maternal contribution to the placenta, for implantation. While the required transcriptional changes may alter depending on species or implantation type (Table 1.5), these signals are necessary for the establishment of pregnancy in the eutherian mammals.

Table 1.5 Molecules associated with early pregnancy and the transcriptional alterations they induce in species representative of invasive (human) and non-invasive implantation and placental types.

Implantation Strategy	Signal	Species	Number of differentially expressed genes (DEGs)	Study
Invasive	P4	*Homo sapiens*	653 (454↑, 199↓)	(Chi *et al.*, 2019)
	MRP (hCG)	*Homo sapiens*	173 (130↑, 43↓)	(Li *et al.*, 2020)
	Conceptus produced signals	*Homo sapiens*	340 (184↑, 156↓)	(Altmäe *et al.*, 2012)
Non-invasive	P4	*Bos taurus*	4421 (2313↑, 2108↓)	(Forde *et al.*, 2009)
	MRP (IFNA)	*Bos taurus*	212 (173↑, 39↓)	(Bauersachs *et al.*, 2012)
	Conceptus produced signals	*Bos taurus*	764 (514↑, 250↓)	(Forde *et al.*, 2011)

The establishment of eutherian pregnancy requires extensive communication between the maternal endometrium and the conceptus. This communication is facilitated by molecules such as (i) progesterone (P4) secreted by the corpus luteum after ovulation; (ii) maternal recognition of pregnancy, a signal emitted by the conceptus to signal its presence to the maternal uterus, which is hCG in humans and IFNT in ruminants; and (iii) Other conceptus produced signals which may aid in promoting uterine receptivity during implantation. These signals induce specific transcriptional changes, with the number of differentially expressed genes (DEGs) associated with each signal outlined.

1.4.4 Placental invasion and formation

After implantation the embryo undergoes gastrulation, whereby three distinct embryonic structures form; the endoderm, or inner layer; the mesoderm, or middle layer; and the trophectoderm, or outer layer (Bellomo *et al.*, 1996). Early placentation begins with the formation of the yolk sac from the embryonic trophectoderm and endoderm. As the mesoderm layer attaches to the endoderm, blood vessels begin to differentiate, providing a vascular system for the yolk sac, or choriovitelline placenta (Freyer and Renfree, 2009). In *Rodentia* and

Lagomorphia, the yolk sac remains throughout pregnancy and inverts, with the endoderm layer differentiating into an epithelial layer that faces the maternal endometrium, facilitating exchange (Chavatte-Palmer and Tarrade, 2016). In other mammals, once the yolk sac placenta is established, the allantois, a fluid filled extra-embryonic sac, grows and displaces the yolk sac, merging with the trophoblast to form a chorioallantoic placenta. In primates, the yolk sac degenerates quickly, without a true choriovitelline placenta forming, before a chorioallantoic placenta is established (Carter and Enders, 2004).

Aims of this Thesis

The eutherian placenta is an individualised tissue with a specialised function. It evolved once on the eutherian stem lineage and was never subsequently lost. While the main function of the eutherian placenta remains the same, placental morphology varies across *Eutheria*. The overall aim of this thesis is to investigate innovation in both protein coding and regulatory elements at the root of placental mammals and the role of those innovations in early pregnancy events.

The specific aims are:

1) To identify protein coding genes that underwent selective pressure on the stem eutherian lineage
2) To identify miRNAs that emerged in therian and eutherian mammals using a high-quality miRNA database
3) To identify putative regulatory interactions between novel therian and eutherian miRNAs and positively selected genes identified
4) To determine the expression of stem lineage miRNAs in uterine endometrium
5) To assess if miRNA expression was regulated by molecules that are important for early pregnancy success and if they are different in species with different implantation strategies.
6) To investigate whether positively selected targets of stem lineage miRNAs were biologically regulated in models of early pregnancy

Chapter 2: Assessment of selective pressure variation on the placental stem lineage – a protein coding perspective

2.1 Aims of this chapter

1. To assess selective pressure variation in single gene orthologs on the stem Eutherian lineage
2. To determine if there are positively selected sites in single gene orthologs that arose on the stem Eutherian lineage and that were subsequently conserved in all extant mammals
3. To identify specific functional categories of any genes under positive selection on the stem Eutherian lineage

2.2 Introduction

The emergence of the placenta as an organ in mammals likely required major alterations in both protein coding and regulatory elements. In this chapter we explore the role of selective pressure variation in facilitating novel function on the stem mammal lineage in protein coding genes. This altered function may be directly or indirectly involved in placentation or indeed in other aspects of placental mammal evolution such as lactation, indeed these functional shifts may be due to exposure to new pathogens or other environmental factors. We will attempt to tease apart these different contributing factors based on functional annotation of the protein coding elements and their known involvement in specific processes/pathways.

Selective pressure variation has been previously investigated in *Eutheria*, across the entire clade and on a lineage-specific basis. Evidence of selective pressure variation was found in the

vertebrate-specific Cysteine-Rich Secretory Proteins (CRISPs), proteins which are highly expressed in male reproductive tissues, with CRISP2 expressed in the testis and acting as an essential unit in the head of sperm (Koppers, Reddy and O'Bryan, 2011); and CRISP1 and CRISP3 secreted into the epididymal lumen during sperm maturation (Gibbs, Roelants and O'Bryan, 2008). Evidence of positive selection was found in all three mammalian CRISPs, with the strongest signature in CRISP3 (Vicens and Treviño, 2018). Signatures of positive selection were also found in Abnormal Spindle-like Microcephaly-associated (ASPM) and CDK5 Regulatory Subunit Associated Protein 2 (CDK5RAP2), two genes associated with microcephaly, leading to the hypothesis that selective pressure on these genes may have contributed to the evolution of larger brain size (Montgomery and Mundy, 2014).

A study using six high-quality mammalian genomes found 400 genes with signatures of positive selection (Kosiol *et al.*, 2008). This study identified functional enrichment within these positively selected genes for Gene Ontology categories related to immunity, including NK-mediated immunity, cytokine-mediated immunity and complement-mediated immunity. Other enriched immune-related pathways include coagulation, apoptosis and the membrane attack complex. The significant enrichment of these immune-related functions within positively selected genes in mammals may indicate adaptation to novel pathogens or environments as mammals diversified. This hypothesis is supported by further evidence of positive selection in mammalian Toll-like receptors (Webb *et al.*, 2015) cell surface receptors which serve to recognise pathogens and initiate the adaptive immune response (Areal, Abrantes and Esteves, 2011); the dipeptidyl peptidase 4 (DPP4) tumor suppressor gene (Sironi *et al.*, 2015); and in the receptor binding region of Interleukin-2 (IL-2) (O'Connell and McInerney, 2005).

The post-extinction radiation of placental mammals at the end of the Cretaceous was marked by extensive morphological innovation as they diversified and adapted to new environments (Halliday and Goswami, 2016). The levels of positive selection reported from comparative genomic analyses thus far within *Eutheria* may be indicative of this extensive adaptation (Meredith *et al.*, 2011).

There are a number of cases of selective pressure variation in protein coding genes that have had a direct role in placentation. The primate-specific hormone chorionic gonadotropin, the pregnancy recognition signal in humans is comprised of an alpha (CGα) and a beta (CGβ) subunit (Morgan and Canfield, 1971). The CGα subunit is identical to that of luteinizing hormone subunit α. The CGβ subunit originated from a duplication of the luteinizing hormone subunit (LHβ) in the ancestor of anthropoid primates. Following further duplications, with CGβ gene copy number varying between primate species, the 3' tail region of the CGβ gene was under positive selection in New World monkeys, along with weak positive selection on the 3' tail region in the ancestral hominid species. Analysis of site-specific substitutions found an alanine to serine substitution in the ancestral hominid species, leading to the hypothesis that this occurred in order to mitigate metabolic effects on chorionic gonadotropin (Maston and Ruvolo, 2002).

Galectins, proteins implicated in immune tolerance that are expressed at the maternal-fetal interface were also found to be under positive selection in primates, including 5 galectins specifically expressed in the placenta (Than *et al.*, 2009). Evidence of positive selection was also discovered in the cadherins, a family of cell surface adhesion proteins (Summers and Crespi, 2005). Three cadherin genes (E-cadherin, P-cadherin and VE-cadherin) that are expressed in the placenta were found to be under positive selection and all were implicated in placental function: E-cadherin being involved in trophoblast invasion (Floridon *et al.*, 2000); P-cadherin being primarily expressed in the placenta (Haig, 1996); and VE-cadherin modulating trophoblast invasion and angiogenesis (Zhou *et al.*, 1997). Whereas H-cadherin, which is not expressed in the placenta, did not have a signature of positive selection (Summers and Crespi, 2005).

By combining the resolved eutherian phylogeny (Moran, Morgan and O'Connell, 2015; Tarver *et al.*, 2016) with a genome-wide analysis of selective pressure variation on the stem eutherian lineage, we wished to determine the levels of selective pressure variation on single gene orthologs which may have contributed to the evolution of mammal placenta. Conducting a selective pressure analysis has many potential difficulties which must be considered. The method chosen for this analysis, CodeML (Yang, 2007), uses codon-based models of evolution to analyse variation of selective pressure. The use of codon-based models to detect selective pressure has

been debated (Hughes and Friedman, 2008; Zhai *et al.*, 2012). However, CodeML is subject to regular updates and has been extensively validated, with the branch-site model found to be robust even when testing GC-rich sequences (Yang and dos Reis, 2010; Gharib and Robinson-Rechavi, 2013). These publications have, however, emphasized the importance of an accurate phylogeny, the lack of short branches, an appropriate multiple sequence alignment, and rigorous downstream statistical testing.

2.3 Materials and Methods

2.3.1 Assembly of Representative Species Dataset

Taking into account previous studies, noting that selective pressure analyses display higher type I error rates on larger phylogenies (Anisimova, Nielsen and Yang, 2003), we chose 10 placental mammal species from Ensembl 90 (Yates *et al.*, 2016) for this analysis from available eutherian species (Table 2.1). These were chosen (1) to represent major placental morphologies (Figure 2.1), and (2) based on the highest quality data available for each clade (determined by coverage and assembly level). If multiple candidate species within the same clade had the same placental type, the species with the highest quality genome was chosen. Four outgroup species were chosen, representing *Monotremata, Marsupiala, Aves* and *Teleosti* (Table 2.2).

Table 2.1 Candidate Species for Selective Pressure Analysis

Species	Accession Number	Order	Placenta Type	Coverage
Bushbaby	OtoGar3	Primate	Diffuse/Epitheliochorial	137x
Chimpanzee	CHIMP2.1.4	Primate	Haemochorial/villous with an intervillous space	97% of genome
Gibbon	Nleu1.0	Primate	Haemochorial/villous with an intervillous space	5.6x
Gorilla	gorGor3.1	Primate	Haemochorial/villous with an intervillous space	35x
Human	GRCh38.p7	Primate	Haemochorial/villous with an intervillous space	
Macaque	Mmul_8.0.1	Primate	haemochorial/villous with an intervillous space	47.4x
Marmoset	C_jacchus3.2.1	Primate	haemonochorial/villous	6.6x
Mouse Lemur	Mmur_2.0	Primate	mixed: endothelial/labrynthine w/ epitheliochorial ring	221.6x
Olive Baboon	PapAnu2.0	Primate	haemochorial/villous	85x Illumina
Orangutan	PPYG2	Primate	Haemochorial/villous with an intervillous space	
Tarsier	tarSyr1	Primate	haemochorial/discoid	1.82x
Vervet-AGM	ChISab1.1	Primate	haemochorial/villous	95x
Guinea Pig	cavPor3	Rodent	discoidal/haemomonochorial	6.79x
Kangaroo Rat	dipOrd1	Rodent	labrynthine/chorioallantoic	1.85x
Mouse	GRCm38.p4	Rodent	discoid/chorioallantoic/labrynthine?/has trophospongium	
Pika	OchPri2.0	Rodent	Discoid/	1.93x
Rabbit	OryCun2.0	Rodent	labrynthine, hemodichorial, chorioallantoic placenta	7x

Squirrel	spetri2	Rodent	Discoid/	495.1
Tree Shrew	tupBel1	Rodent	Bidiscoidal, labrynthine, endotheliochorial, chroioallantoic placenta	2x
Alpaca	vicPac1	Laurasiatheria	Diffuse/Epitheliochorial	Low - 2.51X
Cat	Felis_catus_6.2	Laurasiatheria	endothelial-chorial/?	12x 454
Cow	UMD3.1	Laurasiatheria	Primarily epitheliochorial + ? Small regions of a syndesmochorial type/villous	9x sanger
Dog	CanFam3.1	Laurasiatheria	endotheliochorial/villous?	7x Sanger
Dolphin	turTru1	Laurasiatheria	Diffuse/epitheliochorial	2.59
Ferret	MusPutFur1.0	Laurasiatheria	endotheliochorial/diffuse?	162x Illumina
Hedgehog	eriEur1	Laurasiatheria	haemonochorial chorioallantoic	1.82x
Horse	Equ Cab 2	Laurasiatheria	diffuse/Epitheliochorial	6.79x
Megabat	pteVam1	Laurasiatheria	Generally discoidal/chorioallantoic/support for both hemochorial and andotheliochorial	2.63x
Microbat	Myoluc2.0	Laurasiatheria	Generally discoidal/chorioallantoic/support for both hemochorial and andotheliochorial	7x
Panda	ailMel1	Laurasiatheria	no description	60x Illumina
Pig	Sscrofa10.2	Laurasiatheria	diffuse/Epitheliochorial	24x Illumin
Sheep	Oar_v3.1	Laurasiatheria	polycotyledonary	142x Illumina

Shrew	sorAra1	Laurasiatheria	Endothelio-endothelial??	1.9X
Elephant	Loxafr3.0	Afrotheria	Endotheliochorial/villous/zonary	7x
Hyrax	proCap1	Afrotheria	hemochorial/villous/zonary	2.19x
Lesser hedgehog tenrec	TENREC	Afrotheria	discoid, hemochorial + region of paraplacenta	2x
Armadillo	Dasnov3.0	Xenarthra	villous/zonary/slightly cotyledonary	6x Sanger
Sloth	choHof1	Xenarthra	endotheliochorial/lamellar/multicotyledonary	2.05x
Tasmanian Devil	Devil_refv70	Marsupial	non-invasive	85x Illumina
Opossum	monDom5	Marsupial	non-invasive	7.33x
Platypus	OANA5	Monotreme	n/a	6x
Zebrafish	GRCz10	Fish	Yolk sac	n/a

Table 2.2 Vertebrate genomes sampled for analysis of heterogeneity in selective pressure.

Clade	Species	Version	Genome Quality
Teleosti	Zebrafish	GRCz11	Full Genome, Chromosome Level
Aves	Chicken	Gallus_gallus-5.0	70X Coverage, Chromosome Level
Monotremes	Platypus	OANA5	6X Coverage, Chromosome Level
Marsupialia	Opossum	monDom5	7.33X Coverage, Chromosome Level
Eutheria	Elephant	Loxafr3.0	7X Coverage, Scaffold Level
	Armadillo	Dasnov3.0	6X Coverage, Scaffold Level
	Mouse	GRCm38.p6	High Quality Reference Assembly
	Human	GRCh38.p12	High Quality Reference Assembly
	Mouse Lemur	Mmur_3.0	221.6X Coverage, Chromosome Level
	Bushbaby	OtoGar3	137X Coverage, Scaffold Level
	Cat	Felis_catus_9.0	72X Coverage, Chromosome Level
	Microbat	Myoluc2.0	7X Coverage, Scaffold Level
	Horse	Equ Cab 2	6.79X Coverage, Chromosome Level
	Cow	UMD3.1	9X Coverage, Chromosome Level

Candidate vertebrate genomes were selected from Ensembl 90 (Yates *et al.*, 2016) *based on clade, placental type (within Eutheria) and genome quality. A final group of 14 genomes were chosen, with four outgroups from Fish, Aves, Monotremata, Marsupialia and 10 eutherian species. If multiple candidate species had the same placental type within a lineage, the species with the highest quality genome was chosen.*

Ensembl annotated gene families were accessed from Ensembl 90 (accessed 28th November 2017) (Yates *et al.*, 2016), with each sequence appended with the protein family identifier, gene ID and transcript ID. There was a total of 7,614 gene families which had gene family members in all 14 species. Nucleotide sequences for each of these genomes were also accessed to construct nucleotide alignments for CodeML analysis. Using the 'clean' function in the VESPA pipeline, sequences were checked to ensure they had complete codons (Webb, Walsh and O'Connell, 2016a).

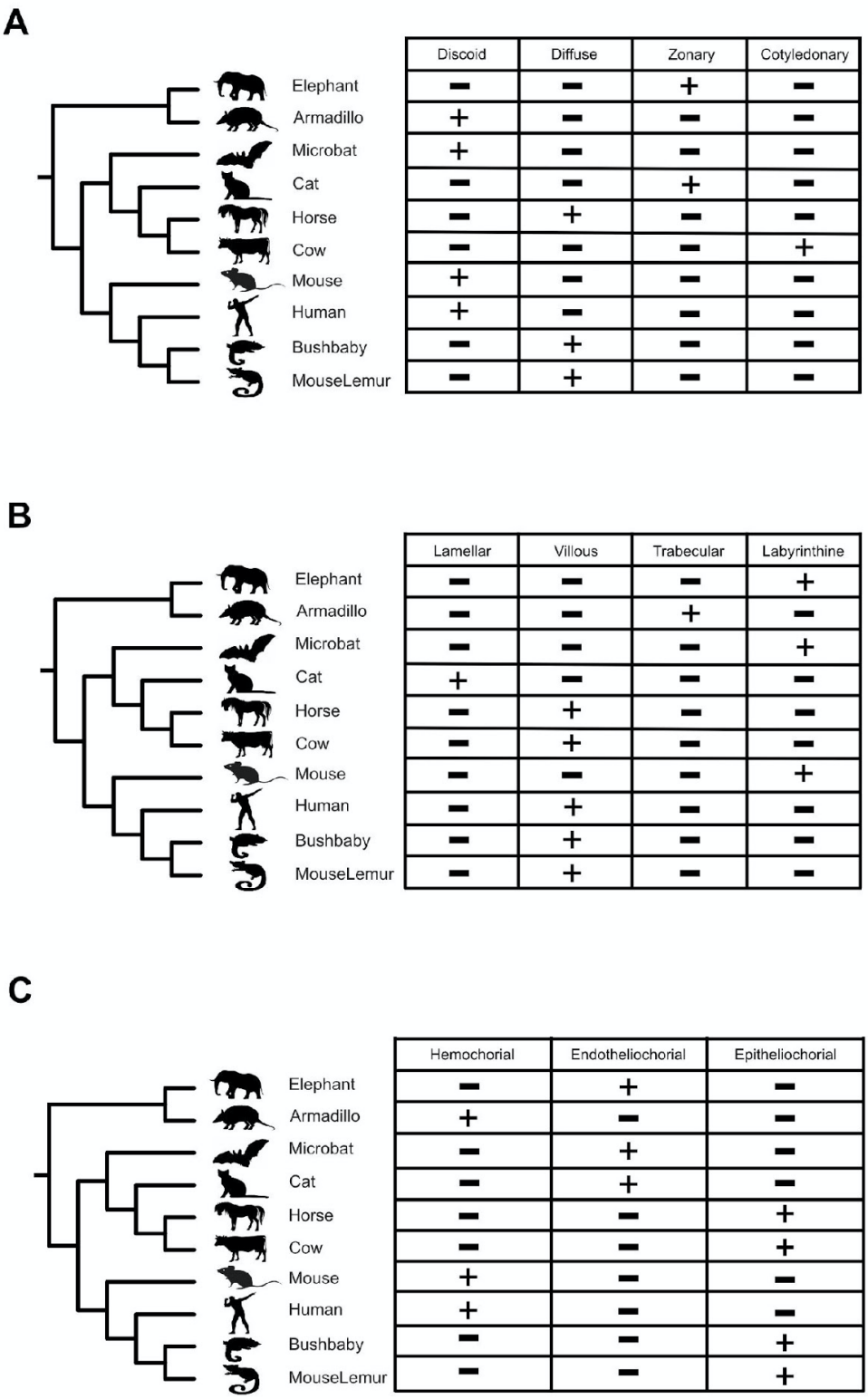


Figure 2.1 Phylogeny and placental morphology of representative species included in selective pressure analysis. *14 representative species (Four outgroups and 10 eutherian species) were chosen from Ensembl 90* (Yates *et al.*, 2016) *based on genome quality within their respective*

*clades and the placental morphology they represented. Placental morphologies are represented in a presence absence matrix showing **A)** shape; **B)** interdigitation; and **C)** interhaemal barrier.*

2.3.2 Single Gene Ortholog Annotation

Single gene orthologs (SGOs) across *Eutheria* were identified using the program 'GetSGO.py' (Electronic Appendix: VESPA Analysis "GetSGO.py"), which searches for families with one sequence from each species with the gene of interest. In this way, 2656 single gene orthologous families were identified.

2.3.3 Protein Family Alignment and Filtering

Previous work has determined that the accuracy of a selective pressure analysis is dependent on the inclusion of sufficient branches in each gene tree, finding an increase of 30% in the power of Likelihood Ratio Tests (LRTs) in detecting positive selection in trees with 8 taxa versus trees with 4 taxa (Anisimova and Yang, 2007). Therefore, single gene orthologous families with ≥7 members were selected, resulting in 1,704 /2,656 SGOs for further analysis .

Multiple sequence alignments (MSAs) were generated for each of the 1,704 SGOs using MUSCLE v3.8.1551 (Edgar, 2004) and MAFFT v7.310 (Katoh *et al.*, 2002) using default settings. The MUSCLE and MAFFT alignments were subsequently compared using the 'metal_compare' function in the VESPA pipeline (Webb, Walsh and O'Connell, 2016a). This function first scores alignments using the d_{pos} metric in metAL (Blackburne and Whelan, 2011), examining the difference in gap positions between two multiple sequence alignments. Secondly, the function uses norMD (Thompson *et al.*, 2001) to evaluate sequence similarity within an alignment on a column-by-column basis. Using these combined scores, the optimal MSA for each candidate gene family was selected. Alignments produced using MAFFT (Katoh *et al.*, 2002) were statistically significant for 87% (1,488) of SGOs, whilst MUSCLE (Edgar, 2004) was the best scoring alignment method for 13% (216) SGOs.

2.3.4 Selection Analysis

The 1,704 aligned candidate gene families were mapped onto the resolved placental mammal phylogeny (Moran, Morgan and O'Connell, 2015; Tarver *et al.*, 2016) using the 'infer_genetree' function in the VESPA pipeline (Webb, Walsh and O'Connell, 2016a). The aligned amino acids were used in combination with the corresponding nucleotide data to create aligned nucleotide files using "map_alignments" function in VESPA, which uses amino acid MSAs to map nucleotide sequences to their corresponding positions, conserving the alignment structure.

The CodeML directory structure (Figure 2.2A) was created using the "codeml_setup" function in VESPA, using the mapped nucleotide alignments and the inferred gene tree. As selective pressure variation was being investigated on a lineage level, a branch label table (Figure 2.2B) was created using the "create_branch" function, specifying the eutherian lineage.

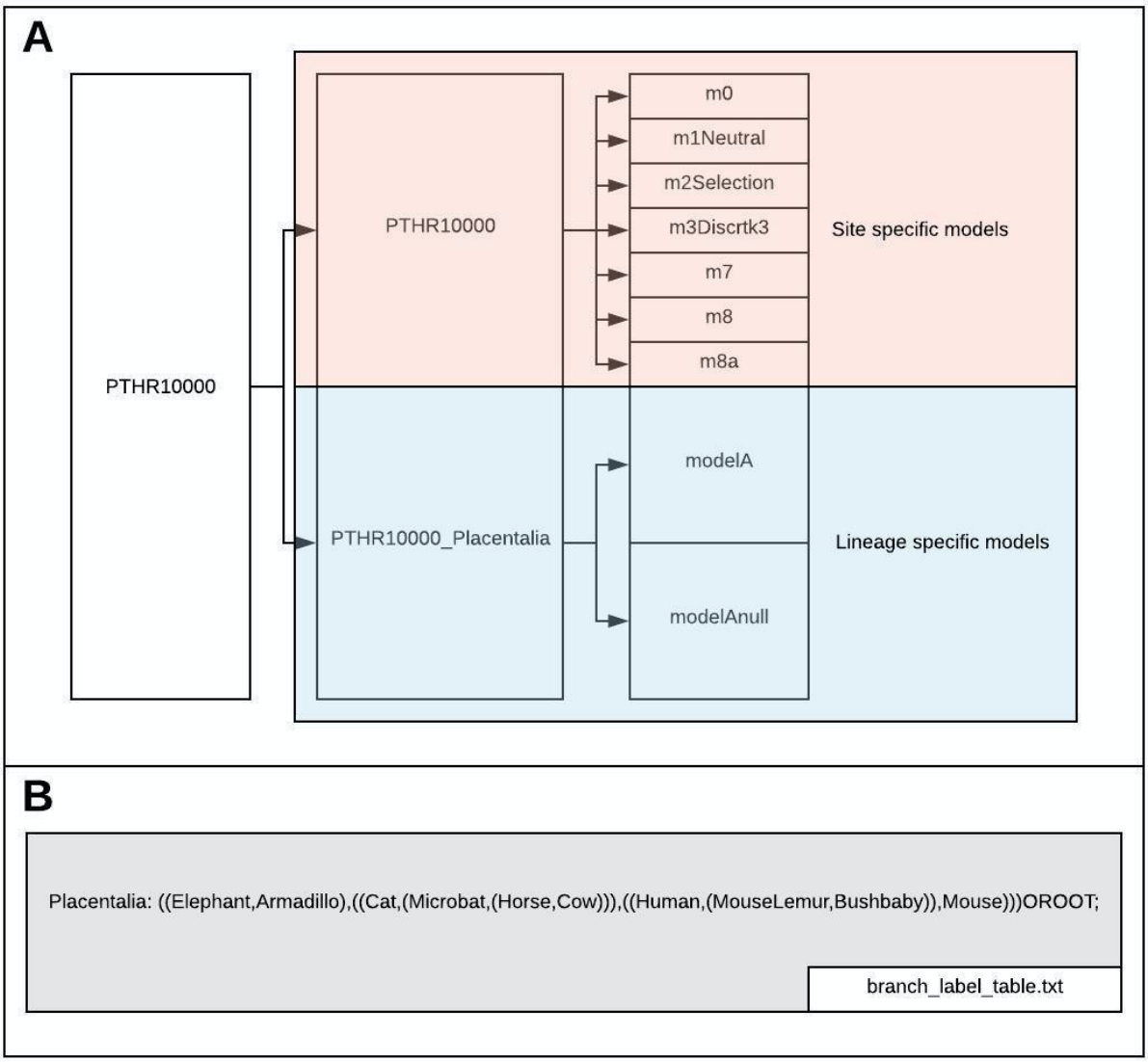


Figure 2.2 Example nested directory structure in CodeML, reproduced from Chapter 1. A: *Graphical depiction of CodeML directory structure, with gene family PTHR10000 having two nested directories (PTHR10000, PTHR10000_Placentalia) containing site specific and lineage specific models. Lineage specific models are created by labelling the branch of interest, in this case Eutheria (or Placentalia) with a branch label table as in* ***B.***

Selective pressure variation is assessed by comparing the amount of nonsynonymous nucleotide substitutions per nonsynonymous site (dN) to synonymous nucleotide substitutions per synonymous (dS) and quantifying the dN/dS ratio(ω); where a ratio of greater than 1 suggests positive selection (Zhang, Nielsen and Yang, 2005; Yang, 2007). The ratio of positive selection can be identified on a site-specific basis or, in this case, a lineage specific basis. The models used are described in detail in Sections 1.2.3.3 and 1.2.3.4 and are listed in Table 2.3.

Table 2.3 Vertebrate genomes sampled for analysis of heterogeneity in selective pressure.

Model	Free parameters	Foreground	Background
M1Neutral	p_0 , $\omega_0 < 1$	n/a	n/a
modelA	p_0 ,ω_0, p_2 , ω_2	$\omega_2 > 1$	$0 < \omega_0 < 1$, $\omega_1 = 1$
modelAnull	p_0 ,ω_0, p_1	$\omega_1 = 1$	$0 < \omega_0 < 1$, $\omega_1 = 1$

Table of codon models used to analyse selective pressure variation on the stem Eutherian lineage in the CodeML package, where ω_n is the dN/dS ratio, and p_n is the proportion of sites under the corresponding ω ratio (Yang, 2007).

Once the models had been compared to candidate SGOs and their log likelihood (lnL) calculated, the CodeML output was processed using the 'codeml_reader' function in the VESPA pipeline. This extract the lnL values and associated parameters for each model tested in order for likelihood ratio tests (LRTs) to be performed, in order to determine which model is the best fit for the data.

For each gene family, m1Neutral was compared to modelA using likelihood ratio tests, which determine the fit of a model by calculating 2 times the difference in log likelihood (lnL) between these models (2Δl), followed by a χ^2 test with varying degrees of freedom depending on the model used. The degrees of freedom of lineage-specific models are: M1neutral (df = 1), modelA (df = 10) and modelAnull (df = 9). If 2Δl> χ^2, the model with a greater number of free parameters (in this case, modelA) is determined to be significant. If modelA is determined to fit the data better than m1Neutral, modelA is then compared to modelAnull under the same parameters. If the null model is rejected, modelA is again determined to be significant. LRTs were performed on both alternate models as per the PAML manual (Yang, 1997) as the m1Neutral model may introduce false positives, while the null model has been documented as a more conservative alternative.

If modelA is determined to be significant, the posterior probability of any particular codon in the alignment being under positive selection is calculated using the Bayes Empirical Bayes (BEB) method (Zhang, Nielsen and Yang, 2005), whereby if the probability of the ω > 1 is >0.5, the site is expected to be under positive selection.

Of the 1,704 SGOs with ≥7 members, 267 had no non-Eutherian outgroup and could not be analysed .. . These 267 gene families could not be analysed by CodeML, as there was no ancestral codon state to compare to. This is required to determine the occurrence of substitution.

2.3.5 Assessment of levels of conservation of positively selected residues across Eutheria

For this analysis, substitutions that were introduced on the stem Eutherian lineage and became fixed in all extant species were of interest. The program 'post_reader.py' (Orr, 2018), when the 'U' option is selected, analyses the alignments of gene families under putative selective pressure to identify unaltered substitutions unique to the species analysed where the corresponding position in outgroup species was not blank. Gene families that passed this filtering step were assessed in order to determine the level of selective pressure variation present within the data.

2.3.6 Retrieval of annotated functions of filtered gene families under positive selection on the stem Eutherian lineage

Gene families that i) rejected the null and neutral models in likelihood ratio tests, ii) underwent an unreversed substitution on the stem Eutherian lineage, and iii) are conserved in the analysed species, were subject to functional annotation using GO Slim Biological Process terms and an enrichment analysis in PANTHERv.14 (Muruganujan *et al.*, 2018). To determine whether the data was significantly enriched for interactions among the gene families we used the STRINGv.11 database of interactions (Szklarczyk *et al.*, 2019). STRINGv.11 extracts data on gene-gene interactions from experimental (BIND, DIP, GRID, HPRD, IntAct, MIND and PID) and curated database sources (Biocarta, BioCyc, GO, KEGG and Reactome) (Szklarczyk *et al.*, 2019).

2.4 Results

2.4.1 115 conserved gene families underwent positive selection on the stem Eutherian lineage

Once candidate gene families had been assessed for selective pressure variation, 318 gene families had rejected the m1neutral and modelANull models in likelihood ratio tests. Gene families with substitutions that were reversed, or families with alignment gaps in outgroup species were removed, leaving 237 gene families (Supplementary Table 2). Of these 237, 115 gene underwent a substitution on the stem Eutherian lineage that remained fixed and was advantageous and will be referred to as PSStemFixed (Positively Selected Stem lineage Fixed substitutions) throughout this thesis (Supplementary Table 3) and 81 gene families were found to be have undergone positive selection on the stem Eutherian lineage, with the substitution later being altered or reverting to an ancestral state (depicted in Figure 2.3). As highlighted in Figure 2.3, outgroup residues includes AGC, GAT and GAC. This residue is altered in *Eutheria to* AGT. However, this residue is altered in Cow (AGC). The progressive filtering of this dataset is summarised in Table 2.4. This analysis was focussed on the 115 genes under positive selection where the positively selected residue/s were subsequently fixed and present in all extant Eutherian lineages.

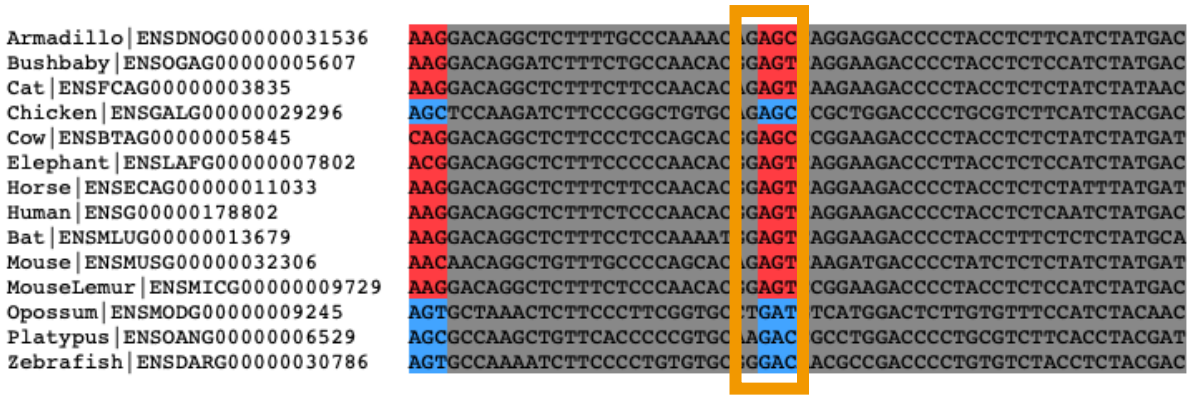


Figure 2.3 Reversed substitution in eutherian gene family. *Evidence of residues shared between eutherian species (red) and outgroups (blue) in PTHR10309. The residue highlighted in orange is the same in Chicken and Cow. Gene families were filtered using Post Reader* (Orr, 2018)*, to remove substitutions that reverted to an ancestral state.*

.

Table 2.4 Stepwise filtering of candidate gene families during selective pressure analysis.

Dataset	Number of Gene Families
All Families	7,596
Single Gene Orthologues	2656
CodeML suitable	1704
CodeML positively selected	318
Filtered positively selected	237
Positively selected and fixed in *Eutheria*	115

Table showing loss of gene families from the pipeline as it progressed. Analysis began with 7596 gene families. Upon filtering for single gene orthologues, 2656 gene families remained. Of these, families with ≥7 members were selected, as families with less members do not have the statistical power in CodeML to detect selective pressure variation. After analysis using CodeML, 318 families best fit models of positive selection. After removing gene families with substitutions that later reverted to the ancestral state, 237 gene families were found to have undergone selective pressure variation on the eutherian stem lineage, with 115 of these families remaining conserved in sampled species.

2.4.2 Functional Assessment

Functional annotation was performed in PANTHERv.14 (Muruganujan *et al.*, 2018) (Figure 2.4), in order to examine conserved positively selected genes on a functional basis. Statistically significantly enriched genes are detailed in Table 2.5.

Table 2.5 Statistically significantly enriched functions in SGOs that underwent positive selection on the stem Eutherian lineage where the positively selected residue was subsequently fixed in all extant Eutheria

Function	Implicated Genes	P value
Establishment of protein localization to telomere	WRAP53, ACD, TERT	$2.7e^{-4}$
Dyskeratosis congenita	WRAP53, ACD, TERT	$1.3e^{-3}$
Telomere maintenance via telomerase	WRAP53, ACD, TERT	$2.7e^{-3}$

PANTHERv.14 (Muruganujan *et al.*, 2018) *was used to analyse 115 conserved gene families that underwent positive selection on the stem eutherian lineage in order to determine enriched functions. 'Function' refers to annotated GO term from PANTHER. 'Implicated Genes' refers to the genes implicated in significantly enriched functions. 'P-value' is the p-value obtained from PANTHER.*

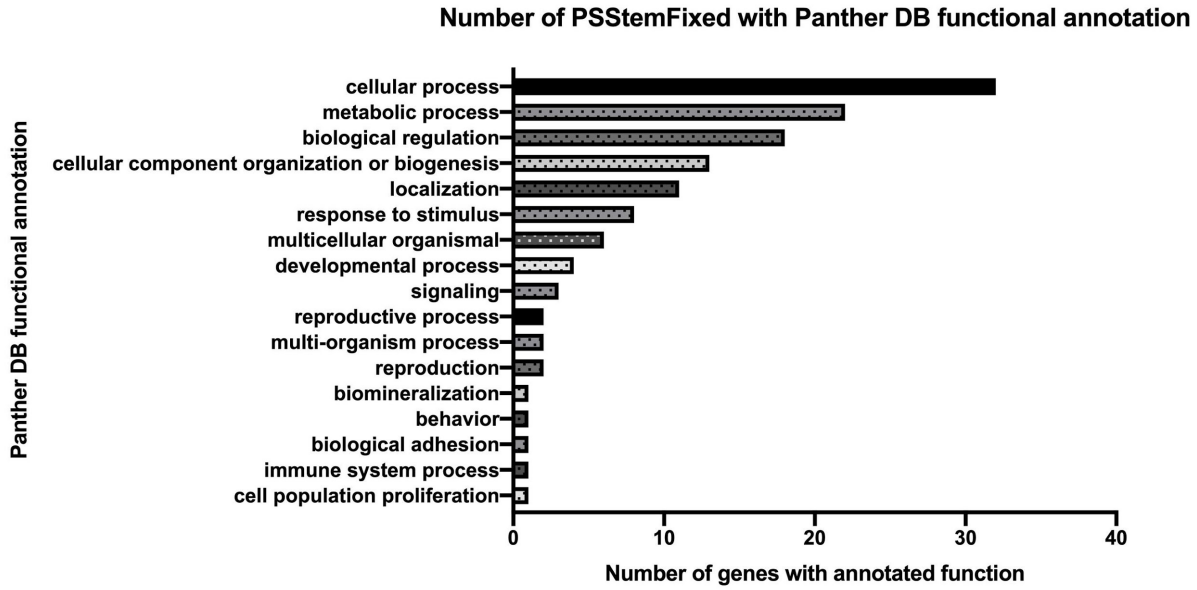


Figure 2.4 Functional annotation using Gene Ontology Biological Process terms of115 SGOs that underwent positive selection on the stem eutherian lineage and where the amino acid substitution was fixed on all extant Eutheria tested. *The absolute number (out of 115) of positively selected genes in a given category are shown in the X axis and the functional annotations on the Y-axis.*

These 115 PSStemFixed were found to be enriched for high-confidence (>0.700 interaction score) gene-gene interactions, based on experimental and database evidence (p=0.0468) in STRINGv.11 (Szklarczyk *et al.*, 2019) (Figure 2.5).

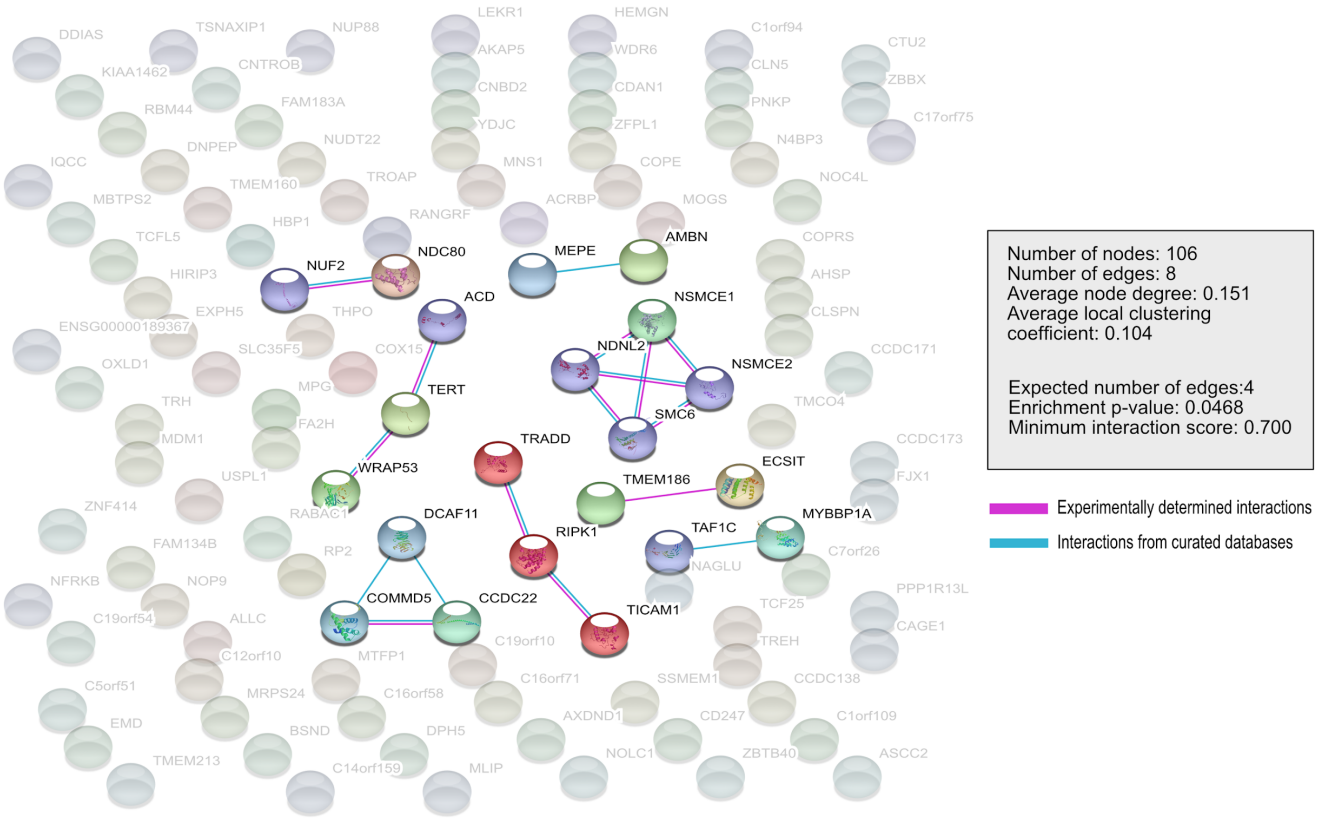


Figure 2.5 String interaction network of the 115 PSStemFixed genes*. Background node, with no high confidence interactions from experimental and database sources are translucent. Nodes with high confidence interactions from experimentally determined (pink lines) or curated database (blue lines) sources are depicted in colour. The network was found to be significantly enriched for gene-gene interactions (p=0.0468).*

2.5 Discussion

In summary, 237 gene families were found to have undergone positive selection on the stem Eutherian lineage with the substitution not later being altered. One hundred and fifteen of these 237 gene families were shared between all sampled Eutherian species.

The outcome of a selective pressure analysis is dependent on the quality of the data and the fit of the models used. Multiple sequence alignments have been previously shown to impact the outcome of a selective pressure analysis, with incorrect alignment methods increasing the number of nonsynonymous substitutions identified (Schneider *et al.*, 2009). The comparison of different alignment methods to identify the best fit to the data has been recommended in order to prevent alignment-related issues in a selective pressure analysis (Anisimova, M., Cannarozzi, G., Liberles, 2010). By implementing alignment construction using both MUSCLE (Edgar, 2004) and MAFFT (Katoh *et al.*, 2002), and by scoring and comparing alignments using 'metal_compare' (Thompson *et al.*, 2001; Blackburne and Whelan, 2011) implemented in the VESPA pipeline (Webb, Walsh and O'Connell, 2016b), we ensured that the best available alignment for each gene family was used.

The 115 PSStemFixed families comprises 6.74% of the 1,704 SGOs that underwent selective pressure analysis in CodeML and 1.5% of the total 7,565 orthologous gene families shared between the species studied. This agrees with previous publications on the level of positive selective pressure in mammalian lineages which have reported between 0 and 16% of protein coding genes with signatures of positive selection, discussed below. Studies of positive selection in primates found 0 loci to be under positive selection in *Homo sapiens* and 0.11-20% of loci under positive selection in *Pan troglodytes* (Eyre-Walker, 2006). Independently, other studies of simian primates identify 3% of protein coding genes to be under positive selection, comprising 0.014% of the genome (van der Lee *et al.*, 2017). Positive selection has been detected at a comparable level (i.e. between 1 and 16 %) in a variety of Laurasiatherian mammals. For example, in the bat lineage (*Chiroptera) a total of* 16% of analysed genes displayed signatures of positive selection (Hawkins *et al.*, 2019) and a similar analysis of 4 marine mammal species (*O. rosmarus, O. orca,*

T. manatus latirostris and T. truncates) found 1.13% displaying signatures of positive selection (Foote *et al.*, 2015). The variance in levels of positive selection detected in different Laurasiatherian lineages may be a product of genome quality (Booker, Jackson and Keightley, 2017) or issues with phylogeny resulting in false positive results (Anisimova, Nielsen and Yang, 2003; Hallström *et al.*, 2011). Due to these observations, previous studies may not provide an adequate benchmark of the expected amount of positive selective pressure in a genome, the results of this study investigating selective pressure at the stem Eutherian lineage fall within the range previously reported for Mammals.

We find an enrichment for functions related to telomere activity in the 115 genes that underwent positive selection on the stem mammal lineage, where the substitution is unique, unreversed, and conserved on all branches of the Eutherian tree (i.e. the PSStemFixed gene set). In addition, the analysis of gene-gene interactions showed an enrichment for interactions related to telomere activity in our positively selected gene set.

Telomeres are 'TTAGGG' repeats that cap the ends of chromosomes, preventing chromosomal alterations by a DNA damage response (de Lange, 2002). Telomeres are appended to chromosome ends by the ribonucleoprotein telomerase, which undergoes progressive reduction in expression in mammalian somatic cells until the telomeres reach the critical length, or the shortest length required to prevent chromosomal instability (Liu *et al.*, 2019). This process, known as 'replicative senescence' has evolved in certain longer-lived placental mammals and is thought to prevent the uncontrolled cell division associated with cancerous phenotypes (Seluanov *et al.*, 2007). The role of telomeres in regulating cell division has led to the examination of the role of telomeres in the placenta, which undergoes rapid cell differentiation during its development, with cell differentiation continuing in the placenta until birth. Progressive shortening of telomeres in cells of the placenta has led to the hypotheses that altering telomere length may act as a cellular 'count-down' as the placenta ages, initiating the process of birth (Phillippe, Sawyer and Edelson, 2019). This hypothesis is supported by comparison of term placenta telomere lengths to first trimester placental villi, where telomere length is increased in first trimester samples (Novakovic *et al.*, 2016). Shorter telomere length was found in

pregnancies complicated with maternal diabetes and intra-uterine growth restriction, indicating that altered telomerase expression may result in differences in cell proliferation and health, resulting in placental disorders (Jirkovska, Korabecna and Lassakova, 2019).

The enrichment for telomere related functions is reflected in the results of an analysis of enrichment for gene-gene interactions in STRINGv.11 (Szklarczyk *et al.*, 2019). Three functionally related genes; WD Repeat Containing Antisense to TP53 (WRAP53), telomerase reverse transcriptase (TERT), and ACD shelterin complex subunit and telomerase recruitment factor (ACD), are all involved in telomerase function, with WRAP53 interacting with TERT, and ACD acting as part of the telosome/shelterin telomeric complex to protect the telomere ends (Xin, Liu and Songyang, 2008). The WRAP53 protein has dual function and also serves as an antisense transcript of tumor protein p53, positively regulating p53 expression by targeting the 5'UTR of p53 (Mahmoudi *et al.*, 2009). P53 expression has been linked to the expression of human chorionic gonadotropin (hCG), the human maternal pregnancy recognition signal, by upregulating the hCGbeta7 (CGB7) gene, allowing the human endometrium to become receptive to an implanting blastocyst (Sohr and Engeland, 2011). The interactions of NSE1 Homolog, SMC5-SMC6 Complex Component (NSMCE1), NSE3 Homolog, SMC5-SMC6 Complex Component (NDNL2), NSE2 (MMS21) Homolog, SMC5-SMC6 Complex SUMO Ligase (NSMCE2) and Structural maintenance of chromosomes protein 6 (SMC6) provide further evidence of adaptive evolution acting on genes involved in chromosomal maintenance and DNA repair on the stem Eutherian lineage, mediated in this case by the SMC5/6 complex (Fernandez-Capetillo, 2016).

Other high confidence interactions identified in our STRING (Szklarczyk *et al.*, 2019) analysis include Transmembrane protein 186 (TMEM186) and its interaction with ECSIT Signalling Interactor (ECSIT). ECSIT a signalling protein implicated in the Toll signalling pathway, has been implicated in ovarian folliculogenesis, displaying 5 fold upregulation in ovaries afflicted with atresia, a process by which follicles degenerate and fail to ovulate (Terenina *et al.*, 2016). Furthermore, ECSIT has been found to be required for healthy cell differentiation and embryo development during murine early pregnancy, with ECSIT knockout mice experiencing embryonic fatality by day 7.5 (Xiao *et al.*, 2003).

Other gene-gene interactions that are enriched include a pathway implicated in the inflammatory immune response, whereby Coiled-Coil Domain Containing 22 (CCDC22) and COMM Domain Containing 5 (COMMD5) interact to activate the Nuclear Factor kappa-light-chain-enhancer of activated B cells (NF-κB) proinflammatory signalling pathway. This interaction has been shown to be necessary for the activation of the NF-κB pathway, with decreased CCDC22 expression preventing adequate activation of the NF-κB pathway (Starokadomskyy *et al.*, 2013). The activation of the NF-κB proinflammatory signalling pathway by the interaction of CCDC22 and COMMD proteins in turn activates Cullin 4B (CUL4B) formation of the Cullin4B-RING E3 ligase complex (CRL4B), which in turns binds DCAF11, the conserved positively selected gene implicated in STRING annotated interactions, to form $CRL4B^{DCAF11}$ E3 ligase, which positively regulates cell proliferation and cell cycle progression (Chen *et al.*, 2017). The interaction of these three proteins has been studied in metastatic cancer cells, and the link between cancer and placentation phenotypes has been established, due to the rapid cell proliferation that is common to both states (Costanzo *et al.*, 2018). This link to cancer continues through the interaction of PSStemFixed genes Tumor Necrosis Factor Receptor Type 1-Associated DEATH Domain Protein (TRADD), Receptor (TNFRSF)-Interacting Serine-Threonine Kinase 1 (RIPK1) and Toll Like Receptor Adaptor Molecule 1 (TICAM1). Both TRADD and RIPK1 are receptors associated with tumor necrosis factor receptor 1 (TNFR1), with TRADD belonging to the family of 'death' receptors that induce apoptosis in association with TNFR1. However, these 'death' receptors can also induce cell proliferation, along with the NF-κB proinflammatory signalling pathway. TRADD is also important in the recruitment of RIP kinases to assist in the activation of the NF-κB proinflammatory signalling pathway, among them being RIPK1 (Pobezinskaya and Liu, 2012). Like TRADD, RIPK1 is also involved in apoptosis and cell death in association with TNFR1 (Degterev, Ofengeim and Yuan, 2019). RIP kinases are in turn involved in the recruitment of TICAM-1 to regulate the activity of Toll-like receptor 3, which can induce cell death via necrosis or apoptosis; or mediate cell survival by the production of Type I Interferons and the localisation of proinflammatory cytokines (Seya *et al.*, 2012).

The gene-gene interactions discovered between 115 PSStemFixed genes show adaptive evolution acting upon groups of genes involved in cell cycle, cell maturation, chromosome maintenance, DNA repair, inflammatory responses, cancer and programmed cell death. While these functions are not explicitly related to placentation and reproduction, the role of cell proliferation, inflammation and regulation of the cell cycle in the formation of the placenta have been documented (Hauguel-de Mouzon and Guerre-Millo, 2006; Woods, Perez-Garcia and Hemberger, 2018; Gal *et al.*, 2019).

2.6 Conclusion

Using high quality vertebrate genomes, along with alignment comparison methods and a resolved phylogeny, 237 gene families were found to have undergone positive selection on the stem Eutherian lineage, with the substitution being unaltered once it entered the Eutherian lineage. Of these gene families, 115 genes (PSStemFixed) were found to be conserved in extant eutherians.

These genes were found to be enriched for functions related to telomere activity, and showed statistically significant enrichment for gene-gene interactions. The functions of these gene-gene interactions are implicated in pathways related to cell cycle regulation, cell death, proliferation and immune functions, suggesting that genes from multiple pathways underwent positive selection for during the evolution of mammalian placenta, underlining the complex innovations required for the emergence of a novel organ.

Chapter 3: A comparative genomics approach to identify therian specific microRNAs

3.1 Aims of this chapter

1. Using database sources, compile a set of miRNAs which emerged on the stem mammal lineage.
2. Place these miRNAs onto the mammal phylogeny using a gain-loss analysis
3. Predict the targets of miRNAs that emerged at significant points in evolution to determine their biological roles

3.2 Introduction

Micro RNAs (miRNAs), are short non-coding regulatory elements, that control gene expression post-transcriptionally. They predominantly work via downregulating gene expression by binding to a seed sequence in the 3'UTR region of the target and degrading target mRNAs, or obstructing translation (Yuan *et al.*, 2010); and upregulating gene expression by binding to a promotor region in the target 3'UTR (Zhang *et al.*, 2018) or by engaging in competitive binding with a repressor miRNA (Vasudevan, 2012). Since their discovery, miRNAs have been of interest in the study of morphological evolution, due to the observation of significant expansions in miRNA populations at times of morphological innovation (Hertel *et al.*, 2006). By comparing experimentally verified miRNA sequences to representatives of *Bilateria, Porifera* and *Cnidaria*, 30 miRNAs were found to have emerged in the bilaterian ancestor and remained fixed in extant *Bilateria* (Prochnik, Rokhsar and Aboobaker, 2007). This method of searching for homologous miRNA sequences in genomes of interest was used to discover 41 conserved miRNAs gained in the vertebrate clade but not present in any invertebrate (Heimberg *et al.*, 2008).

Expansions in miRNA repertoire concurrent with increased complexity and/or major transitions in animal evolution have led to investigations of the role of miRNAs in the development of novel

tissues and organs. Using Northern Blotting and *in-vitro* mutagenesis two *C.elegans* miRNAs; *lin-4* and *let-7* were identified as temporally expressed regulators during *C.elegans* development, with *lin-4* repression of the *lin-14* locus enabling the transition from the first to the second larval stage, while *let-7* downregulation of *lin-41* facilitates the transition from late larval to early adult cell fates during *C. elegans* development (Lee, Feinbaum and Ambros, 1993; Pasquinelli *et al.*, 2000). Temporally expressed miRNAs were also implicated in mouse lung development, using microarray hybridization analyses and distinct time-points of mouse development, from Day 12 of gestation into adulthood, finding 117 miRNAs with differential expression at different stages of lung development. Using target prediction software and mRNA microarray hybridization, target transcripts and/or proteins of these temporally expressed miRNAs were found to be correspondingly downregulated, suggesting that miRNAs may drive dynamic genetic alterations during organogenesis (Dong *et al.*, 2010).

miRNAs are known to play a role in healthy placental development and successful pregnancy. For example, miR-675 was found to be exclusively expressed in the placenta, and using RNAi knockdowns and expression profiling was found to have increased expression during placental development (Keniry *et al.*, 2012). By knocking down miR-675 in mouse models, miR-675 was found to negatively regulate proliferation during placental growth, with miR-675 knockout mice exhibiting placental overgrowth (Keniry *et al.*, 2012). miRNAs are also implicated as regulators of trophoblast cells, with temporal expression of members of the miR-17~92 cluster observed during trophoblast cell differentiation while miR-17~92 miRNAs found to be downregulated during syncytiotrophoblast differentiation, with overexpression of members of this cluster associated with reduced differentiation (Kumar *et al.*, 2013). miRNAs may also enhance trophoblast cell differentiation, for example, by the action of miR-378a-5p on the Nodal signalling pathway, downregulating Nodal and thereby preventing its negative action on trophoblast differentiation and migration during early pregnancy, along with preventing Nodal-mediated placental apoptosis during early gestation (Luo *et al.*, 2012).

While miRNAs may promote or facilitate healthy placental development and maintenance, aberration in miRNA expressions may contribute to diseases of the placenta. miRNAs have been

found to be differentially expressed in preeclampsia, with miR-210, miR-1233 and miR-574 elevated in serum samples from preeclampsia patients (Munaut *et al.*, 2016). Multiple miRNAs affect trophoblast invasion, including miR-16, miR29b, miR-34aa, miR-155 and miR-210, leading to issues with spiral artery remodelling (Bounds *et al.*, 2017). Inhibition of syncytiotrophoblast differentiation is a direct effect of upregulation of members of the miR-17~92 cluster, with said cluster upregulated in placentas affected by preeclampsia (Kumar *et al.*, 2013). miRNAs are also implicated in recurrent pregnancy loss (RPL), with overexpression of miR-133a in recurrent miscarriages. This miRNA affects histocompatibility antigen (HLA-G) levels, reducing maternal immune tolerance to the developing allograft-like fetus (Santamaria and Taylor, 2014).

Using homology-based searches and a set of carefully chosen candidate genome sequences, our aim is to identify miRNA cohorts that emerged at specific points during mammal evolution.

3.3 Materials and Methods

3.3.1 Dataset Construction

3.3.1.1 Species sampling

Ensembl 92 (Zerbino *et al.*, 2017) was mined for candidate genomes from each vertebrate clade, with representatives from *Fish, Amphibia, Reptilia, Aves, Monotremata, Metatheria and Eutheria*. Within each vertebrate clade, representative species were chosen from candidates based on genome coverage, when compared to other members of their clade. Within *Eutheria*, candidate species were assessed based on genome coverage, in order to include as much genomic information in a homology search, and placental morphology (Chapter 2, Table 2.1), in order to include representatives of each placental type. Selected species, along with their identifier and coverage are detailed in Table 3.1. Species selected were the best quality genomes available within their taxonomic group, however, when comparing between clades (i.e. *Laurasiatheria* vs *Primates*), genome quality varied. This variation in genome quality and coverage may impact the ability to detect miRNAs if sections of the genome, such as introns, are incomplete.

Table 3.1 Vertebrate Species sampled: the genome version used and the coverage and completion level.

Clade	Species	Version	Genome Quality
Fish	Zebrafish	GRCz11	Full Genome, Chromosome Level
	Stickleback	BROAD S1	11x Coverage, Chromosome Level
	Fugu	FUGU 4.0	8.5x Coverage, Scaffold Level
	Coelacanth	LatCha1	77.5x Coverage, Scaffold Level
Amphibians	Xenopus	JGI 4.2	7.65X Coverage, Scaffold Level
Reptiles	Anole Lizard	AnoCar2.0	7.10X Coverage, Chromosome Level
Birds	Zebra Finch	taeGut3.2.4	6.X Coverage, Chromosome Level
	Chicken	Gallus_gallus-5.0	70X Coverage, Chromosome Level
Monotremes	Platypus	OANA5	6X Coverage, Chromosome Level
Metatheria	Opossum	monDom5	7.33X Coverage, Chromosome Level
Eutheria	Elephant	Loxafr3.0	7X Coverage, Scaffold Level
	Armadillo	Dasnov3.0	6X Coverage, Scaffold Level
	Mouse	GRCm38.p6	High Quality Reference Assembly
	Human	GRCh38.p12	High Quality Reference Assembly
	Mouse Lemur	Mmur_3.0	221.6X Coverage, Chromosome Level
	Bushbaby	OtoGar3	137X Coverage, Scaffold Level
	Cat	Felis_catus_9.0	72X Coverage, Chromosome Level
	Microbat	Myoluc2.0	7X Coverage, Scaffold Level
	Horse	Equ Cab 2	6.79X Coverage, Chromosome Level
	Cow	UMD3.1	9X Coverage, Chromosome Level

Using Ensembl 92 (Yates *et al.*, 2016)*, a dataset of genomes representative of (i) vertebrate outgroup clades, or (ii) variations in placental morphology in metatherian and eutherian mammals. For each clade, genomes were chosen based on genome coverage for downstream homology searching. 'Clade' refers to the taxonomic group of each included species. 'Species' denotes the included species, by their common name. 'Version' denotes the genome assembly version included in this analysis. 'Genome Quality' refers to the coverage and assembly level of each species included in this analysis.*

3.3.1.2 Constructing the subject database of regions that potentially code for miRNAs

miRNAs have been found to be encoded throughout the genome, including intergenic and intronic regions (Olena and Patton, 2010). In order to place miRNAs of interest onto the Therian phylogeny (containing *Metatheria* and *Eutheria*), the 5'UTRs, 3'UTRs, exonic and intronic regions for each species listed in Table 3.1 were accessed from Ensembl92 (accessed April 2018) (Yates *et al.*, 2016) and constructed into subject datasets, based on genomic region. The properties of each dataset detailed in Table 3.2. In summary, the 'Introns' dataset contained the highest total number of sequences (6472686), the longest sequence (4693080) and the longest average sequence length (4108.0 base pairs) when compared to other genomic regions, while the '3'UTR' dataset contained the smallest total number of sequences (311524) and the '5'UTR dataset' had the shortest average sequence length at 232.1 base pairs (Table 3.2).

Table 3.2 Properties of each subject database used.

Dataset	Total Number of Sequences	Longest Sequence (base pairs)	Average Sequence Length (nt)
3'UTR	311524	122583	1021.9
5'UTR	337821	20453	232.1
Introns	6472686	4693080	4108.0
Exons	5314509	205012	243.3

For each of the selected representative Vertebrate genomes, the 5'UTRs, 3'UTRs, introns and exons were accessed from Ensembl92 (Yates *et al.*, 2016). *These regions were constructed into subject datasets for homology searching in BLASTn2.6.0* (Altschul *et al.*, 1990)

3.3.1.3 miRNA Sequence Acquisition

All miRNAs included in this analysis were predicted to have emerged on the therian and eutherian ancestral lineages. This resulted in 112 miRNA candidate gene families (Fromm *et al.*, 2018). For each miRNA, the stem-loop sequence (~100nt) was downloaded from MirGeneDB, the curated miRNA database (accessed 15th April 2018). Searching for sequence homology at the nucleotide using short sequences (such as the 18-22nt mature miRNA sequence) can be problematic. As there are only 4 nucleotides, the probability of finding homology in short sequences due to random chance is high, producing a higher number of false positives (Pearson, 2013). Therefore,

the stem loop sequence (i.e. the long pre-miRNA sequence) was used, rather than the mature miRNA sequences for the homology searches performed.

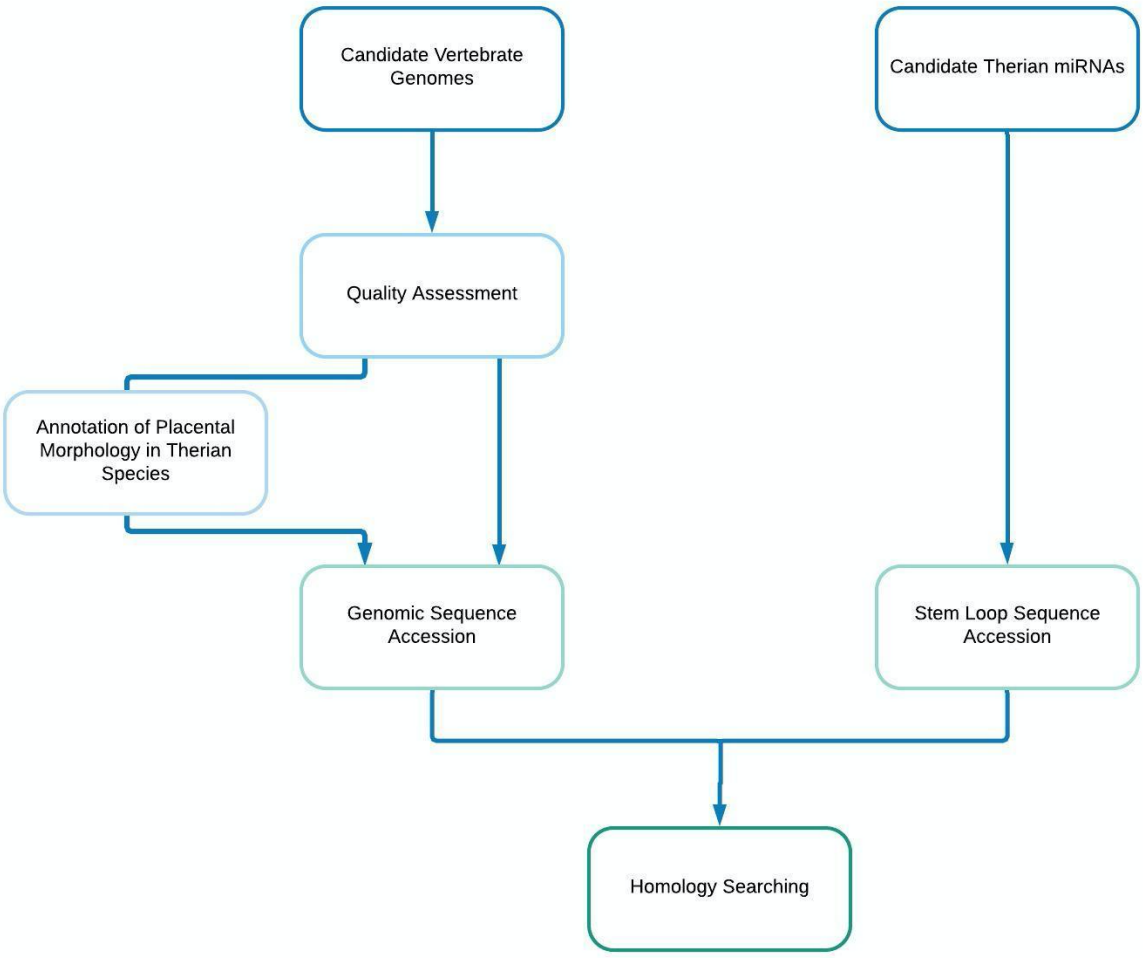


Figure 3.1 Flow diagram showing the stages of assembling the high quality vertebrate genomic regions and the stem loop miRNA regions. *Candidate vertebrate genomes representing (i) major vertebrate outgroup clades or (ii) variations in placental morphology in metatherian mammals were accessed from Ensembl92* (Yates *et al.*, 2016)*. Genomes were selected based on genome coverage in order to have sufficient genomic information for downstream homology searching. 5'UTRs, 3'UTRs, introns and exons were accessed from Ensembl92 for each representative vertebrate genome and subject databases were created for BLASTn analysis. Stem loop miRNA sequences were accessed from miRGeneDB2.0* (Fromm *et al.*, 2018)*, with the longer stem loop sequence used preferentially as the shorter mature miRNA sequence may result in false positive results from homology searching* (Pearson, 2013)*. Homology searches were performed using BLASTn 2.6.0* (Altschul *et al.*, 1990)*.*

3.3.2 Searching for miRNA homologs in *Vertebrata*

Homology searches were performed in BLASTn (version 2.6.0+) (Altschul *et al.*, 1990). The recommended minimum sequence length for BLASTn searches is 30nt (NCBI, 2017). The average length of the pri-miRNA sequences used in this analysis is 60.4nt, with the shortest sequence

being 52nt, thereby satisfying the criteria for length in an attempt to bypass the issues of false positive homology caused by short query sequences, described above.

Each miRNA stem loop query sequence was compared to each genomic region to determine (i) if candidate miRNA homologues existed in non-therian species and (ii) if candidate miRNAs showed homology to genomic regions from therian species. Results were compared to the smallRNA sequencing information from miRGeneDB (Fromm *et al.*, 2018) for candidate miRNAs to determine the efficacy of our method (Fromm *et al.*, 2018). We applied two e-values (e^{-7} and e^{-10}) for homology searches, whereas the default BLAST e-value is e^{-10}, the relaxed e-value of e^{-7} was also used to account for possible false negatives due to possible annotation issues in poorly sampled clades (e.g. *Marsupialia* and *Laurasiatheria*). This was to accommodate varying quality of genomes, sequence changes over the vertebrate clade (when comparing human, mouse or cow sequences to non-mammalian vertebrates), and the length of query sequences. Results of the two sequence similarity searches using these e-values were compared to find an optimal e-value for miRNA annotation across clades.

3.3.3 Establishing the phylogenetic distribution of candidate miRNAs

The phylogenetic distribution of miRNA gene families originating within extant therians and eutherians was assessed from the patterns of presence and absence on the current mammal species phylogeny (Morgan *et al.*, 2013; Tarver *et al.*, 2016). Using the parsimony-based approach implemented in TNT (Giribet, 2005) the precise phylogenetic placement of each miRNA was estimated on the *Therian* phylogeny. TNT was chosen due to (i) the resolved phylogeny used in this analysis and (ii) the moderate number of candidate miRNA sequences (112).

3.3.4 Investigation of miRNA targets

TargetScan (Agarwal *et al.*, 2015) was used to predict human targets for conserved miRNAs that emerged at the base of the *Therian* and *Eutherian* lineages (hereafter referred to as stem lineage miRNAs) in the human genome. Previous benchmarking, as discussed in Chapter 1, (Walsh, 2013; Fan and Kurgan, 2015) showed that, while there is no optimal miRNA target prediction software, TargetScan (Agarwal *et al.*, 2015) is among the highest quality programs available. It was

developed using benchmark datasets from human, mouse and *Drosophila*, and searches introns, exons and 5'UTRs along with 3'UTRs for miRNA binding sites (Lewis, Burge and Bartel, 2005b). Benchmarking studies of miRNA target prediction software showed TargetScan outperforming Pita and RNA22 in sensitivity and outperforming miRanda-mirSVR and Pita in specificity (Oliveira *et al.*, 2017), providing an excellent compromise between the two metrics.

Therefore, the TargetScan output was filtered by level of complementary binding between the miRNA seed region and target site. Targets with ≥7mer complementary binding to the seed region were selected, with this level of miRNA-target binding indicative of an efficient binding site (Bartel, 2009). Of these filtered targets, levels of co-operative and antagonistic binding involved in stem lineage miRNA regulation were analysed for transcripts with >1 predicted regulatory miRNA, using binding site co-ordinates predicted in TargetScan. Co-operative binding was identified using previously observed optimal seed site differences for miRNA co-operative binding (Sætrom *et al.*, 2007), with stem lineage miRNA seed sites within 13-35nt of each other annotated as putative co-operative sites. miRNAs with overlapping binding sites on the same transcript were annotated as involved in possible competitive or antagonistic binding (sites within 0-7nt of each other) (Yu *et al.*, 2008; Koscianska *et al.*, 2015). Analysis of the functional enrichment of stem lineage miRNA targets was performed in PANTHER v.14 (Muruganujan *et al.*, 2018).

3.3.5 Statistical overrepresentation of predicted targets of conserved miRNAs

The filtered predicted gene targets of stem lineage miRNAs were analysed for overrepresented pathways using the Reactome Pathway Database (Jassal *et al.*, 2019). Predicted miRNA targets (15,958 genes) were compared to the full Reactome Pathway annotation of the human genome, containing 10,867 protein-coding gene annotations, 24,849 alternate forms of proteins due to post-translational modifications and 1,803 pathways, along with 1,856 annotations of substrates, and regulatory molecules, to identify statistically overrepresented pathways, with statistical results corrected for false-discovery using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995) by the Reactome Pathway program itself.

3.3.6 Binding dynamics analysis of positively selected miRNA targets

As discussed in Chapter 2, 237 gene families were found to have signatures of positive selection on the stem eutherian lineage. 115 of these families were present in all extant mammals, with the amino acid substitution that occurred on the stem eutherian lineage being fixed in extant species. In order to test the hypothesis that genes under positive selection on the stem eutherian lineage are more likely to be regulated by the stem lineage miRNAs, these 115 gene families were analysed for stem lineage miRNA binding sites. Of these gene families, 84 were predicted targets of the stem lineage miRNAs. The number of computationally predicted miRNA binding sites was determined for the human orthologue of the 84 positively selected genes for the 14 miRNAs that emerged on the stem therian and eutherian lineages. The mean number of stem lineage miRNA binding sites on these 84 gene families were then compared with the mean number of predicted stem lineage binding sites for all other human genes that did not undergo positive selection on the stem eutherian lineage. Differences in stem lineage binding sites per transcript were considered statistically significant when $p \leq 0.05$, using a two-sample t-test with unequal variance, determined using Levene's test of variance of populations (Levene, 1960).

3.4 Results

3.4.1 E-Value Assessment

To ensure that the homology searching method (BLASTn) was sufficiently sensitive to detect homology in all of the sampled species, BLAST output files were analysed to determine whether a stringent (e^{-10}), or relaxed (e^{-7}) e-value was optimal for miRNA annotation in 20 vertebrate species. The default BLAST e-value (e^{-10}) retrieved all but one hit. This BLAST hit (Human miR-362 pri-miRNA sequence against Armadillo intronic sequence) had an e-value of $3.66e^{-09}$. This instance was a false negative, as miR-362 has been annotated in Armadillo through small-RNA sequencing (Tarver *et al.*, 2016). Therefore, an e-value of e^{-7} was found to be more suitable for cross-species miRNA annotation as the default of e^{-10} failed to retrieve known sequences. Using an e-value of e^{-7}, the miRNA annotations of miRGeneDB were successfully replicated, enabling downstream analysis of phylogenetic placement of the 112 therian miRNAs.

3.4.2 Six miRNAs emerged on the Therian stem lineage and are present in extant Eutherians

Of the 112 assessed miRNAs annotated as originating on the therian and eutherian stem lineage (Fromm *et al.*, 2018), a heterogeneous distribution was found across eutherian mammals, with multiple gains and losses evident (detailed in Figure 2A). However, there are 6 miRNA families that emerged on the Therian stem lineage; miR-340, miR-483, miR-671, miR-675, miR-1251, and miR-3613. Of the 6 miRNAs of therian origin, miR-340 remained fixed throughout the therian and eutherian species sampled, whereas miR-671, is present in all sampled species with the exception of *Monodelphus domestica*. Considering miR-671 is present in *Sarcophilus harrisii,* which diverged after *M. domestica* during marsupial evolution, and all representative high-quality Eutherian species sampled, the lack of miR-671 annotation in *M. domestica* may be due to issues with annotation and sequencing quality. Taking this into account, miR-671 was included in downstream analyses as a 'stem lineage miRNA'.

3.4.3 106 miRNAs emerged on the stem Eutherian lineage

A total of 106 miRNAs emerged in *Eutheria* and 12 of these remained conserved in extant *Eutheria*. A total of 29/106 miRNAs had various patterns of lineage- or species-specific loss (detailed in Figure 2B), such as in *Laurasiatheria (*e.g. *Bos taurus, Equus caballus* and *Myotis lucifugus*) (miR-760) or in *Rodentia* (e.g. *Mus musculus*) (miR-450) and a further 66/106 miRNA families experienced lineage- or species-specific gain and loss patterns. For example, the *Euarchontoglires clade* (e.g. *Mus musculus* and *Homo sapiens*) gained 2 miRNAs (miR-134 and miR-1247) whilst miR-1842 was present only in *Ungulata* (e.g. *Bos Taurus* and *Equus caballus*).

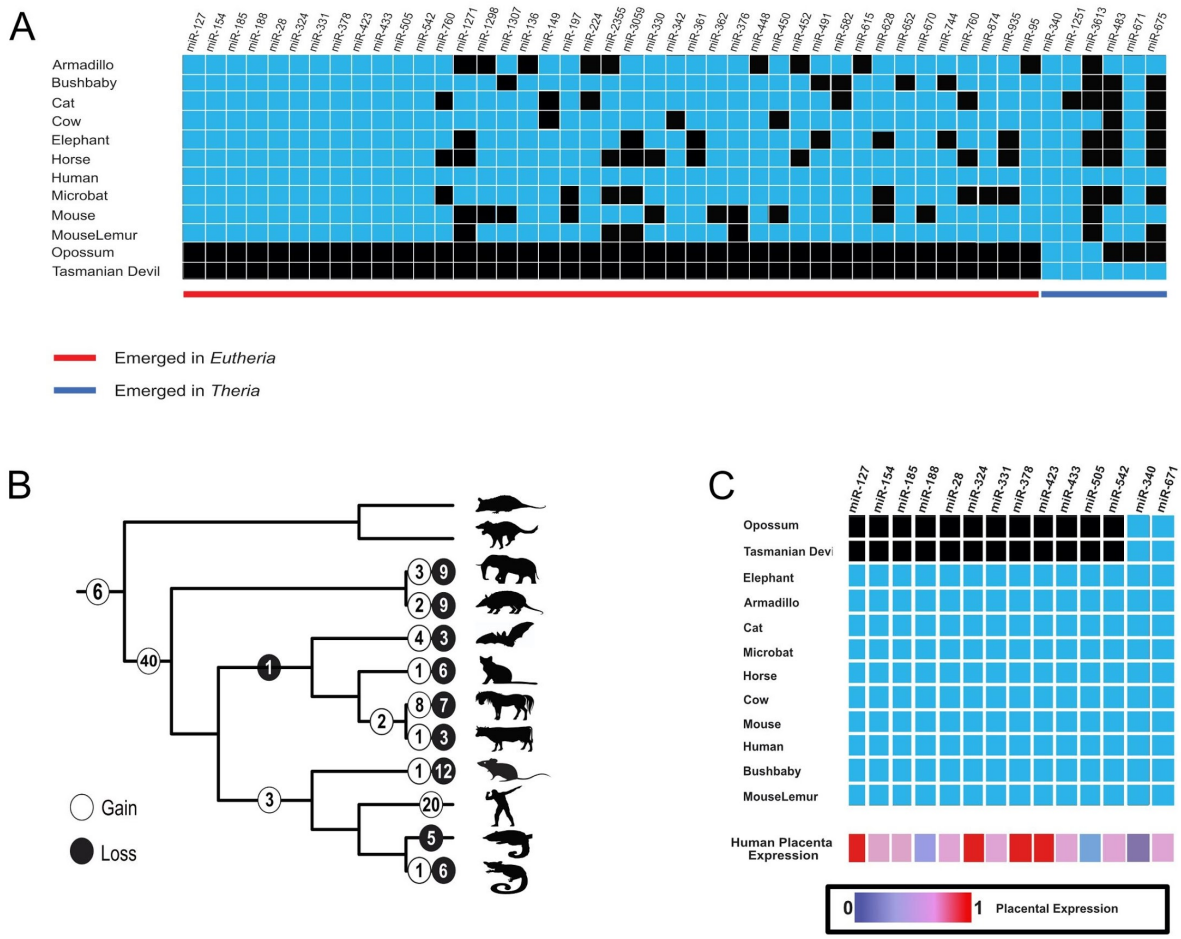


Figure 3.2 Emergence of miRNAs specific to therian and eutherian mammals. A: *Representation of patterns of novel miRNA gain (blue) and loss (black) across the Therian and Eutherian clades;* ***B****: Species and lineage specific miRNA gains and loss depicted on the Therian phylogeny, with species used being representative of genus;* ***C:*** *Representation of the emergence of stem lineage miRNAs that emerged at the root of therian mammals and eutherian mammals. Expression of each miRNA in the human placenta is shown below, ranging from highly expressed (red) to low expression (blue)*

3.4.4 Two miRNAs emerged on the therian stem lineage, followed by 12 on the eutherian stem lineage, that are conserved in all extant mammals

Two miRNA families emerged on the Therian stem lineage and may be conserved in all extant Therian mammals. A further 12 miRNA families emerged on the Eutherian stem lineage and are conserved in all extant Eutherian mammals. In addition, miRGeneDB annotations of miRNAomes from publicly available expression datasets (Fromm *et al.*, 2018), indicate that these miRNAs are expressed in multiple tissues, including placental tissue (expression levels indicated in Figure 3.2c). The origin and evolution of these 14 miRNAs is coincident with the origin of mammals and a variety of mammal specific traits and they may be implicated in alterations in gene regulation that facilitated the evolution of mammal placenta.

3.4.5 Target prediction and regulatory dynamics of stem lineage miRNAs

TargetScan (Agarwal *et al.*, 2015) output was mined for predictions of transcripts regulated by the 14 stem lineage miRNAs. Of 29,264 annotated 3'UTRs in the TargetScan database, stem lineage miRNAs were predicted to target 22,911 (when filtered for binding sites with ≥7 seed sequence match). However, the levels of regulation for each miRNA varied, with miR-127 predicted to target only 520 transcripts, in contrast to miR-185, which had annotated binding sites in 20,754 transcripts (Figure 3.3).

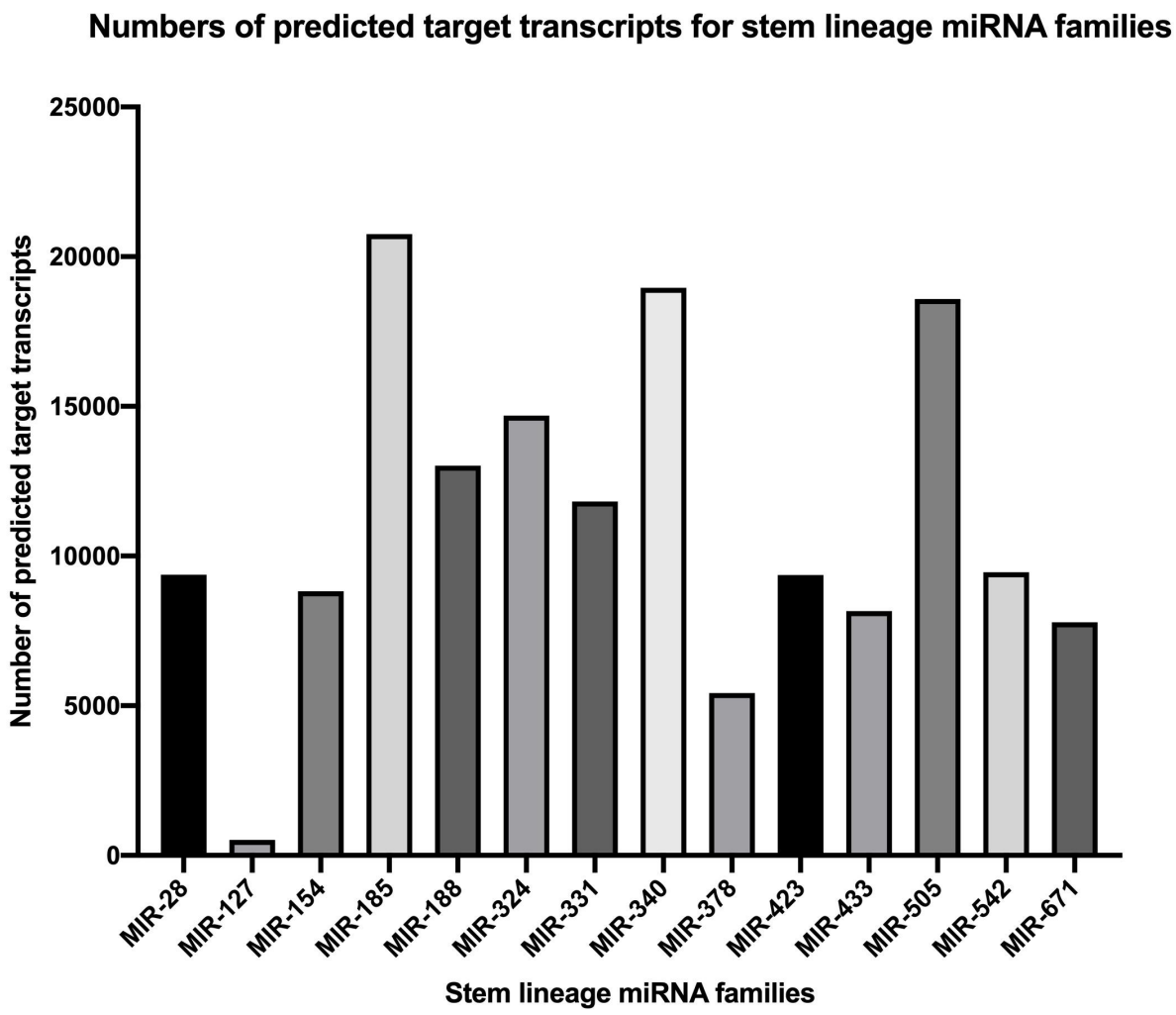


Figure 3.3 Number of predicted human targets for each stem lineage miRNA family, filtered for ≥7-mer seed sequence matches*. Targets were predicted for each of the 14 stem lineage miRNAs using TargetScan70* (Agarwal *et al.*, 2015)*. TargetScan output was filtered for targets with ≥7-mer seed sequence matches, as seed sequence pairing of ≥7-mer is indicative of efficient binding* (Bartel, 2009)*.*

Co-operative and antagonistic binding sites were identified for each predicted transcript (Table 3.3). Co-operative binding is defined as miRNAs with binding sites within 13-35nt of each other working together to increase their regulatory effect. Antagonistic binding is defined as miRNAs with overlapping binding sites competing against each other to bind a target transcript (Sætrom *et al.*, 2007; Yu *et al.*, 2008; Koscianska *et al.*, 2015). A number of predicted target transcripts (898) showed signatures of both cooperative and antagonistic binding patterns along their sequence, indicating dynamic regulation involving multiple miRNAs. Putative cooperative binding sites were found on 2111 transcripts, while 999 transcripts had signs of antagonistic binding. 17029 transcripts showed no indication of either cooperative or antagonistic binding.

Table 3.3 Binding dynamics of predicted targets of stem lineage miRNAs.

Type of binding	Number of transcripts
Co-operative	2111
Antagonistic	999
Co-operative and antagonistic	898
No result	17029

Using filtered TargetScan70 (Agarwal *et al.*, 2015) *output, predicted binding site regions for each stem lineage miRNA were analysed to find signatures of 'Cooperative' binding (where stem lineage miRNA binding sites are located within 13-35nt of each other* (Sætrom *et al.*, 2007) *or 'Antagonistic' binding (where stem lineage miRNA binding sites overlap* (Yu *et al.*, 2008; Koscianska *et al.*, 2015)*). 'No result" refers to predicted targets without signatures of co-operative or antagonistic binding.*

3.4.6 miRNA Target prediction – investigation of functional annotation

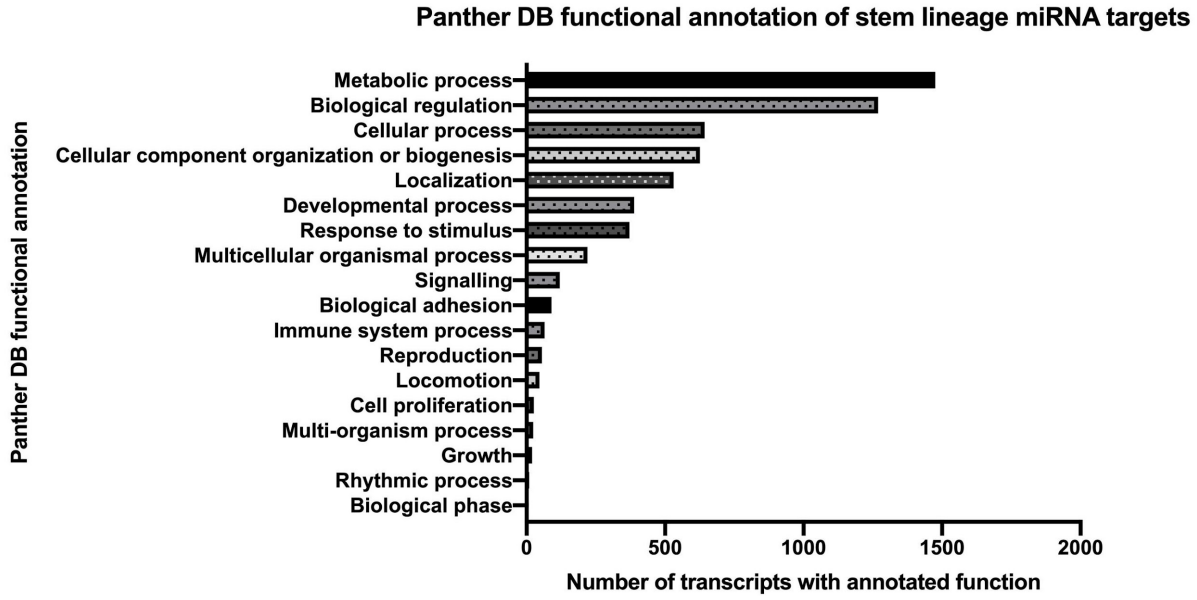


Figure3.4 Functional annotation of predicted targets of the 14 stem lineage miRNAs. *Filtered stem lineage miRNA target transcripts were analysed for functional enrichment using PANTHERv.14* (Muruganujan *et al.*, 2018)*, where PANTHERv.14 found functional annotations to be significantly enriched when p≤0.05.*

While only 55 stem lineage miRNA target transcripts were found to be implicated in reproductive functions, the highly represented functions - Metabolic process (1476 transcripts), Biological regulation (1269 transcripts), etc) are integral to cellular health and biological processes.

3.4.7 Stem lineage miRNAs preferentially target genes that underwent positive selection on the stem Eutherian lineage

The 84 positively selected gene families were found to be significantly enriched for binding sites for the conserved therian miRNAs ($p=9.27737^{e-12}$), with a mean of 6.77 (Figure 3.5) and a median of 4.0 binding sites per transcript, when compared to genes that did not undergo positive selection on the stem eutherian lineage (mean=1.70, median=1.0). From this data, it is clear that the stem lineage miRNAs preferentially target the positively selected gene families.

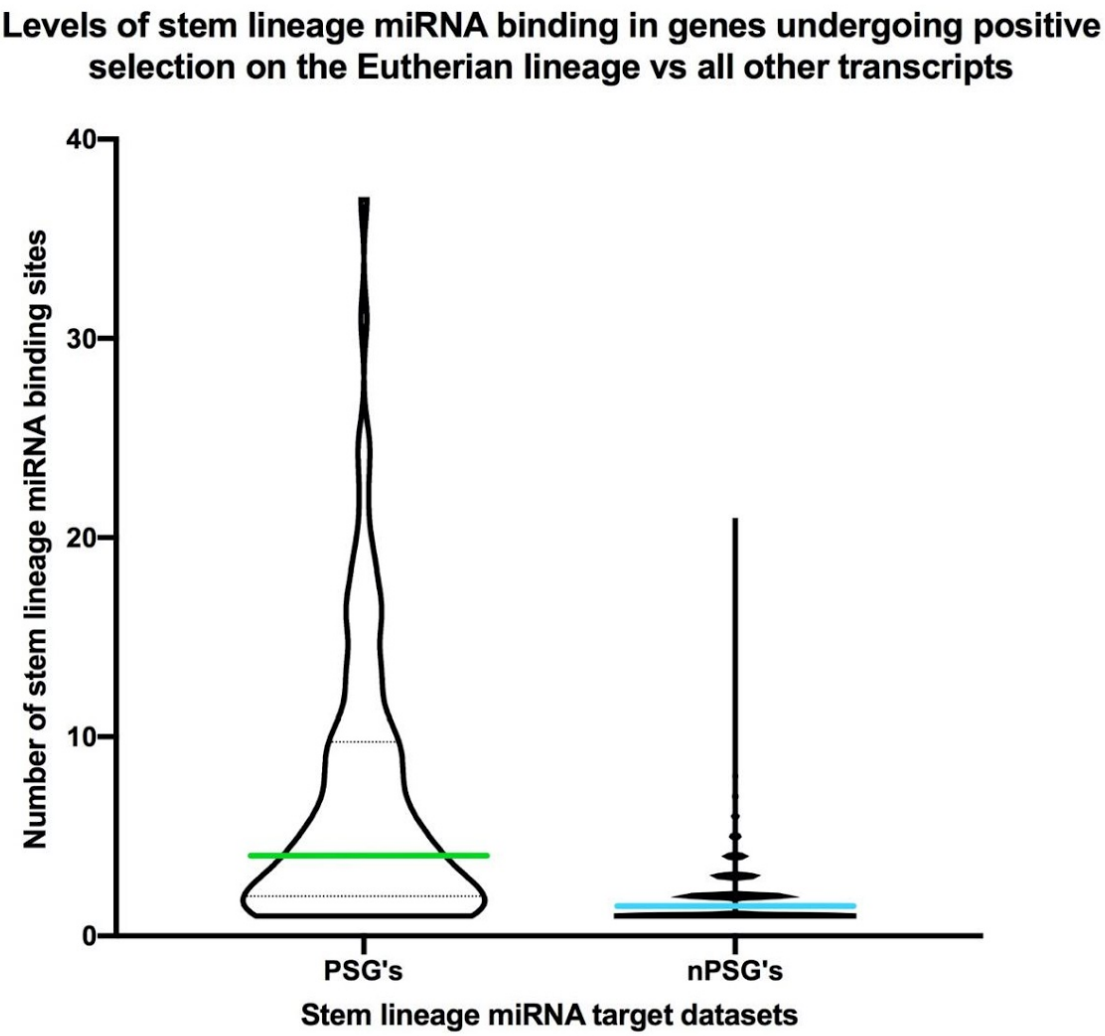


Figure 3.5 Violin plot showing the number of stem lineage binding sites per transcript in genes that underwent positive selection on the stem Eutherian lineage (PSGs) compared to genes that did not undergo positive selection on the stem Eutherian lineage (nPSG)*. For each of the 84 gene families that underwent positive selection on the stem eutherian lineage, the mean number of stem lineage miRNA binding sites (depicted in green) was determined for each transcript of the human gene orthologue. This was compared to the mean number of binding sites for each predicted human target transcript that did not undergo positive selection on the stem eutherian lineage (depicted in blue). The mean number of binding sites was determined to be significantly different between the two datasets when p≤0.05, determined using a two-sample t-test with unequal variance.*

3.5 Discussion

Through the use of the miRGeneDB (Fromm *et al.*, 2018) curated database of miRNA annotations and stem loop sequences 112 miRNAs annotated in miRGeneDB were placed onto the agreed mammal phylogeny. Arguably, the curated nature of miRGeneDB greatly improved the process of selection of candidate miRNAs. As it was created using publicly available small-RNA sequencing datasets to annotate miRNAs and uses a strict set of criteria to identify miRNAs (discussed in Chapter 1) and miRNA gene families, the available data was high quality. This improvement in miRNA annotation and nomenclature removes much of the issues associated with databases that allow user submitted miRNAs, which has resulted in incorrectly identified miRNAs, or members of the same miRNA family being annotated as separate miRNAs (Fromm *et al.*, 2015).

The quality of the stem loop miRNA sequences, along with the availability of small-RNA sequencing information from miRGeneDB allowed the comparison of the results of this study to miRGeneDB annotations, in order to examine the accuracy of this analysis. This information proved valuable in the assessment of the e-value used in homology searching. The default e-value used in BLASTn (Altschul *et al.*, 1990) is e^{-10}, which identified sequence homology between the vertebrate genomic regions used in this study and 111 of the 112 miRNAs examined. However, as stated, the e-value was too stringent to identify the Armadillo homologue of miR-362 (e-value $3.66e^{-9}$), which had been annotated using small-RNA sequencing (Tarver *et al.*, 2016). This result indicates that a more relaxed e-value (e.g. e^{-7}) may be more suitable when identifying miRNA homologues across species, possibly due to either the short sequence (~100nt) or issues with genome assembly in poorly sampled clades. When using an e-value of e^{-7}, the only difference in identified miRNAs was the Armadillo homologue of miR-362. All other results were identical to those obtained using an e-value of e^{-10}. If similar approaches are taken to identify miRNAs in different species or clades, it would be beneficial to compare the results of homology searches to published datasets, where applicable, in an attempt to mitigate false positive or negative results.

The available sequencing data for these miRNAs and the strict annotation required for inclusion in the miRGeneDB database allowed confidence in the data when performing downstream

analyses. The quality of annotation, along with the observed lack of homoplasy in miRNAs (Sempere *et al.*, 2006; Tarver *et al.*, 2018) were factors in the decision to use a parsimony-based method (Giribet, 2005) to place miRNAs on the mammalian phylogeny. While their emergence in *Theria* and *Eutheria* was documented in miRGeneDB, placing 112 miRNAs onto the agreed mammalian phylogeny allowed patterns of miRNA emergence to be visualised within *Mammalia*. This evolution appears heterogeneous in nature, with varying patterns of gain and loss across the phylogeny. This gain and loss may provide evidence for homoplasy in miRNA datasets, a theory which has been discussed previously in relation to *Vertebrata* (Thomson *et al.*, 2014). However, this may in actuality be a result of varying quality in the genomic regions included in this analysis, a point which has previously been made in reference to apparent homoplasy in miRNA datasets (Tarver *et al.*, 2018). While efforts were made to overcome this issue, such as comparing genome coverage within each sampled clade and choosing the highest quality representatives in each case, the quality of vertebrate genomes remains variable, with reference genomes (e.g. *H.* sapiens) used alongside scaffold level genome assemblies (e.g. *D. novemcinctus*). Therefore, as available genomic data improves, the ability to detect miRNA gain and loss may also improve.

This study investigated the phylogenetic placement of Therian and Eutherian miRNAs, finding 2 miRNAs that emerged on the stem Therian lineage and remained fixed in all extant mammals. 12 further miRNAs emerged on the stem Eutherian lineage and remained fixed in all extant mammals. Each of these miRNAs have, to some extent, placental expression. miR-340, which emerged on the Therian lineage, has localised expression to the materno-fetal inferface of the junctional zone of the murine placenta (Schroeder *et al.*, 2018), suggesting that it may play a role in the transfer of nutrients across the placenta. miR-671, which emerged on the Therian lineage and is predicted to be conserved in all extant mammals, has been observed to be upregulated in first trimester placentas, indicating a role in the establishment of the placenta during early pregnancy (Gu *et al.*, 2013).

Of the conserved miRNAs that emerged on the Eutherian stem lineage, 5 miRNAs (miR-154, miR-185, miR-188, miR-423 and miR-542) are all differentially expressed in placentas from pregnancies affected by preeclampsia (Fu *et al.*, 2013; Harapan and Andalas, 2015; Hosseini *et*

al., 2018). Furthermore, miR-154 has been found to be differentially expressed in extracellular vesicles from pre-term vs full-term labours (Pineles *et al.*, 2007), along with miR-505 (Fallen *et al.*, 2018). miR-505 is also implicated in trophoblast invasion, showing differential expression in first-trimester vs third trimester trophoblast cells (Morales-Prieto *et al.*, 2012). miR-542 has also been implicated in a second role in the endometrium, mediating the decidualisation of human endometrial stromal cells (Tochigi *et al.*, 2017).

miR-127 is processed from the eutherian-specific retrotransposon-like 1 gene (*RTL1*) antisense transcript, has previously been shown to regulate *RTL1* levels, with miR-127 knockout mice displaying placentomegaly (Ito *et al.*, 2015b). The *RTL1* gene itself is expressed in fetal capillaries and is involved in feto-maternal interactions and placental function in late development (Kaneko-Ishino and Ishino, 2012). miR-433 is derived from a primary transcript which overlaps with the miR-127 pri-miRNA transcript (Song and Wang, 2008). In miR-127 knockout mice, miR-433 is upregulated and is downregulated in *RTL1* knockouts, suggesting a compensatory mechanism to suppress *RTL1* effects on the placenta.

Differential expression of miR-28 has been observed in the placenta throughout pregnancy, with low expression in 1st trimester placentas and high expression in term placenta (Farrokhnia *et al.*, 2014). Upregulation of miR-324 has been associated with a higher likelihood of healthy gestational weight (Rahman *et al.*, 2018). miR-331 has been associated with alterations in ATP-binding cassette reporters in the placentas of pregnancies afflicted with intra-amniotic infection (do Imperio *et al.*, 2018). Finally, miR-378 has been shown to negatively regulate Nodal, preventing Nodal-mediated placental apoptosis (Luo *et al.*, 2012).

From the literature, each of these conserved stem lineage miRNAs are implicated in placental development, early pregnancy and pregnancy maintenance. This, along with their phylogenetic placement, establishes them as possible future targets for investigation into health pregnancy and placental development. The predicted targets of stem lineage miRNAs include 15379 genes that evolved endometrial expression at the stem lineage of *Eutheria* (Lynch *et al.*, 2015), along with 130 of 287 gene families that emerged at the Eutherian stem lineage (Dunwell, Paps and

Holland, 2017). The simultaneous evolution of novel expression patterns or gene families; and binding sites for miRNAs that emerged at the same phylogenetic root indicates the formation of novel regulatory networks. These novel regulatory networks, or organ identity networks, are hypothesized to emerge at times of alterations in morphology, with the novel network supporting the variation required for the emergence of a novel organ or tissue (Wagner, 2015).

In terms of annotated functional roles for these gene targets, the Reactome Pathway analysis implicates the targets of these stem lineage miRNAs in ADP ribosylation factor like GTPase 13B (ARL13B)-mediated ciliary trafficking of inositol polyphosphate-5-phosphatase E (INPP5E), a phosphoinositide 5-phosphatase, which is involved in ciliary signalling (Dyson *et al.*, 2017) and implicated in the regulation of platelet-derived growth factor receptor α (PDGFRα), which mediates blood vessel formation and cell growth (Hellstrom *et al.*, 1999), factors which are necessary during placental invasion and formation in early pregnancy. Stem lineage miRNA targets are also overrepresented in SMAD3/4 mutations and loss of function of SMAD4 in cancer, leading to alterations in transforming growth factor beta (TGF-β) signalling. TGF-β signalling has been previously linked to NODAL signalling and implantation (Li, 2014), and its disruption during the rapid cell proliferation during cancer may be mirrored during the rapid cell proliferation and alterations occurring during early pregnancy and placental invasion.

Regulation of tumor protein p53 (tp53) the gene encoding the p53 protein, a tumor suppressor, is also among the statistically overrepresented pathways in the Reactome Pathway analysis of stem lineage miRNA targets. P53 has been implicated in the induction of human chorionic gonadotropin (hCG) expression by upregulating the hCGbeta7 (CGB7) gene (Sohr and Engeland, 2011), thereby increasing the receptivity of the human endometrium and promoting successful implantation. Another overrepresented pathway involves Neurophilin interactions with VEGF and VEGFR. Neurophilin itself has been implicated in angiogenesis via transcription factor Ets-1 signalling in human umbilical vein endothelial (HUVEC) cells (Teruyama *et al.*, 2001). Vascular endothelial growth factor (VEGF) and its receptor (VEGFR) are also implicated angiogenesis, specifically novel blood vessel formation in the placenta (Chen and Zheng, 2014).

That the targets of these novel Therian and Eutherian miRNAs include genes that underwent alterations in expression, or emerged for the first time; along with genes that are implicated in pathways related to cell differentiation, maternal recognition of pregnancy and angiogenesis further identifies these miRNAs as potential targets for future investigation into their role in altering the maternal endometrium in early pregnancy, along with the invasion and maintenance of the healthy placenta.

Among the predicted targets of stem lineage miRNAs are 84 conserved gene families which underwent positive selection on the stem Eutherian root. The binding sites for these miRNAs are conserved in species representative of major Eutherian clades. An analysis of the regulatory dynamics of stem lineage miRNAs and these conserved gene families that underwent positive selection on the stem Eutherian lineage showed that stem lineage miRNAs have on average, 6.77 binding sites on these positively selected genes, compared to an average of 1.0 binding sites in genes that did not undergo positive selection on the stem Eutherian lineage. This result suggests a degree of preferential binding between stem lineage miRNAs and these conserved positively selected genes, providing further evidence of a novel regulatory network that emerged at the root of placental mammals. Taking this, along with the previously mentioned hypothesis discussed by Wagner (Wagner, 2015), indicates that this regulatory network may have been vital in the transformation of the uterus and reproductive tissues that facilitated the evolution of mammalian placenta.

3.6 Conclusion

From these analyses, it is clear that novel regulatory elements emerged at the base of *Mammalia*. This is not surprising given the large body of literature on miRNA expansions (with miRNA populations increasing at the base of *Bilateria* and *Vertebrata*). Mapping the 112 miRGeneDB miRNAs onto the agreed mammalian phylogeny led to the identification of 14 conserved *Therian* and *Eutherian* miRNAs which arose once and were fixed in the representative species sampled.

Each of these 14 miRNAs were found to be related to placentation and reproductive pathology in the literature, which confirmed that these miRNAs were indeed involved in the placenta and not in other mammalian characteristics (e.g. hair). The targets of stem lineage miRNAs were found to be overrepresented in pathways involved in cell proliferation, hCG expression and angiogenesis - functions which are necessary for the establishment and maintenance of pregnancy . Furthermore, these 14 stem lineage miRNAs were found to target previously published datasets of genes that evolved endometrial expression at the Eutherian root, along with novel eutherian gene families.

Combining the findings from Chapter 2 with these data we have shown that gene families that underwent positive selection on the eutherian lineage have a significantly higher rate of microRNA target predictions for the 14 miRNAs than genes that did not undergo positive selection on the Eutherian stem lineage. Together these results suggest that the combined influence of transcripts undergoing positive selection and simultaneously evolving binding sites for these novel miRNAs may have played a role during the evolution of eutherian mammals. While the placenta did emerge during this time, other morphological innovations (such as the evolution of hair) occurred during eutherian evolution. However, these miRNAs and their targets are candidates for future investigations into regulatory innovations during eutherian evolution.

Chapter 4: Expression Analysis of Conserved Stem Lineage miRNAs

4.1 Aims of this chapter

1. Confirm phylogenetic placement of stem lineage miRNAs using expression analyses.
2. Confirm that stem lineage miRNAs are expressed in the maternal endometrium in selected representative species
3. Investigate stem lineage miRNA response to early pregnancy events in bovine and human model systems *in vitro*.

4.2 Introduction

As discussed in Chapter 1, maternal recognition of pregnancy and successful embryo implantation involve complex signalling pathways and large-scale changes in the transcriptomic profile of the endometrium. These dynamic alterations in gene expression, therefore, must be carefully regulated in order to establish pregnancy. miRNAs have previously been linked to uterine alterations and early pregnancy events. miRNA profiling of biopsies of proliferative and mid-luteal human endometrium revealed distinct expression patterns in each endometrial phase, with 49 miRNAs significantly differentially expressed between the two groups, suggesting a role of miRNAs in modulating shifts in the phases of the endometrium (Kuokkanen *et al.*, 2010). miRNAs found to be upregulated in mid-secretory phase endometrium were found to target cell-cycle regulators, possibly downregulating or preventing normal cell turnover during the luteal phase (Kuokkanen *et al.*, 2010). miRNAs have been associated with stromal cell decidualization in the human endometrium and the formation of the decidua during early pregnancy. miRNA profiling of menstrual decidua from patients diagnosed with secondary infertility and decidua from pregnancies terminated at the 8-week point found 9 miRNAs to be differentially expressed in post-implantation decidua when compared to menstrual decidua, including 2 stem lineage miRNAs; miR-423 and miR-532, a paralogue of miR-188 originating on the stem eutherian lineage

(Lv *et al.*, 2016). During equine maternal recognition of pregnancy, 58 miRNAs were found to be significantly differentially expressed in pregnant vs non-pregnant mares, with predicted targets implicated in adheren junctions, hippo signalling, cell proliferation and cell cycle functions. Of these, 3 miRNAs (miR-8986b, miR-19a and miR-539) had previously been found in extracellular vesicles – membrane bound vesicles which can transport molecules such as miRNAs, with miR-19a acting to reverse repression of JAK-STAT signalling to promote cell proliferation (Klohonatz *et al.*, 2019). miRNAs have also been implicated in bovine implantation. A 2014 study comparing endometria of high fertility (displayed signs of pregnancy at Day 28 and sustained pregnancy through the first trimester) cattle vs low fertility cattle (those without the mentioned criteria), found 11 miRNAs significantly differentially expressed between high- and low-receptivity cattle at Day 7 of pregnancy, indicating these miRNAs may be involved in the preparation of the endometrium for implantation and successful pregnancy (Ponsuksili *et al.*, 2014). This link between miRNAs and uterine receptivity was also established in a study of IVF patients who experienced recurrent implantation failure (RIF) (Revel *et al.*, 2011). When compared to the endometrial miRNA profile of fertile women, 13 miRNAs were found to be differentially expressed in RIF patients (10 significantly upregulated and 3 downregulated). The predicted targets of these miRNAs were enriched for cell adhesion, cell cycle pathways and Wnt signalling, functions known to be implicated in embryo implantation (van Mourik, Macklon and Heijnen, 2009). miRNAs have also been implicated in recurrent miscarriage in humans, comparing circulating miRNA levels in blood plasma from women suffering from recurrent miscarriage to women who were not pregnant. Six miRNAs were found to be differentially expressed in women suffering from recurrent miscarriage, with miR-127-3p, a stem lineage miRNA originating on the eutherian lineage, significantly downregulated in the blood plasma of recurrent miscarriage patients (Yang *et al.*, 2018), indicating that these miRNAs may promote implantation when expressed at sufficient levels.

With increasing interest in the study of miRNAs, many methods have been developed to study their expression. Popular miRNA quantification methods include reverse-transcription-quantitative PCR (RT-qPCR), RNA-sequencing and hybridization techniques (microarrays). In

2014, a miRNA quality control study (miRQC) was carried out (Mestdagh *et al.*, 2014), which analysed the three methods mentioned across 9 different vendors. Each platform was first assessed on their titration response, or the ability to accurately measure increasing miRNA expression in increasing RNA concentrations. While RNA-sequencing and hybridization methods displayed accurate titration response across platforms, the response of RT-qPCR methods varied by vendor; with miRCury, miScript and SmartChip RT-qPCR methods performing on a level with RNA-sequencing and hybridization technologies. However, TaqMan and OpenArray methods showed a much lower titration response, indicating an inability to quantify small expression changes. The titration response also correlated with reproducibility, with miRCury and SmartChip methods performing on a similar level with Affymatrix and RNA-Sequencing methods, while TaqMan and OpenArray technologies having the lowest reproducibility. RT-qPCR methods also showed higher accuracy and sensitivity, though these are restricted to known primers.

RT-qPCR-based miRNA quantification methods are favourable, due to their sensitivity and accuracy. In addition, RT-qPCR based miRNA quantification can be performed on standard qPCR machines and allow the screening of multiple samples at a lower cost when compared with RNA-sequencing methods. While RNA-sequencing can identify novel miRNAs, when an experiment involves known miRNAs, RT-qPCR provides a flexible and accurate platform for miRNA quantification and differential expression analysis. The miRCury RT-qPCR method, improved sensitivity and specificity through the use of locked nucleic acid (LNA) oligonucleotide technology. LNA oligonucleotides are a group of RNA analogues in which the ribose ring is fixed in the C3'-endo conformation by a methylene bridge, increasing the melting temperature (T_m) even in short LNA sequences, a feature that is necessary when analysing mature miRNAs (18-22 nucleotides long) (GrÜnweiler and Hartmann, 2007). A Northern Blot experiment, comparing LNA probes to DNA probes, found a ten-fold increase in sensitivity when LNA probes were used, enabling the probe to differentiate between miRNAs with similar sequences, including those with single nucleotide differences (Válóczi *et al.*, 2004). The increase in sensitivity provided by LNA probes, along with the flexibility of RT-qPCR when conducting expression analysis on a small set of known miRNAs, made this the method of choice for this analysis.

That miRNAs are involved in early pregnancy in placental mammals has been established. However, it is the specific signals of early pregnancy and how they affect miRNA expression that is of particular interest. By investigating the impact of MRP signals and molecules involved in implantation on stem lineage miRNA expression, it may be possible to gain a fuller understanding of how miRNAs contribute to the transcriptional changes induced at each stage of early pregnancy. It is our hypothesis that miRNAs that emerged on the therian and eutherian stem lineages are biologically regulated by molecules involved in early pregnancy events such as maternal recognition of pregnancy and implantation.

4.3 Materials and Methods

In order to investigate the role of miRNAs computationally predicted to have emerged on the Therian and Eutherian stem lineages (referred to as stem lineage miRNAs) in early pregnancy events, the expression of miRNAs of interest were evaluated in endometrial/uterine tissues of five species representing *Marsupiala* (opossum), *Suidae* (pig)*, Ruminanta* (cow)*, Rodentia* (mouse) and *Primates* (human). Stem lineage miRNA expression was also analysed in marsupial mammary gland tissue, in order to analyse therian miRNA expression in glands that emerged on the therian stem lineage. *In vitro* cell systems (detailed below) were used to test effect of specific molecules important for early pregnancy on miRNA expression in human and bovine endometrial cells. These molecules are the bovine MRP signal, IFNT; the conserved implantation-related protein, protein-disulphide isomerase (P4HB), with 96% sequence homology between human and bovine sequences; the conserved implantation-related protein, macrophage capping protein (CAPG), with 91% sequence homology between human and bovine sequences; and the key hormone of pregnancy progesterone (P4). P4HB and CAPG were chosen for this study due to a) their high sequence conservation, b) evidence for their secretion by the conceptus prior to implantation and c) their influence on the endometrium during implantation is not yet fully understood. Unless otherwise stated all chemicals were sourced from Sigma Aldrich.

4.3.1 Isolation of Uterine and Endometrial Tissue for Confirmation of miRNA Phylogenetic Placement

4.3.1.1 Porcine Tissue Samples

Porcine uteri were collected from 3 cyclic female pigs at bacon weight (80-100kg, weight reached at 8-10 months of age) of their life. Following slaughter at a commercial abattoir (Cranswick abattoir, Preston) the reproductive tracts were transported back to the laboratory on ice. The uterine horns were dissected away from connective tissue. Each uterine horn was opened by inserting sterile curved scissors into the uterine lumen and cutting towards the oviduct to expose the endometrium. The exposed endometrium was separated away from the underlying myometrium using sterile curved scissors. Strips of endometrial tissue were placed into

RNase/DNase free Eppendorfs, snap-frozen in liquid nitrogen and stored at -80°C prior to subsequent processing.

4.3.1.2 Murine Tissue Samples

Uteri from 3 wild-type female mice were a kind gift from Dr Lynn McKeown. Tissue was placed into Eppendorfs and snap-frozen in liquid nitrogen. Samples were stored at -80°C .

4.3.1.3 Opossum Mammary Gland and Uterine Tissue Samples

Mammary gland and uterine tissue from 5 wild type female opossums (mated and unmated) were a kind gift from Dr James Turner of the CRICK Institute. Samples were collected and snap-frozen in liquid nitrogen, and stored at -80°C prior to subsequent processing.

4.3.2 Primary Cell and Explant Collection

4.3.2.1 Bovine Endometrial Explant Culture

Bovine reproductive tracts (n=3), sourced from John Penny & Sons abattoir, Leeds, were selected from a batch (Figure 4.2). Late luteal tracts were selected in order to model maternal recognition of pregnancy and implantation, occurring during the late luteal phase (Ireland, Murphee and Coulson, 1980) and the following criteria used:

1. Presence of corpus luteum at the appropriate stage of the cycle
2. Presence of dominant follicle on ovary.

The uterine horn ipsilateral to the CL was dissected by inserting sterile curved scissors into the uterine lumen and cutting towards the oviduct to expose the endometrium and washed once with Gibco Phosphate Buffered Saline (PBS) with the addition of 1% Gibco Penicillin-Streptomycin-Glutamine 100X (GSP) (Thermo Fisher Scientific, USA). 90 5mm long explants were taken from the selected horn by excising 5mm long sections of endometrium using tweezers and small dissection scissors. Explants were washed twice in Gibco Hanks' Balanced Salt Solution (HBSS) with 1% Gibco GSP (Thermo Fisher Scientific, USA). Explants were cultured in 6 well plates with 2 explants per well in 2mL of Gibco RPMI 1640 (Thermo Fisher Scientific, USA) media with

the addition of 10% (v/v) PAA FBS Gold (PAA Laboratories, Austria) and 1% Sigma-Aldrich Antibiotic Antimycotic Solution (ABAM), maintained for 4hr in a 37°C incubator at 5% CO_2.

Figure 4.1 Late luteal stage bovine reproductive tracts (n=3). *Bovine reproductive tracts were sourced from John Penny and Sons, Leeds. Bovine reproductive tracts with a stage III, bright yellow or orange corpus luteum, along with a dominant follicle on the ovary, indicating late luteal stage of estrous were selected* (Ireland, Murphee and Coulson, 1980)*. Dissected corpus luteum with orange-yellow internal colouring is visible on (1) the right ovary, (2) the right ovary, and (3) the left ovary.*

4.3.2.2 Bovine Epithelial and Stromal Cell Isolation

Bovine reproductive tracts (n=3) were selected based on luteal phase. Late luteal tracts were selected based on criteria discussed in Section 5.3.2.1. Early luteal tracts were selected based on the presence of a corpus luteum during the early luteal phase (Ireland, Murphee and Coulson, 1980), in order to assess the effects of progesterone treatment prior to the rise in endogenous levels of progesterone indicated by the corpus luteum . Examples of late luteal stage tracts and early luteal tracts are displayed in Figures 4.2 and 4.3, respectively.

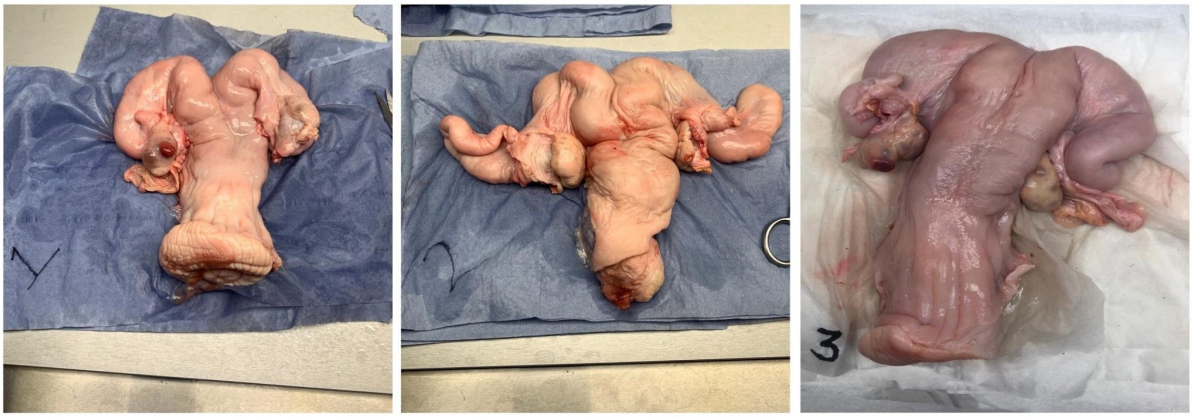


***Figure 4.2 Early luteal stage bovine reproductive tracts (n=3)**. Bovine reproductive tracts were sourced from John Penny and Sons, Leeds. Bovine reproductive tracts with a stage I, red, corpus luteum* (Ireland, Murphee and Coulson, 1980)*, indicating recent ovulation is visible on (1) the left ovary, (2) the right ovary, and (3) the base of the left ovary.*

Ipsilateral horns were dissected by inserting sterile curved scissors into the uterine lumen and cutting towards the oviduct to expose the endometrium and washed once with Gibco Phosphate Buffered Saline (PBS) with the addition of 1% Gibco Penicillin-Streptomycin-Glutamine 100X (GSP) (Thermo Fisher Scientific, USA). The endometrium was dissected from the myometrium in sheets using sterile scissors and tweezers and washed in Gibco HBSS with 1% Gibco GSP (Thermo Fisher Scientific, USA). Endometrial tissue was dissected into 3-5mm fragments and digested for 1 hour at 37°C in 40mL HBSS with the addition of 40mg BSA, 20mg collagenase II, 100µl 4% DNase I and 400µl 100X Trypsin solution. The digested endometrial tissue was first passed through a 40µM cell strainer to isolate epithelial cells, followed by a 70µM strainer to isolate stromal cells. Late luteal stromal cells were maintained in T75 flasks, in Gibco RPMI 1640 media with the addition of 10% (v/v) PAA FBS Gold and 1% Sigma-Aldrich ABAM in a 37°C incubator at 5% CO_2 for 6 days, passaging every 3 days, until the cells reached 70% confluency. Late luteal epithelial cells were maintained under the same conditions for 36 days, until they reached 70% confluency. Early luteal epithelial and stromal cells were maintained in culture for 14 days, passaging every 3 days until cells reached 70% confluency.

4.3.3 Cell Culture

Ishikawa cells (ECACC: #99040201) were incubated at 37°C and 5% CO_2 in a T75 flask. Ishikawa cells (n=3) were maintained in Gibco Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (DMEM/F-12) (Thermo Fisher Scientific, USA) with the addition of 10% (v/v) Gibco One-shot EV-

depleted FBS (Thermo Fisher Scientific, USA) (charcoal stripped using 1g Sigma-Aldrich dextran-coated charcoal) and 1% Gibco GSP. Cells were passaged every 3 days, until they reached 70% confluency.

4.3.3.1 Treatment of Bovine Endometrial Explants

Bovine endometrial explants (n=3 biological replicates) were incubated in 6-well plates, in 37°C and 5% CO_2. After 4 hours, wells were replenished with 2mL Gibco RPMI 1640 media with the addition of 10% (v/v) PAA FBS Gold and 1% Sigma-Aldrich ABAM. Explants were then treated with

i) control (100⌠l Gibco RPMI 1640 media with the addition of 10% (v/v) PAA FBS Gold and 1% Sigma-Aldritch ABAM).
ii) vehicle control (200⌠l PBS).
iii) 100⌠l 1000ng/mL ovine Interferon Tau (oIFNT).

Treated explants were incubated for 24 hours at 37°C and 5% CO_2. After 24 hours, explants were removed from wells using sterile tweezers, then placed in sterile Eppendorfs and snap-frozen in liquid nitrogen. Frozen explants were stored at -80°C.

4.3.3.2 Treatment of Primary Bovine Epithelial and Stromal Cells

Cultured epithelial and stromal cells (n=3 biological replicates) were maintained until cells had reached 70% confluency. Live cells were counted using Trypan Blue exclusion dye and a haemocytometer. Epithelial cells were diluted to 100,000 cells/mL and stromal cells were diluted to 25,000 cells/mL using Gibco RPMI 1640 media with 10% (v/v) PAA FBS Gold and 1% Sigma-Aldrich ABAM. Cells were plated at 2mL/well in 6 well plates and treated with

i) control (20⌠l Gibco RPMI 1640 media with 10% (v/v) PAA FBS Gold and 1% Sigma-Aldrich ABAM).
ii) vehicle control (20⌠l PBS).
iii) 100⌠l 1000ng/mL oIFNT.
iv) 100⌠l 1000ng/mL bCAPG.
v) 100⌠l 1000ng/mL bP4HB.

Treated cells were incubated at 37°C and 5% CO_2. After 24 hours cells were trypsinized, pelleted and lysed using the Invitrogen *mir*Vana miRNA extraction lysis buffer (from Invitrogen *mir*Vana miRNA Isolation Kit, with phenol). Lysed cells were transferred to a sterile labelled Eppendorf and were snap-frozen in liquid nitrogen. Frozen cells were then stored at -80°C.

4.3.3.3 Treatment of Primary Bovine Epithelial and Stromal Cells with P4

Cultured epithelial and stromal cells (n=3 biological replicates) were maintained until cells had reached 70% confluency. Live cells were counted using Trypan Blue and a haemocytometer. Epithelial cells were diluted to 100,000 cells/mL and stromal cells were diluted to 25,000 cells/mL using Gibco RPMI 1640 media with 10% (v/v) PAA FBS Gold and 1% Sigma-Aldrich ABAM. Cells were plated at 2mL/well in 6 well plates and treated with

i) control (10⌠l Gibco RPMI 1640 media with 10% (v/v) PAA FBS Gold and 1% Sigma-Aldrich ABAM).
ii) vehicle control (10⌠l ethanol).
iii) 10⌠l 0.1⌠g/mL P4.
iv) 100⌠l 1.0⌠g/mL P4
v) 100⌠l 10⌠g/mL P4.

Treated cells were incubated at 37°C and 5% CO_2. After 24 hours cells were trypsinized, pelleted and lysed using the Invitrogen *mir*Vana miRNA extraction lysis buffer (from Invitrogen *mir*Vana miRNA Isolation Kit, with phenol). Lysed cells were transferred to a sterile labelled Eppendorf and were snap-frozen in liquid nitrogen. Frozen cells were then stored at -80°C.

4.3.3.4 Treatment of Human Ishikawa Cells with bCAPG and bP4HB

Human Ishikawa cells (n=3 technical replicates) were maintained until cells had reached 70% confluency. Live cells were counted using Trypan Blue and a haemocytometer. Ishikawa cells were diluted to 100,000 cells/mL using DMEM/F-12 media with the addition of 10% (v/v) Gibco One-shot EV-depleted FBS (charcoal stripped using 1g Sigma-Aldrich dextran-coated charcoal) and 1% Gibco GSP. Cells were plated at 2mL cells/well in a 6 well plate and treated with

i) control (20⌠l DMEM/F-12 media with the addition of 10% (v/v) charcoal stripped Gibco One-shot EV-depleted FBS and 1% Gibco GSP).

ii) vehicle control (20⌠l PBS).

iii) 20⌠l 1000ng/mL bCAPG.

iv) 20⌠l 1000ng/mL bP4HB.

Treated cells were incubated at 37°C and 5% CO_2. After 24 hours cells were trypsinized, pelleted and lysed using the Invitrogen *mir*Vana miRNA extraction lysis buffer (from Invitrogen *mir*Vana miRNA Isolation Kit, with phenol). Lysed cells were transferred to a sterile labelled Eppendorf and were snap-frozen in liquid nitrogen. Frozen cells were then stored at -80°C.

4.3.4 RNA Extractions

Total RNA was extracted from bovine endometrial explants treated with oIFNT (9 total samples); bovine endometrial epithelial cells treated with oIFNT, bCAPG and bP4HB (15 total samples); bovine endometrial epithelial cells treated with P4 in a dose-dependent manner (15 total samples); bovine endometrial stromal cells treated with oIFNT, bCAPG and bP4HB (15 total samples); bovine endometrial stromal cells treated with P4 dose-dependent manner (15 total samples); human Ishikawa immortalised endometrial epithelial cells treated with bCAPG and bP4HB in a dose-dependent manner (24 total samples); and human Ishikawa immortalised endometrial epithelial cells treated with P4 in a dose-dependent manner (15 total samples); using the Invitrogen *mir*Vana miRNA Isolation Kit (Thermo Fisher Scientific, USA). Prior to bovine endometrial explant extraction, 50mg snap-frozen bovine endometrial explants were mixed with 500µl (10X volume/tissue mass) *mir*Vana miRNA extraction lysis buffer and homogenised using a mechanical homogenizer for 10-15 seconds. Following homogenization, RNA extraction was performed under the same conditions as RNA extraction from bovine and human endometrial cells, as follows:

Fifty µl (1/10 volume) miRNA Homogenate Additive was added to each homogenate and vortexed for 20sec. The homogenate mixture was incubated on ice for 10 minutes, after which 500µl (equal to the original tissue lysate volume) of Acid-Phenol:Chloroform was added to each sample and was vortexed for 30sec. The sample was then centrifuged at 10,000xg for 5 minutes at room temperature (RT) and the aqueous phase, which contained the RNA, was collected and the volume noted. One and a quarter volumes of 100% ethanol was added to the collected

aqueous phase, mixed and added to a filter column (700µl at a time, centrifuging at 10,000xg for 15 seconds, adding any remaining mixture to the column and centrifuging again). The filter column was washed with 700µl miRNA Wash Solution 1 and centrifuged at 10,000xg for 20 seconds, followed by 500µl Wash Solution 2/3 and a second centrifugation at 10,000xg for 20 seconds. The filter cartridge was dried by centrifuging at 16,000xg for 1 minute. RNA was eluted into a fresh collection tube using 50µl of 95°C nuclease-free water centrifuging at 10,000xg for 30 seconds. (Note: stromal cell RNA was eluted into 30µl of 95°C nuclease-free water. Previous experiments using stromal cell RNA had noted lower RNA concentrations in stromal cells. Therefore, RNA was eluted using less nuclease-free water, to avoid diluting the resulting RNA). Eluted RNA was DNase treated using the Invitrogen DNA-free kit, adding 5µl 10X DNase I buffer along with 1µl rDNase I and incubated at 37°C for 30 minutes. The reaction was stopped using 5µl DNase Inactivation Reagent. Samples were mixed by pipetting and incubated at room temperature for 2 minutes, then centrifuged at 10,000xg for 15 minutes. The aqueous phase was collected and RNA content immediately quantified using the NanoDrop N1000 (Thermo Fisher Scientific, USA). Extracted RNA was stored at -80°C.

4.3.5 Reverse Transcription of Extracted RNA

Reverse transcription was performed on extracted total RNA from bovine, murine and opossum uterine and endometrial tissue; bovine primary epithelial and stromal cells; and human Ishikawa cells using the miRCURY LNA RT kit (Qiagen, UK), according to the manufacturer's instructions, as detailed below. The process of reverse transcription using the miRCURY LNA RT kit allows simultaneous polyadenylation of mature miRNAs for stabilisation and reverse transcription of mature miRNAs for PCR amplification.

Extracted RNA was diluted to 5ng/µl using sterile DNase/RNase free water. 2µl diluted RNA was added to 2µl of 5x miRCURY RT Reaction Buffer, 1µl 10x miRCURY RT Enzyme Mix (containing Reverse Trancription enzyme and Poly(a) polymerase), and 5µl DNase/RNase free water. A RT control was prepared for each treatment set, by pooling 2µl of each diluted RNA sample, vortexing for 30 seconds and adding 2µl pooled RNA to 2µl of 5x miRCURY RT Reaction Buffer

and 6µl DNase/RNase free water (where 1µl 10x miRCURY RT Enzyme Mix was replaced with equal amounts of DNase/RNase free water). Samples were incubated at 42°C for 60 minutes to allow reverse transcription. The reaction was inactivated by incubating at 95°C for 5 minutes. Resulting cDNA was stored at -20°C.

4.3.6 miRNA Expression Profiling Using miRCURY LNA miRNA Custom PCR Panels

Quantification of stem lineage miRNA expression in untreated bovine, murine, porcine, opossum and human samples and treated human and bovine samples was performed using miRCURY LNA miRNA Custom PCR Panels in a 4x24 setup, containing 20 well-characterised primers for a custom set of miRNAs (5p and 3p mature stem lineage miRNA sequences, along with paralogues of stem lineage miRNAs that also originated on the Eutherian lineage), along with 2 normalisation genes and 2 spike-ins (Qiagen, UK; configuration #YCA21533). Quantitative, Real-Time PCR was performed using the miRCURY LNA SYBR Green PCR kit (Qiagen, UK), according to manufacturer's instructions, as detailed below.

cDNA samples were diluted 1:80 using DNase/RNase free water (2µl cDNA in 158µl water). A reaction mix was prepared for each sample (control, vehicle control, and treatment sample (s)): 112µl diluted cDNA was added to 140µl 2x miRCURY LNA SYBR Green PCR Master Mix (containing miRCURY SYBR Green PCR Buffer with SYBR Green I, dNTP mix and QuantiNova DNA Polymerase) and 28µl DNase/RNase free water. 10µl of this reaction mix was applied to 24 wells using the Gilson DISTRIMAN Repetitive Pipette with 125µl DISTRITIP Syringes. This process was repeated with 4 samples per 96-well plate. Plates were then sealed, centrifuged at 1000xg for 30 seconds and placed into a Roche LightCycler (Roche Life Science, UK). *(Note: RT-qPCR of untreated bovine, murine, porcine, opossum and human sample, and bovine endometrial explants treated with 1000ng/µl oIFNT were performed on a Roche LightCycler 480. RT-qPCR of human Ishikawa and bovine epithelial and stromal cells treated with bCAPG and bP4HB were performed on a Roche LightCycler 96. The same analysis parameters were used on each machine)*.

The plate was first incubated at 95°C for 2 minutes to activate the QuantiNova DNA Polymerase. cDNA strands were denatured by incubating at 95°C for 10 seconds, followed by incubation at 56°C for 60 seconds to allow a combined annealing/extension step. The denaturation and combined annealing/extension steps were repeated for 45 cycles, followed by a melting curve performed using a 95°C incubation for 1 minute to denature the DNA, rapid cooling and incubation at 60°C for 30 seconds, followed by slow ramping of temperature back to 95°C, with fluorescence reading performed at each 1°C temperature increment. At the end of the reaction, the raw Ct values and melt curves (Figure 4.4) were exported from the LightCycler 480/96 software. Ct values of 36 or above were determined to be negative, as per the Qiagen Ct analysis parameters. ΔCt values were obtained using 5S rRNA and U6 snRNA normalisation genes as follows: $\Delta C_T = C_T^{miRNA} - AVG\ C_T^{normalisation\ genes}$. The ΔΔCt for each miRNA was calculated as follows: $\Delta\Delta C_T = \Delta C_T - AVG\ \Delta C_T^{vehicle\ control}$. Fold change was then obtained for the construction of graphs ($2^{-\Delta\Delta C_T}$) (Livak and Schmittgen, 2001). Paired 2-tailed t-tests were performed on dCt values in GraphPad PRISM, comparing vehicle control to treatment samples, where miRNAs were determined to be differentially expressed when p≤0.05.

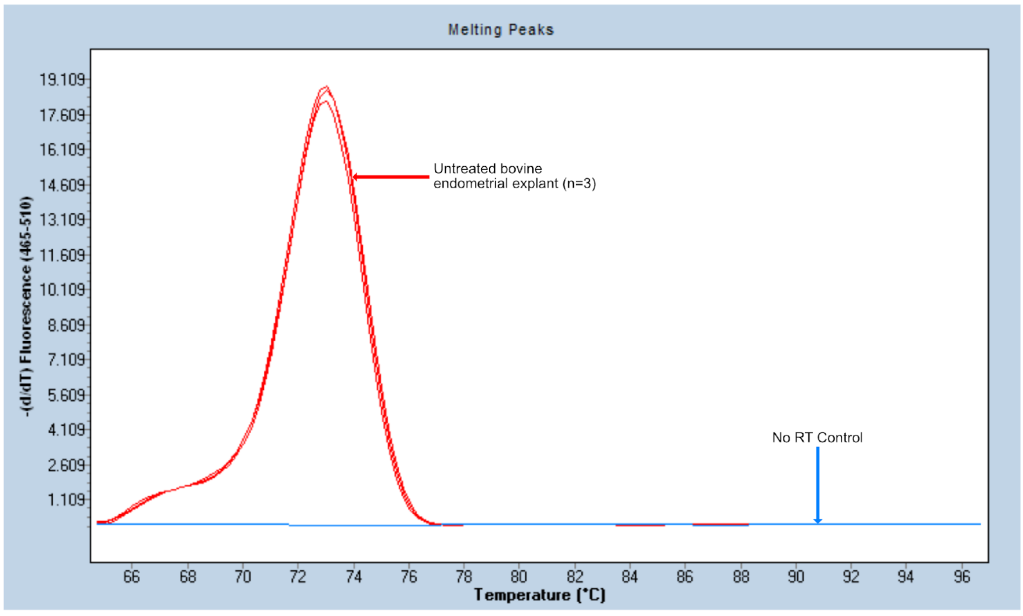


Figure 4.3 Dissociation curve for miR-28-3p in bovine endometrial explant cDNA treated with media control. *Total RNA was extracted from bovine endometrial explants and reverse transcribed into cDNA. The expression profile of this miRNA was investigated using miRCURY LNA miRNA Custom PCR Panels with SYBR Green I. The resulting amplification and dissociation curves of miR-28-3p in n=3 wells containing cDNA and wells containing No RT Controls show normal miR-28 expression and that no genomic DNA contamination was present.*

4.3.7 Reactome Pathway Analysis of the Targets of Differentially Expressed miRNAs

Targets for miRNAs computationally identified as having emerged at the stem Therian and Eutherian lineages that remained fixed in extant mammals were predicted using TargetScan7.2 (Chapter 3). Targets of miRNAs determined to be differentially expressed in *in-vitro* models of early pregnancy in human and bovine samples (when $p \leq 0.05$) were converted from Ensembl transcript identifiers to gene names using Ensembl BioMart (Zerbino *et al.*, 2017). Reactome pathways were constructed from these gene names using the Reactome Pathway Database (Jassal *et al.*, 2019). Separate pathways were reconstructed for the targets of i) each differentially expressed miRNA separately, and ii) pooled targets of all miRNAs differentially expressed in a treatment group (i.e. all significantly differentially expressed miRNAs in human Ishikawa cells treated with 1000ng/µl bP4HB). A statistical overrepresentation analysis was conducted and the 25 most significantly over-expressed pathways (ranked by p value) were assessed for functions related to early pregnancy. If a pathway was within the 25 most significantly over-expressed pathways, but did not have a p-value ≤ 0.05, the pathway was used as a predictor of function for a group of targets rather than a statement of functional overrepresentation.

4.4 Results

4.4.1 Investigation of stem lineage miRNA expression in female reproductive tissue using representative species

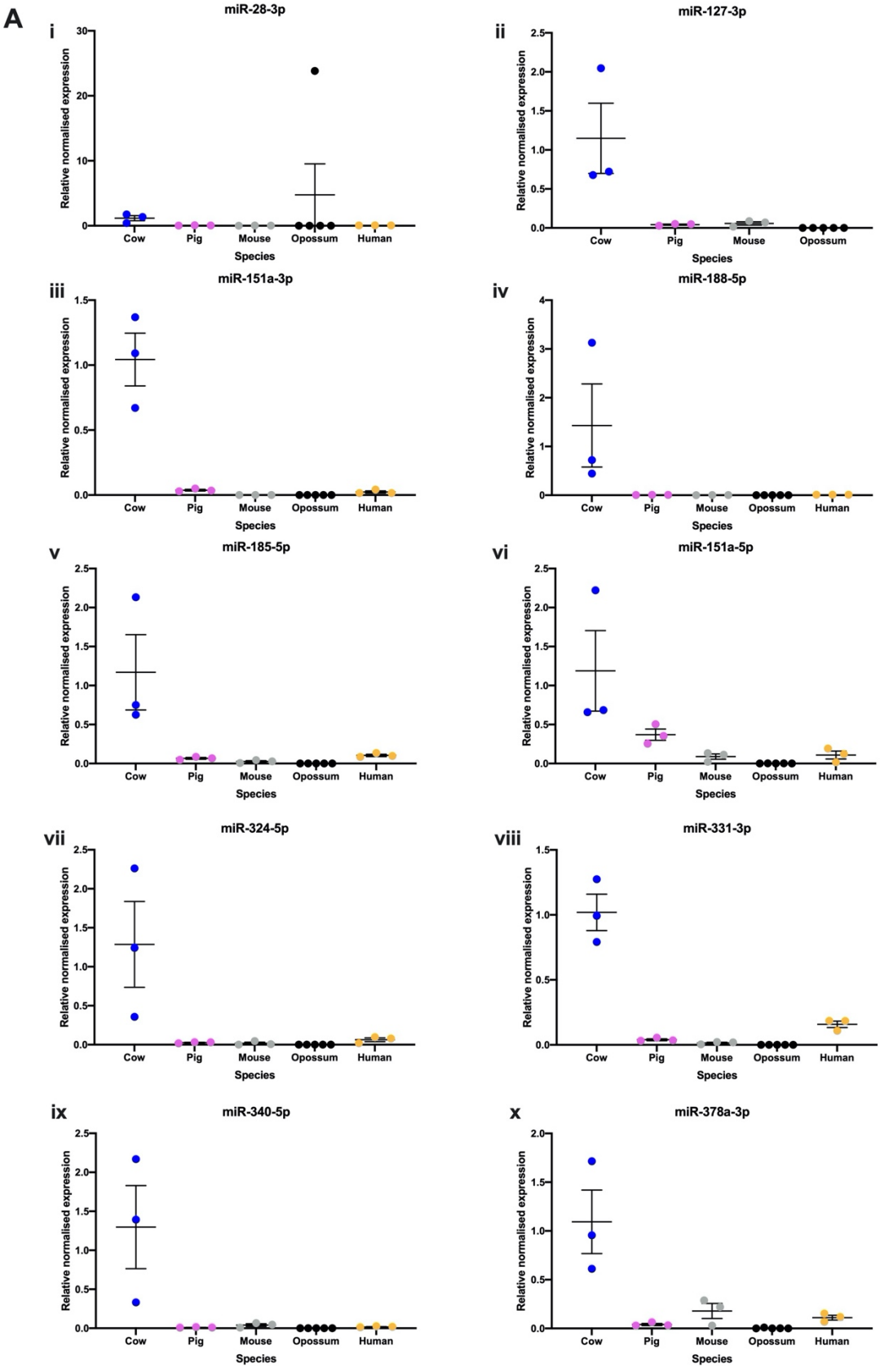
Using samples from five species representative of various branches of the mammalian tree of life (detailed in Table 4.1), expression of miRNAs computationally predicted to have originated at the root of Therian and Eutherian mammals was investigated.

Table 4.1 Species used to investigate miRNA phylogenetic placement, along with species order and mammalian subclass.

Species	Order	Subclass	Placenta Type
Homo sapiens	Primates	Eutheria	Hemochorial, villous, discoid
Mus musculus	Rodentia	Eutheria	Hemochorial, labyrinthine, discoid
Sus domesticus	Suidae	Eutheria	Epitheliochorial, diffuse
Bos taurus taurus	Ruminantia	Eutheria	Epitheliochorial, villous, cotyledonary
Monodelphus domestica	Marsupiala	Metatheria	Yolk sac

Species used to analyse stem lineage miRNA expression patterns in female reproductive tissues. Four eutherian species were used from across the eutherian phylogeny. One outgroup marsupial species, Opossum, was used. Placenta type for each species is given.

All miRNAs were confirmed to be expressed in endometrial and uterine tissue from Eutherian mammal samples (Figure 4.4). Expression values could not be determined for Opossum samples. Ct values of miR-340 and miR-671 (computationally predicted to have emerged on the Therian root) were above the cut-off Ct value of 36. miR-28-3p had a Ct value of 20.22 in opossum uterine sample 88.1g. miR-423-5p had Ct values of 32-40 across all opossum uterine and mammary gland samples. The expression levels of each computationally predicted stem lineage miRNA are shown in Figure 4.4.


A
i
miR-28-3p
Relative normalised expression
Cow
Pig
Mouse
Opossum
Human
Species
ii
miR-127-3p
iii
miR-151a-3p
iv
miR-188-5p
v
miR-185-5p
vi
miR-151a-5p
vii
miR-324-5p
viii
miR-331-3p
ix
miR-340-5p
x
miR-378a-3p

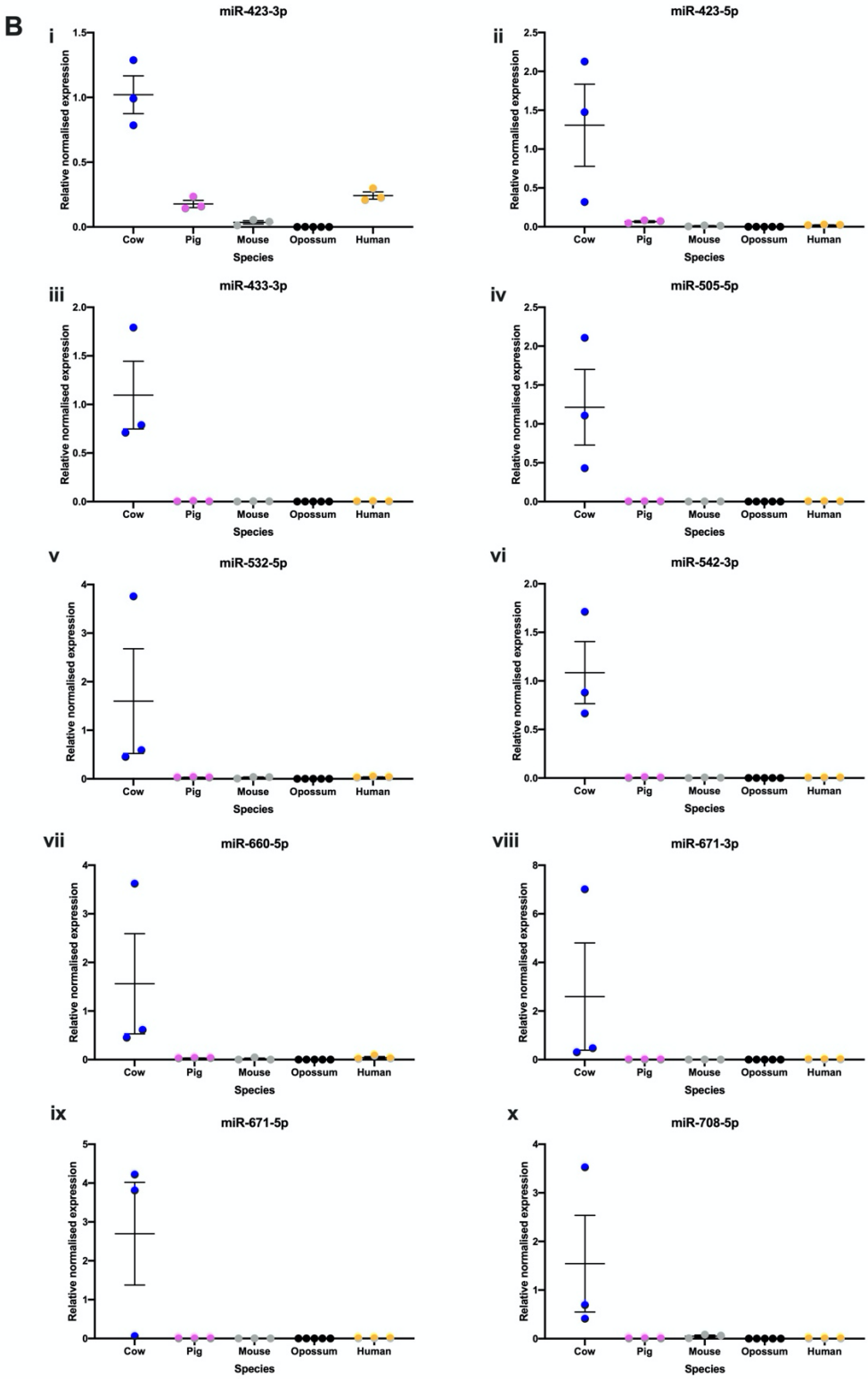
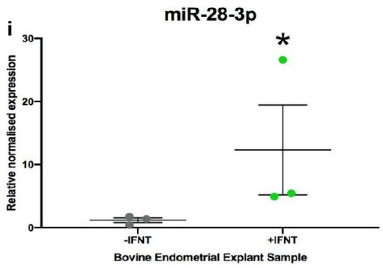
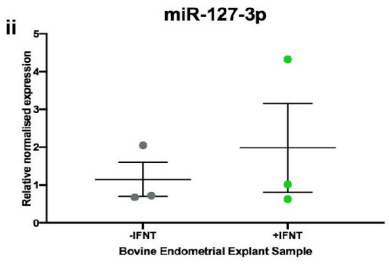


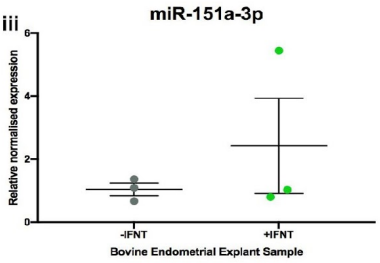
Figure 4.4 Expression of stem lineage miRNAs in representative therian species. *Expression of stem lineage miRNA (i) miR-28-3p, (ii) miR-127-3p, (iii) miR-151a-3p, (iv) miR-151a-5p, (v) miR-185-5p, (vi) miR-188-5p, (vii) miR-324-5p, (viii) miR-331-3p, (ix) miR-340-5p, (x) miR-378a-5p, (xi) miR-423-3p, (xii) miR-423-5p, (xiii) miR-433-3p, (xiv) miR-505-5p, (xv) miR-532-5p, (xvi) miR-542-3p, (xvii) miR-660-5p, (xviii) miR-671-3p, (xix) miR-671-5p and (xx) miR-708-5p in bovine endometrial tissue (blue circle), porcine endometrial tissue (pink circle), murine uterine tissue (grey circle), human Ishikawa immortalized endometrial epithelial cells (yellow circle) and opossum uterine tissue (black circle).*

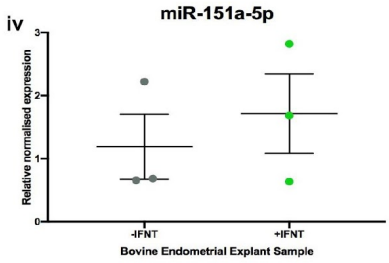
4.4.2 Expression analysis of bovine endometrial explants treated with oIFNT

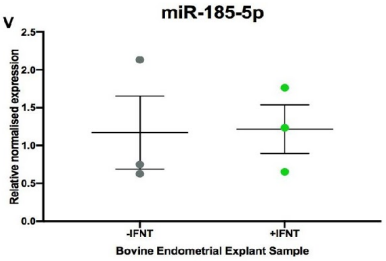
Expression of all miRNAs of interest was detected in this in vitro system/explant type. Treatment of bovine endometrial explants for 24 hr with roIFNT resulted in increased expression of mir-28-3p ($P<0.05$). The was no other effect on any of the other miRNAs examined in this system ($P>0.05$). miRNA expression levels are shown in Figure 4.5.

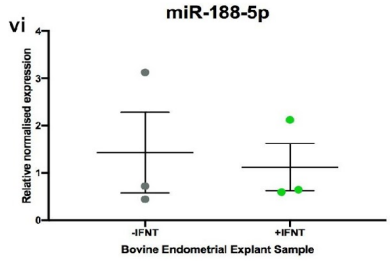

i
miR-28-3p
*
Relative normalised expression
30
20
10
0
-IFNT
+IFNT
Bovine Endometrial Explant Sample

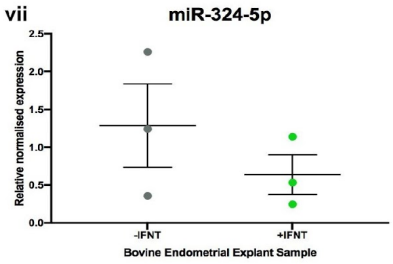

ii
miR-127-3p
Relative normalised expression
5
4
3
2
1
0
-IFNT
+IFNT
Bovine Endometrial Explant Sample

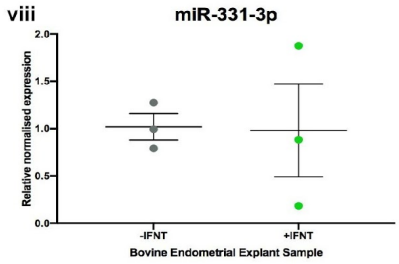

iii
miR-151a-3p
Relative normalised expression
6
4
2
0
-IFNT
+IFNT
Bovine Endometrial Explant Sample

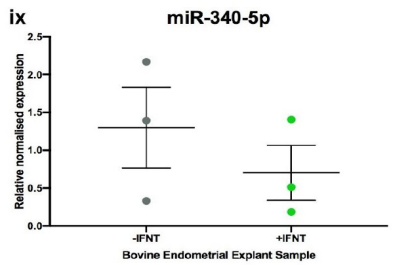

iv
miR-151a-5p
Relative normalised expression
3
2
1
0
-IFNT
+IFNT
Bovine Endometrial Explant Sample

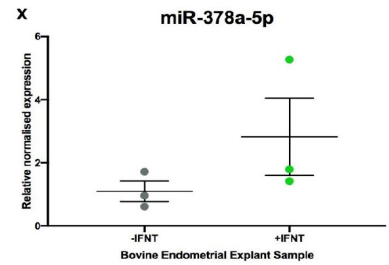

v
miR-185-5p
Relative normalised expression
2.5
2.0
1.5
1.0
0.5
0.0
-IFNT
+IFNT
Bovine Endometrial Explant Sample


vi
miR-188-5p
Relative normalised expression
4
3
2
1
0
-IFNT
+IFNT
Bovine Endometrial Explant Sample


vii
miR-324-5p
Relative normalised expression
2.5
2.0
1.5
1.0
0.5
0.0
-IFNT
+IFNT
Bovine Endometrial Explant Sample


viii
miR-331-3p
Relative normalised expression
2.0
1.5
1.0
0.5
0.0
-IFNT
+IFNT
Bovine Endometrial Explant Sample


ix
miR-340-5p
Relative normalised expression
2.5
2.0
1.5
1.0
0.5
0.0
-IFNT
+IFNT
Bovine Endometrial Explant Sample


x
miR-378a-5p
Relative normalised expression
6
4
2
0
-IFNT
+IFNT
Bovine Endometrial Explant Sample

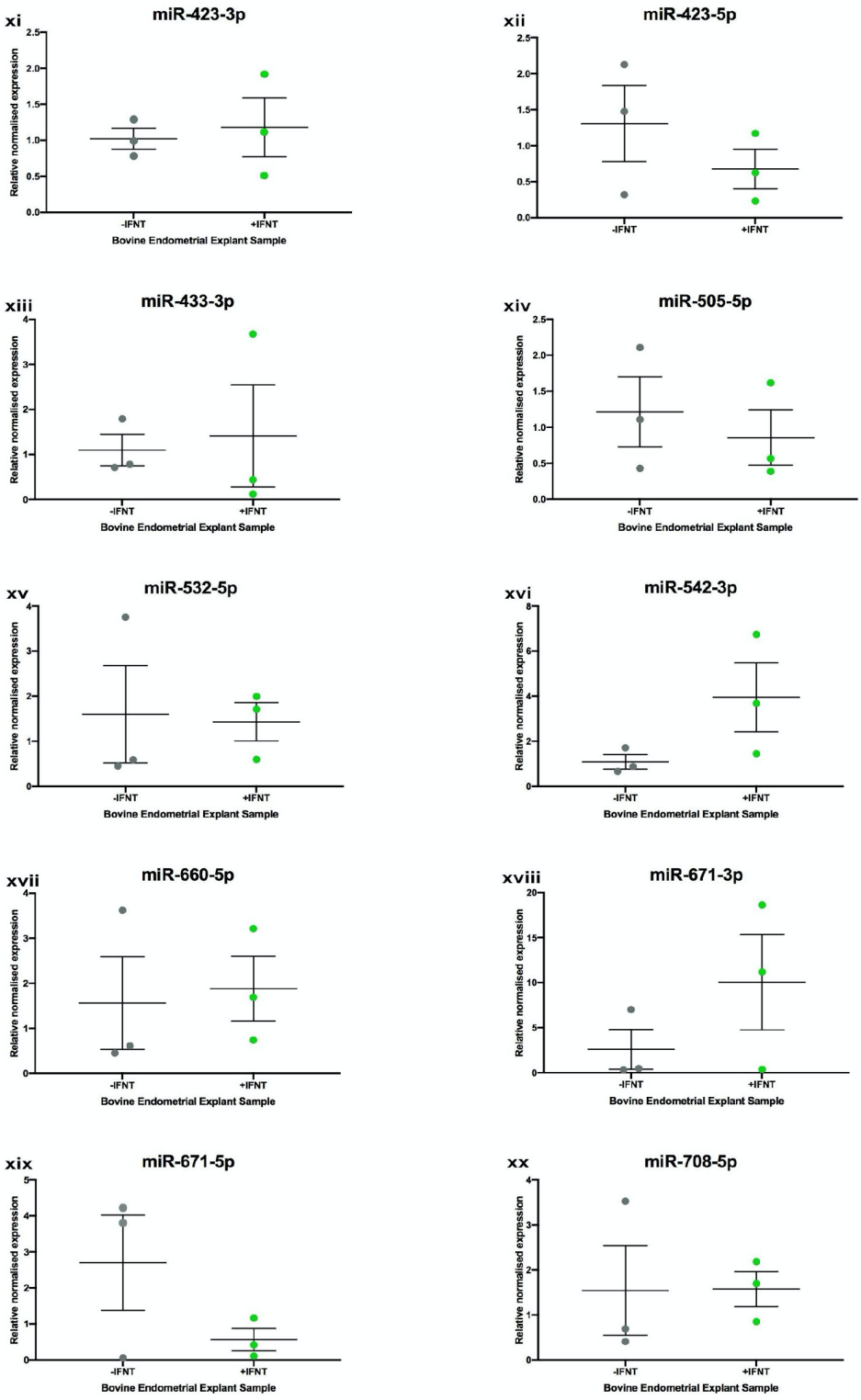


Figure 4.5 Expression of stem lineage miRNAs in bovine endometrial explants treated with oIFNT. *Expression of stem lineage miRNA (i) miR-28-3p, (ii) miR-127-3p, (iii) miR-151a-3p, (iv) miR-151a-5p, (v) miR-185-5p, (vi) miR-188-5p, (vii) miR-324-5p, (viii) miR-331-3p, (ix) miR-340-5p, (x) miR-378a-5p, (xi) miR-423-3p, (xii) miR-423-5p, (xiii) miR-433-3p, (xiv) miR-505-5p, (xv) miR-532-5p, (xvi) miR-542-3p, (xvii) miR-660-5p, (xviii) miR-671-3p, (xix) miR-671-5p and (xx) miR-708-5p in bovine endometrial explants treated with vehicle control (grey circle), or 1000ng/µl oIFNT (green circle) for 24 hours. Significant differences in miRNA expression values determined when p≤0.05 are depicted by an asterisk (*).*

4.4.3 Reactome pathway results for predicted targets of miR-28-3p, a stem lineage miRNA significantly upregulated in bovine endometrial explants treated with oIFNT

Of the 25 most significant pathways for predicted targets of miR-28-3p (ranked by p-value), all were statistically significant (p-value ≤ 0.05), shown in bold in Table 4.2.

Table 4.2 Reactome pathway analysis results for stem lineage miRNAs biologically regulated by oIFNT in bovine endometrial explants.

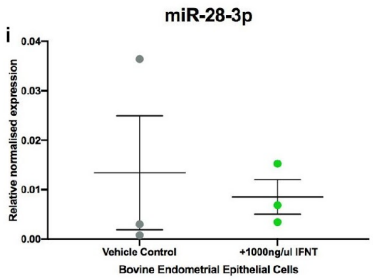
Pathway name	Entities			Reactions		
	found	ratio	p-value	FDR*	found	ratio
MECP2 regulates neuronal receptors and channels	17/32	0.002	0.001	0.774	18/26	0.002
GRB2 events in ERBB2 signaling	12/21	0.001	0.003	0.774	4/4	3.20e-04
GRB2 events in EGFR signaling	10/16	0.001	0.004	0.774	2/3	2.40e-04
Signaling by EGFR	24/60	0.004	0.005	0.774	45/48	0.004
SHC1 events in EGFR signaling	10/19	0.001	0.011	0.774	4/4	3.20e-04
ERBB2 Regulates Cell Motility	10/19	0.001	0.011	0.774	2/2	1.60e-04
Estrogen-dependent nuclear events downstream of ESR-membrane signaling	13/29	0.002	0.015	0.774	7/12	9.61e-04
SHC1 events in ERBB4 signaling	10/21	0.001	0.021	0.774	4/4	3.20e-04
cGMP effects	9/18	0.001	0.021	0.774	4/4	3.20e-04
ERBB2 Activates PTK6 Signaling	9/18	0.001	0.021	0.774	2/2	1.60e-04
MET interacts with TNS proteins	4/5	3.46e-04	0.026	0.774	2/2	1.60e-04
Adherens junctions interactions	14/35	0.002	0.027	0.774	12/16	0.001
PI3K events in ERBB2 signaling	10/22	0.002	0.027	0.774	7/7	5.61e-04
mRNA Splicing - Major Pathway	54/185	0.013	0.029	0.774	9/9	7.21e-04
SMAD4 MH2 Domain Mutants in Cancer	3/3	2.08e-04	0.03	0.774	1/1	8.01e-05
Loss of Function of SMAD4 in Cancer	3/3	2.08e-04	0.03	0.774	1/1	8.01e-05
SMAD2/3 MH2 Domain Mutants in Cancer	3/3	2.08e-04	0.03	0.774	1/1	8.01e-05
L13a-mediated translational silencing of Ceruloplasmin expression	37/120	0.008	0.032	0.774	3/3	2.40e-04
SHC1 events in ERBB2 signaling	14/36	0.002	0.033	0.774	5/6	4.80e-04
Activation of NMDA receptors and postsynaptic events	35/113	0.008	0.034	0.774	66/70	0.006
GAB1 signalosome	10/23	0.002	0.035	0.774	9/11	8.81e-04
mRNA Splicing	56/196	0.014	0.037	0.774	14/14	0.001
EGFR interacts with phospholipase C-gamma	6/11	7.62e-04	0.038	0.774	3/3	2.40e-04
Major pathway of rRNA processing in the nucleolus and cytosol	54/189	0.013	0.04	0.774	7/7	5.61e-04
EGFR downregulation	14/37	0.003	0.04	0.774	22/22	0.002

Reactome pathway results for miR-28-3p, a stem lineage miRNA significantly upregulated in bovine endometrial explants treated with 1000ng/µl oIFNT for 24 hours. 'Entities - found' refers to the number of curated molecules within the named pathway found in the submitted dataset. 'Entities – ratio' refers to the number of curated molecules represented in the pathway compared to the entire database. 'Reactions – found' refers to the number of reactions within the named pathway represented by molecules in the submitted dataset. 'Reactions – ratio' refers to the number of Reactome reactions represented by the reactions in this pathway. False Discovery Rate (FDR) is calculated using the Benjamini-Hochberg method.

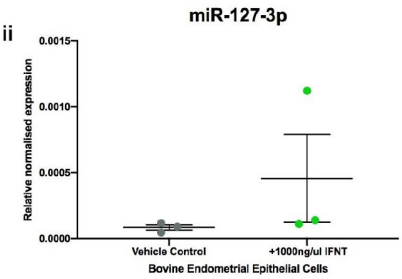
4.4.4 Expression analysis of bovine endometrial epithelial cells treated with oIFNT

No significant differences were found in stem lineage miRNA expression in bovine endometrial epithelial cells treated with oIFNT compared to vehicle control. Relative normalised miRNA expression levels are depicted in Figure 4.6.

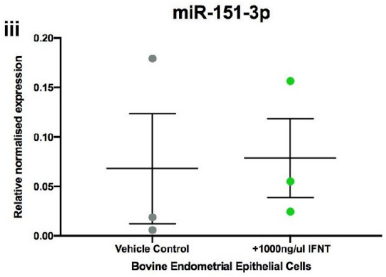
i



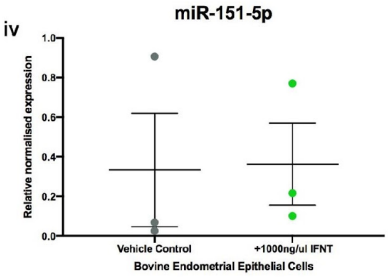
ii



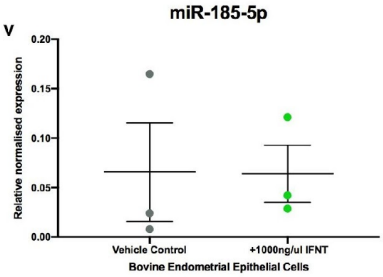
iii



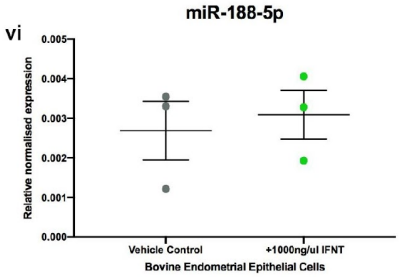
iv



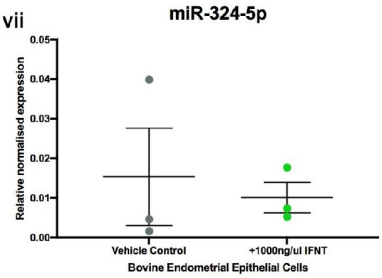
v



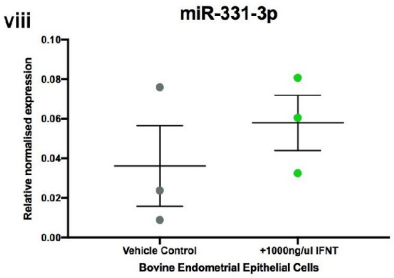
vi



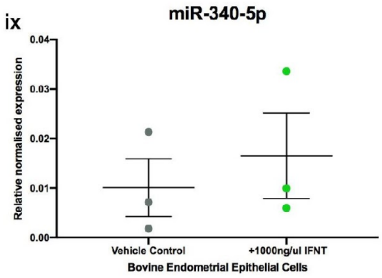
vii



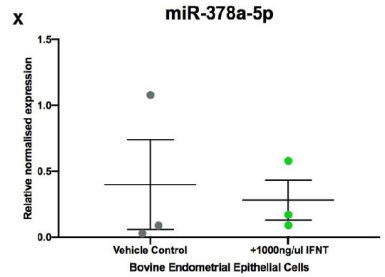
viii



ix



x



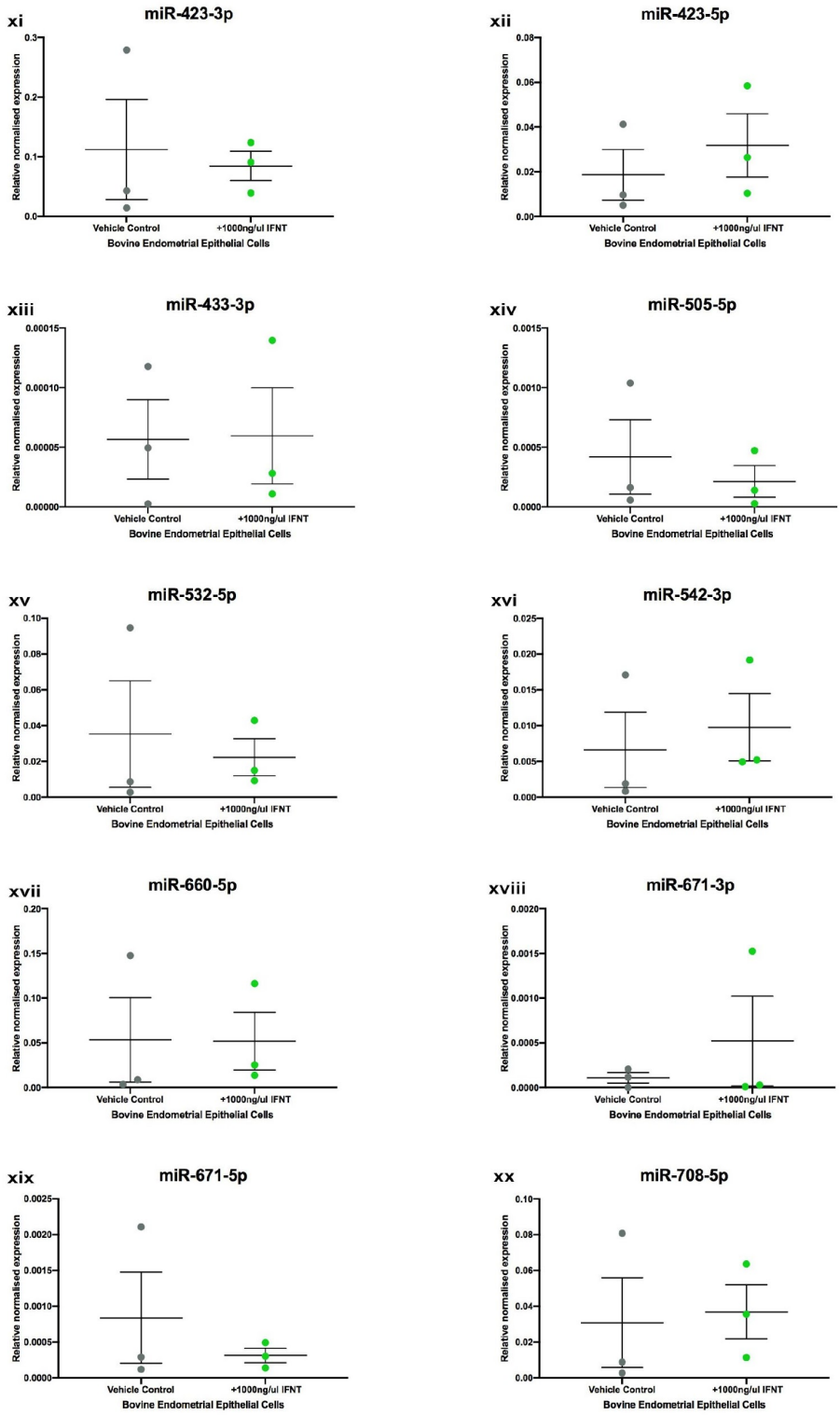
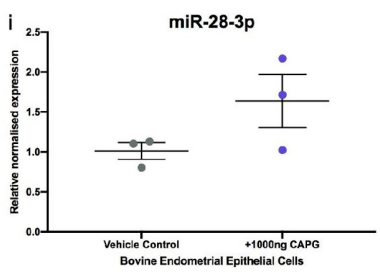
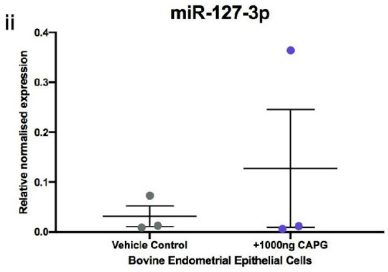


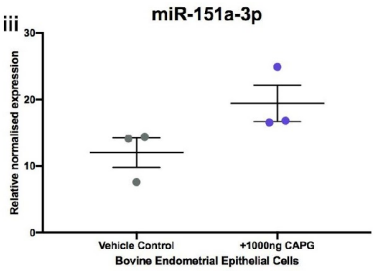
Figure 4.6 Expression of stem lineage miRNAs in bovine endometrial epithelial cells treated with oIFNT. *Expression of stem lineage miRNA (i) miR-28-3p, (ii) miR-127-3p, (iii) miR-151a-3p, (iv) miR-151a-5p, (v) miR-185-5p, (vi) miR-188-5p, (vii) miR-324-5p, (viii) miR-331-3p, (ix) miR-340-5p, (x) miR-378a-5p, (xi) miR-423-3p, (xii) miR-423-5p, (xiii) miR-433-3p, (xiv) miR-505-5p, (xv) miR-532-5p, (xvi) miR-542-3p, (xvii) miR-660-5p, (xviii) miR-671-3p, (xix) miR-671-5p and (xx) miR-708-5p in bovine endometrial epithelial cells treated with vehicle control (grey circle), or 1000ng/µl oIFNT (green circle) for 24 hours. No significant difference in miRNA expression was found.*

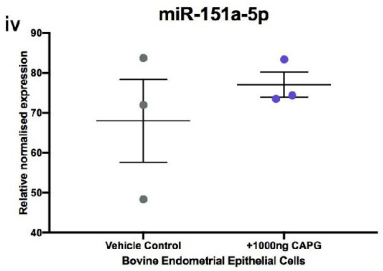
4.4.5 Expression analysis of bovine endometrial epithelial cells treated with bCAPG

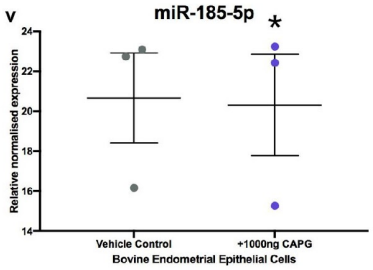
miR-331-3p, miR-324-5p and miR-505-5p showed a significant increase in expression in bovine endometrial epithelial cells treated with 1000ng/µl bCAPG compared to vehicle control. miRNA expression levels are shown in Figure 4.7.

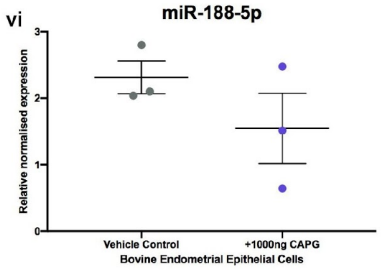

i
miR-28-3p
Relative normalised expression
2.5
2.0
1.5
1.0
0.5
0.0
Vehicle Control
+1000ng CAPG
Bovine Endometrial Epithelial Cells

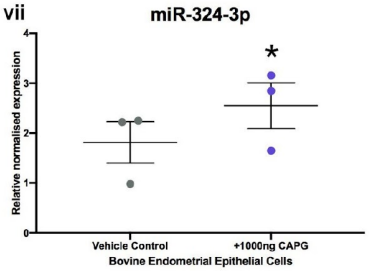

ii
miR-127-3p
Relative normalised expression
0.4
0.3
0.2
0.1
0.0
Vehicle Control
+1000ng CAPG
Bovine Endometrial Epithelial Cells

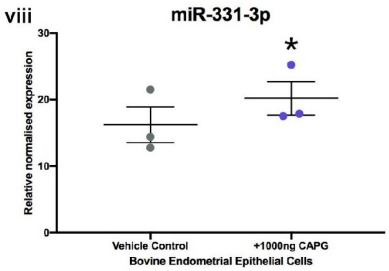

iii
miR-151a-3p
Relative normalised expression
30
20
10
0
Vehicle Control
+1000ng CAPG
Bovine Endometrial Epithelial Cells

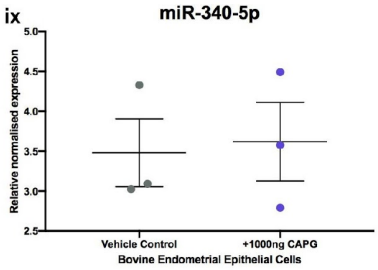

iv
miR-151a-5p
Relative normalised expression
90
80
70
60
50
40
Vehicle Control
+1000ng CAPG
Bovine Endometrial Epithelial Cells

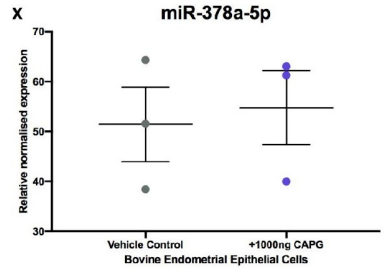

v
miR-185-5p
*
Relative normalised expression
24
22
20
18
16
14
Vehicle Control
+1000ng CAPG
Bovine Endometrial Epithelial Cells


vi
miR-188-5p
Relative normalised expression
3
2
1
0
Vehicle Control
+1000ng CAPG
Bovine Endometrial Epithelial Cells


vii
miR-324-3p
*
Relative normalised expression
4
3
2
1
0
Vehicle Control
+1000ng CAPG
Bovine Endometrial Epithelial Cells


viii
miR-331-3p
*
Relative normalised expression
30
20
10
0
Vehicle Control
+1000ng CAPG
Bovine Endometrial Epithelial Cells


ix
miR-340-5p
Relative normalised expression
5.0
4.5
4.0
3.5
3.0
2.5
Vehicle Control
+1000ng CAPG
Bovine Endometrial Epithelial Cells


x
miR-378a-5p
Relative normalised expression
70
60
50
40
30
Vehicle Control
+1000ng CAPG
Bovine Endometrial Epithelial Cells

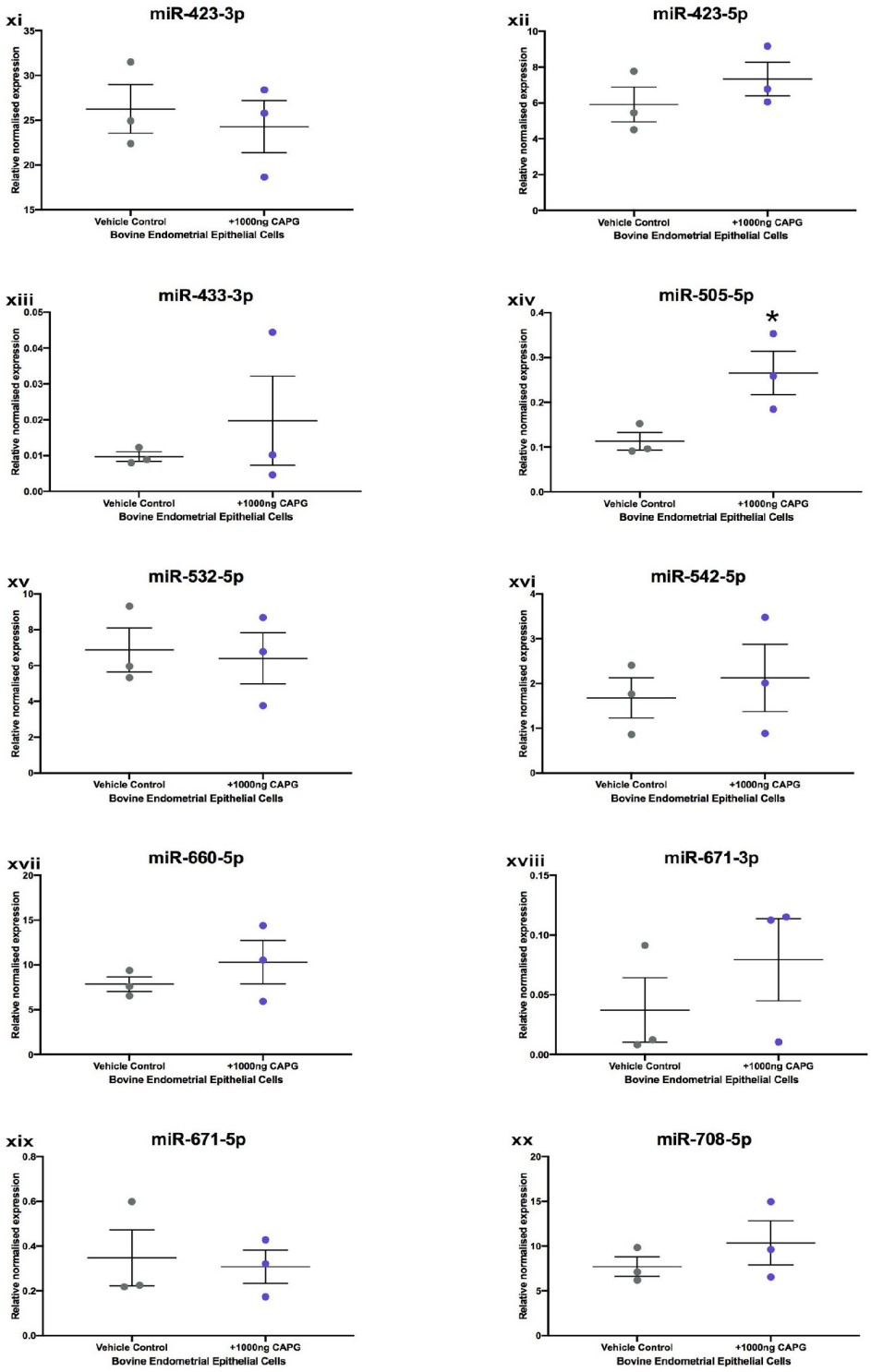


Figure 4.7 Expression of stem lineage miRNAs in bovine endometrial epithelial cells treated with bCAPG. *Expression of stem lineage miRNA (i) miR-28-3p, (ii) miR-127-3p, (iii) miR-151a-3p, (iv) miR-151a-5p, (v) miR-185-5p, (vi) miR-188-5p, (vii) miR-324-5p, (viii) miR-331-3p, (ix) miR-340-5p, (x) miR-378a-5p, (xi) miR-423-3p, (xii) miR-423-5p, (xiii) miR-433-3p, (xiv) miR-505-5p, (xv) miR-532-5p, (xvi) miR-542-3p, (xvii) miR-660-5p, (xviii) miR-671-3p, (xix) miR-671-5p and (xx) miR-708-5p in bovine endometrial epithelial cells treated with vehicle control (grey circle), or 1000ng/µl bCAPG (purple circle) for 24 hours. Significant differences in miRNA expression values determined when p≤0.05 are depicted by an asterisk (*).*

4.4.6 Reactome pathway results for predicted targets of stem lineage miRNAs significantly upregulated in bovine endometrial epithelial cells treated with bCAPG

Of the 25 most significant pathways for predicted targets of stem lineage miRNAs significantly upregulated in bovine endometrial epithelial cells treated with bCAPG (ranked by p-value), none were statistically significant (p-value ≤ 0.05), shown in Table 4.3.

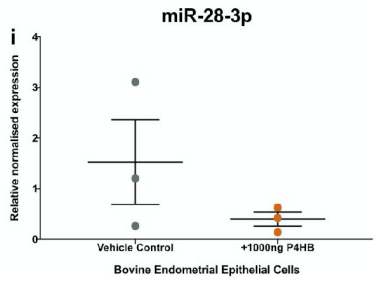
Table 4.3 Reactome pathway analysis results for miRNAs biologically regulated by bCAPG in bovine endometrial epithelial cells

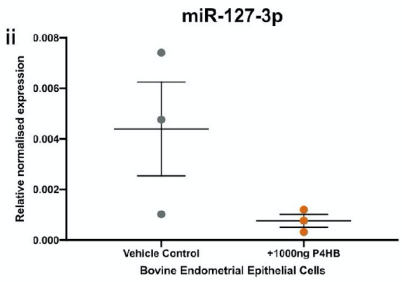
Pathway name	Entities				Reactions	
	found	ratio	p-value	FDR*	found	ratio
RUNX1 interacts with co-factors whose precise effect on RUNX1 targets is not known	32/39	0.003	0.065	0.914	5/5	4.00e-04
Adherens junctions interactions	29/35	0.002	0.07	0.914	16/16	0.001
Neurexins and neuroligins	45/60	0.004	0.105	0.914	19/19	0.002
RUNX1 regulates expression of components of tight junctions	8/8	5.54e-04	0.124	0.914	6/6	4.80e-04
VEGF ligand-receptor interactions	8/8	5.54e-04	0.124	0.914	4/4	3.20e-04
VEGF binds to VEGFR leading to receptor dimerization	8/8	5.54e-04	0.124	0.914	3/3	2.40e-04
Nectin/Necl trans heterodimerization	7/7	4.85e-04	0.144	0.914	8/8	6.41e-04
Apoptotic cleavage of cell adhesion proteins	10/11	7.62e-04	0.145	0.914	10/10	8.01e-04
CRMPs in Sema3A signaling	15/18	0.001	0.149	0.914	5/5	4.00e-04
LGI-ADAM interactions	12/14	9.70e-04	0.159	0.914	5/5	4.00e-04
RNA Polymerase III transcription initiation from type 2 promoter	21/27	0.002	0.165	0.914	4/4	3.20e-04
GRB7 events in ERBB2 signaling	6/6	4.16e-04	0.167	0.914	1/1	8.01e-05
Molecules associated with elastic fibres	28/38	0.003	0.19	0.914	10/10	8.01e-04
SUMOylation of DNA methylation proteins	13/16	0.001	0.191	0.914	3/3	2.40e-04
RUNX1 regulates transcription of genes involved in WNT signaling	8/9	6.24e-04	0.193	0.914	4/4	3.20e-04
Semaphorin interactions	49/70	0.005	0.196	0.914	40/40	0.003
Protein methylation	15/19	0.001	0.198	0.914	9/9	7.21e-04
ERBB2 Regulates Cell Motility	15/19	0.001	0.198	0.914	2/2	1.60e-04
RNA Polymerase III transcription initiation from type 1 promoter	21/28	0.002	0.207	0.914	4/5	4.00e-04
Potassium Channels	70/103	0.007	0.211	0.914	18/19	0.002
Interleukin-33 signaling	4/4	2.77e-04	0.233	0.914	2/2	1.60e-04
Neurophilin interactions with VEGF and VEGFR	4/4	2.77e-04	0.233	0.914	2/2	1.60e-04
RUNX3 regulates RUNX1-mediated transcription	4/4	2.77e-04	0.233	0.914	2/2	1.60e-04
TGFBR1 LBD Mutants in Cancer	4/4	2.77e-04	0.233	0.914	1/1	8.01e-05
BH3-only proteins associate with and inactivate anti-apoptotic BCL-2 members	9/11	7.62e-04	0.239	0.914	4 / 4	3.20e-04

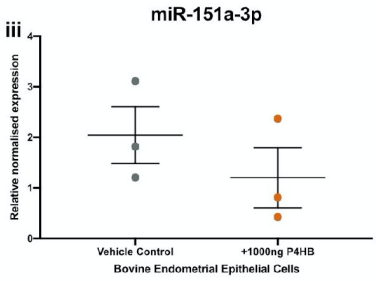
Reactome pathway results for 3 stem lineage miRNAs significantly upregulated in bovine endometrial epithelial cells treated with 1000ng/µl bCAPG for 24 hours. 'Entities - found' refers to the number of curated molecules within the named pathway found in the submitted dataset. 'Entities – ratio' refers to the number of curated molecules represented in the pathway compared to the entire database. 'Reactions – found' refers to the number of reactions within the named pathway represented by molecules in the submitted dataset. 'Reactions – ratio' refers to the number of Reactome reactions represented by the reactions in this pathway. False Discovery Rate (FDR) is calculated using the Benjamini-Hochberg method.

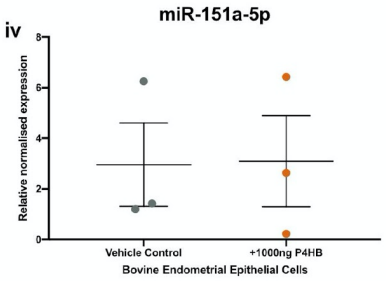
4.4.7 Expression analysis of bovine endometrial epithelial cells treated with bP4HB

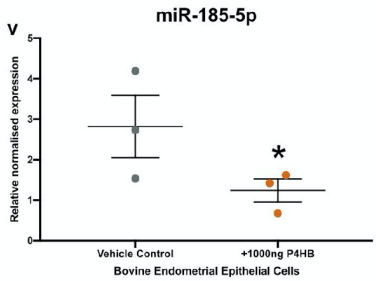
miR-185-5p showed a significant decrease in expression in bovine endometrial epithelial cells treated with 1000ng/µl bP4HB compared to vehicle control. miRNA expression levels are shown in Figure 4.8.

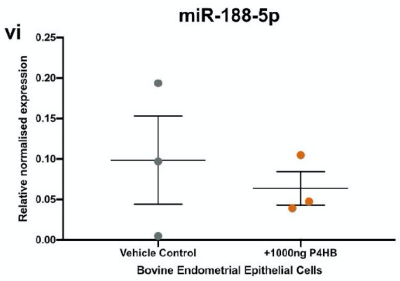

i
miR-28-3p
Relative normalised expression
Vehicle Control
+1000ng P4HB
Bovine Endometrial Epithelial Cells

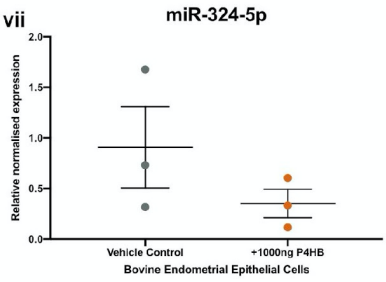

ii
miR-127-3p
Relative normalised expression
Vehicle Control
+1000ng P4HB
Bovine Endometrial Epithelial Cells

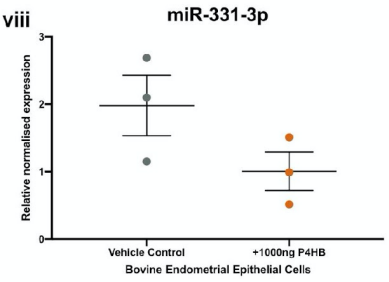

iii
miR-151a-3p
Relative normalised expression
Vehicle Control
+1000ng P4HB
Bovine Endometrial Epithelial Cells

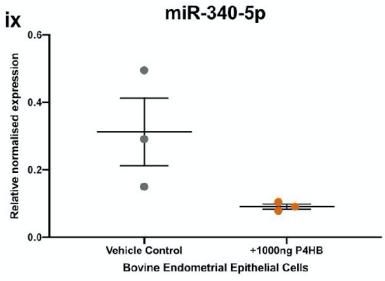

iv
miR-151a-5p
Relative normalised expression
Vehicle Control
+1000ng P4HB
Bovine Endometrial Epithelial Cells

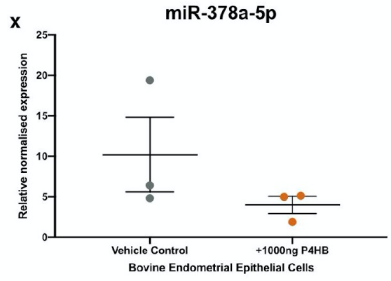

v
miR-185-5p
Relative normalised expression
*
Vehicle Control
+1000ng P4HB
Bovine Endometrial Epithelial Cells


vi
miR-188-5p
Relative normalised expression
Vehicle Control
+1000ng P4HB
Bovine Endometrial Epithelial Cells


vii
miR-324-5p
Relative normalised expression
Vehicle Control
+1000ng P4HB
Bovine Endometrial Epithelial Cells


viii
miR-331-3p
Relative normalised expression
Vehicle Control
+1000ng P4HB
Bovine Endometrial Epithelial Cells


ix
miR-340-5p
Relative normalised expression
Vehicle Control
+1000ng P4HB
Bovine Endometrial Epithelial Cells


x
miR-378a-5p
Relative normalised expression
Vehicle Control
+1000ng P4HB
Bovine Endometrial Epithelial Cells

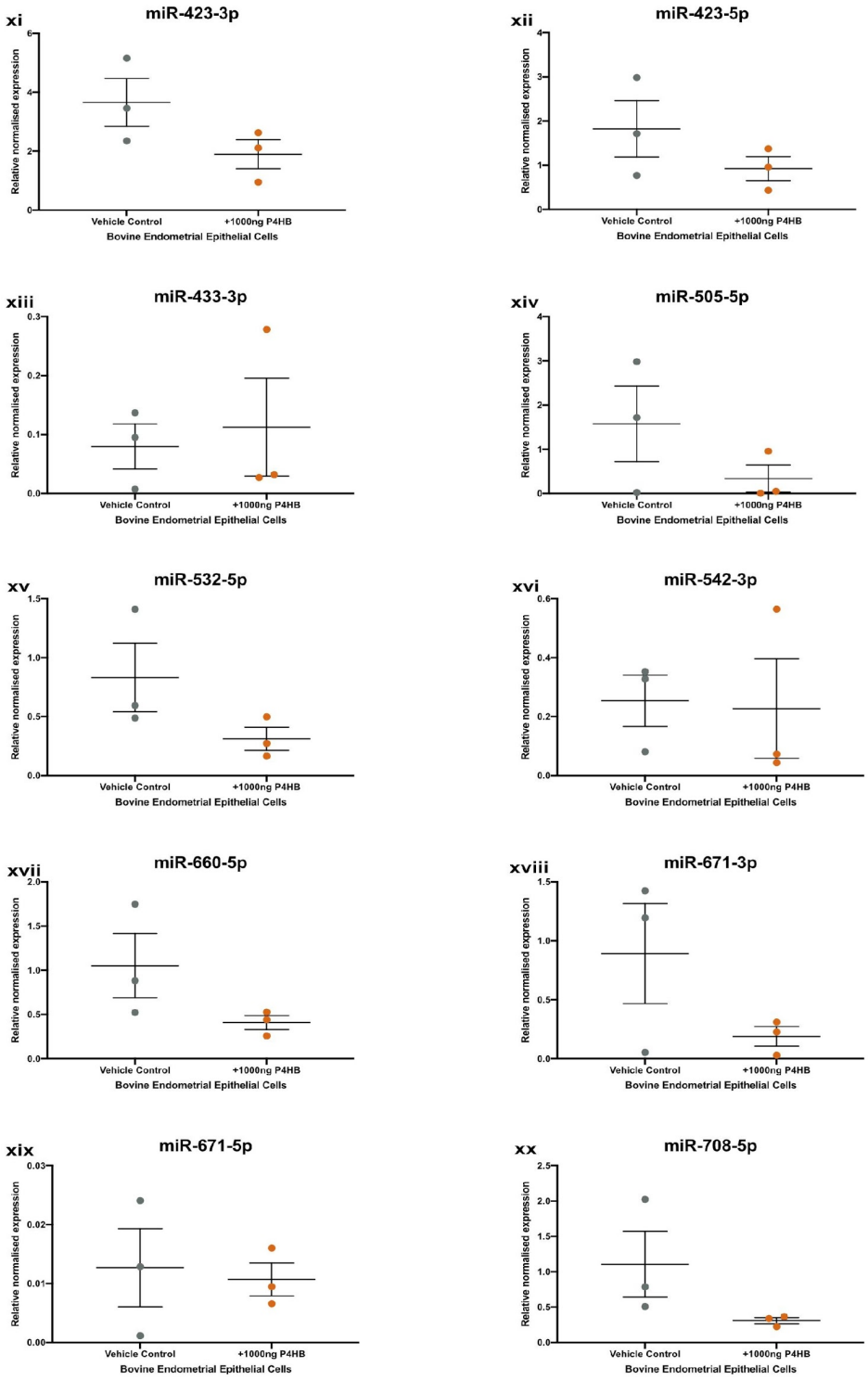


Figure 4.8 Expression of stem lineage miRNAs in bovine endometrial epithelial cells treated with bP4HB. *Expression of stem lineage miRNA (i) miR-28-3p, (ii) miR-127-3p, (iii) miR-151a-3p, (iv) miR-151a-5p, (v) miR-185-5p, (vi) miR-188-5p, (vii) miR-324-5p, (viii) miR-331-3p, (ix) miR-340-5p, (x) miR-378a-5p, (xi) miR-423-3p, (xii) miR-423-5p, (xiii) miR-433-3p, (xiv) miR-505-5p, (xv) miR-532-5p, (xvi) miR-542-3p, (xvii) miR-660-5p, (xviii) miR-671-3p, (xix) miR-671-5p and (xx) miR-708-5p in bovine endometrial epithelial cells treated with vehicle control (grey circle), or 1000ng/µl bP4HB (orange circle) for 24 hours. Significant differences in miRNA expression values determined when p≤0.05 are depicted by an asterisk (*).*

4.4.8 Reactome pathway results for predicted targets of miR-185-5p, a stem lineage miRNA significantly downregulated in bovine endometrial epithelial cells treated with bP4HB

Of the 25 most significant pathways for predicted targets of stem lineage miRNAs significantly upregulated in bovine endometrial epithelial cells treated with bP4HB (ranked by p-value), 3 pathways were statistically significant (p-value ≤ 0.05), shown in bold in Table 4.4.

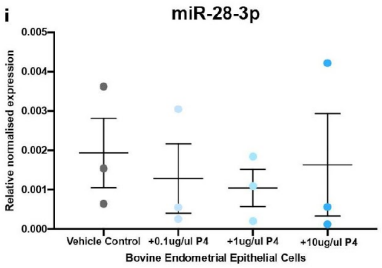
Table 4.4 Reactome pathway analysis results for miRNAs biologically regulated by bP4HB in bovine endometrial epithelial cells.

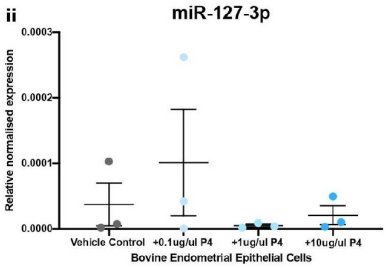
Pathway name	Entities				Reactions	
	found	ratio	p-value	FDR*	found	ratio
Protein-protein interactions at synapses	57/93	0.006	0.018	0.838	33/33	0.003
RUNX3 regulates CDKN1A transcription	8/8	5.54e-04	0.033	0.838	6/6	4.80e-04
SUMOylation of DNA methylation proteins	13/16	0.001	0.035	0.838	3/3	2.40e-04
Transcriptional activation of p53 responsive genes	6/6	4.16e-04	0.059	0.838	5/5	4.00e-04
Transcriptional activation of cell cycle inhibitor p21	6/6	4.16e-04	0.059	0.838	5/5	4.00e-04
RUNX3 regulates BCL2L11 (BIM) transcription	6/6	4.16e-04	0.059	0.838	2/2	1.60e-04
GRB7 events in ERBB2 signaling	6/6	4.16e-04	0.059	0.838	1/1	8.01e-05
Neurexins and neuroligins	36/60	0.004	0.064	0.838	19/19	0.002
Deactivation of the beta-catenin transactivating complex	27/44	0.003	0.079	0.838	14/14	0.001
RUNX3 regulates WNT signaling	8/10	6.93e-04	0.091	0.838	5/5	4.00e-04
RAB GEFs exchange GTP for GDP on RABs	52/94	0.007	0.095	0.838	21/21	0.002
WNT ligand biogenesis and trafficking	18/28	0.002	0.097	0.838	12/12	9.61e-04
RHO GTPases activate KTN1	9/12	8.31e-04	0.103	0.838	2/2	1.60e-04
Caspase-mediated cleavage of cytoskeletal proteins	9/12	8.31e-04	0.103	0.838	9/10	8.01e-04
Loss of Function of TGFBR1 in cancer	6/7	4.85e-04	0.104	0.838	2/2	1.60e-04
Post-transcriptional silencing by small RNAs	6/7	4.85e-04	0.104	0.838	2/3	2.40e-04
Synaptic adhesion-like molecules	15/23	0.002	0.111	0.838	8/8	6.41e-04
Regulation of TP53 Expression	4/4	2.77e-04	0.112	0.838	5/5	4.00e-04
Voltage gated Potassium channels	26/44	0.003	0.114	0.838	1/1	8.01e-05
Intra-Golgi traffic	28/48	0.003	0.116	0.838	10/10	8.01e-04
Signaling by TGF-beta receptor complex in cancer	7/9	6.24e-04	0.121	0.838	7/7	5.61e-04
Adherens junctions interactions	21/35	0.002	0.129	0.838	16/16	0.001
RUNX1 interacts with co-factors whose precise effect on RUNX1 targets is not known	23/39	0.003	0.132	0.838	5/5	4.00e-04
Receptor-type tyrosine-protein phosphatases	13/20	0.001	0.133	0.838	6 / 6	4.80e-04
O-glycosylation of TSR domain-containing proteins	24/41	0.003	0.134	0.838	2 / 2	1.60e-04

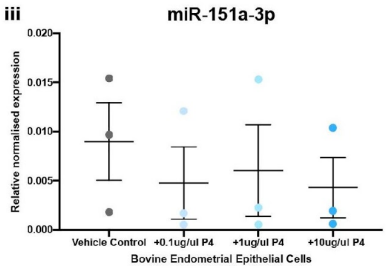
Reactome pathway results for 3 stem lineage miRNAs significantly upregulated in bovine endometrial epithelial cells treated with 1000ng/µl bP4HB for 24 hours. 'Entities - found' refers to the number of curated molecules within the named pathway found in the submitted dataset. 'Entities – ratio' refers to the number of curated molecules represented in the pathway compared to the entire database. 'Reactions – found' refers to the number of reactions within the named pathway represented by molecules in the submitted dataset. 'Reactions – ratio' refers to the number of Reactome reactions represented by the reactions in this pathway. False Discovery Rate (FDR) is calculated using the Benjamini-Hochberg method.

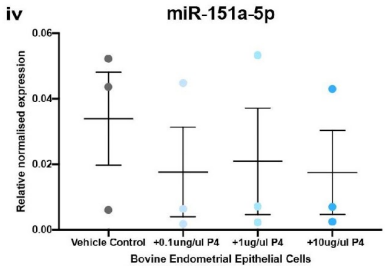
4.4.9 Expression analysis of bovine endometrial epithelial cells treated with P4

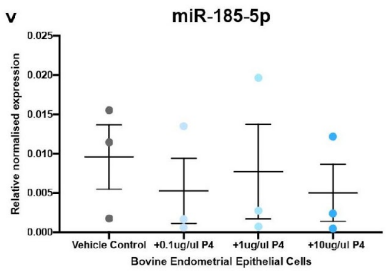
miR-505-5p showed a significant increase in expression in bovine endometrial epithelial cells treated with 10µg/mL P4 in comparison to 0.1µg/mL P4. miRNA expression levels are shown in Figure 4.9.

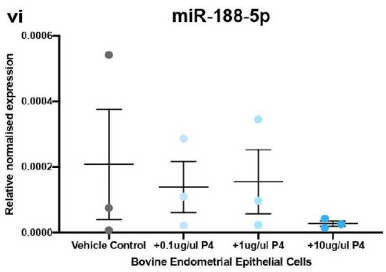

i
miR-28-3p
Relative normalised expression
0.005
0.004
0.003
0.002
0.001
0.000
Vehicle Control
+0.1ug/ul P4
+1ug/ul P4
+10ug/ul P4
Bovine Endometrial Epithelial Cells

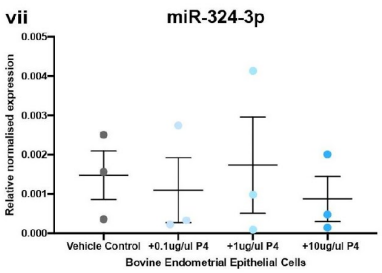

ii
miR-127-3p
Relative normalised expression
0.0003
0.0002
0.0001
0.0000
Vehicle Control
+0.1ug/ul P4
+1ug/ul P4
+10ug/ul P4
Bovine Endometrial Epithelial Cells

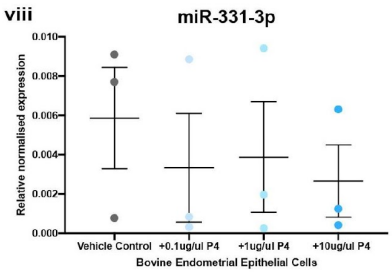

iii
miR-151a-3p
Relative normalised expression
0.020
0.015
0.010
0.005
0.000
Vehicle Control
+0.1ug/ul P4
+1ug/ul P4
+10ug/ul P4
Bovine Endometrial Epithelial Cells

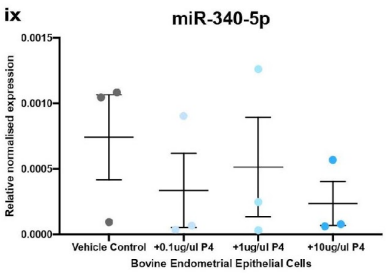

iv
miR-151a-5p
Relative normalised expression
0.06
0.04
0.02
0.00
Vehicle Control
+0.1ug/ul P4
+1ug/ul P4
+10ug/ul P4
Bovine Endometrial Epithelial Cells

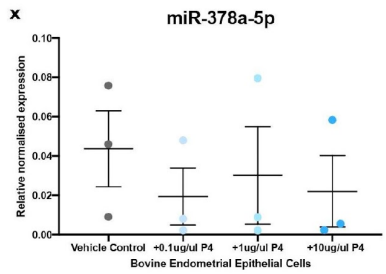

v
miR-185-5p
Relative normalised expression
0.025
0.020
0.015
0.010
0.005
0.000
Vehicle Control
+0.1ug/ul P4
+1ug/ul P4
+10ug/ul P4
Bovine Endometrial Epithelial Cells


vi
miR-188-5p
Relative normalised expression
0.0006
0.0004
0.0002
0.0000
Vehicle Control
+0.1ug/ul P4
+1ug/ul P4
+10ug/ul P4
Bovine Endometrial Epithelial Cells


vii
miR-324-3p
Relative normalised expression
0.005
0.004
0.003
0.002
0.001
0.000
Vehicle Control
+0.1ug/ul P4
+1ug/ul P4
+10ug/ul P4
Bovine Endometrial Epithelial Cells


viii
miR-331-3p
Relative normalised expression
0.010
0.008
0.006
0.004
0.002
0.000
Vehicle Control
+0.1ug/ul P4
+1ug/ul P4
+10ug/ul P4
Bovine Endometrial Epithelial Cells


ix
miR-340-5p
Relative normalised expression
0.0015
0.0010
0.0005
0.0000
Vehicle Control
+0.1ug/ul P4
+1ug/ul P4
+10ug/ul P4
Bovine Endometrial Epithelial Cells


x
miR-378a-5p
Relative normalised expression
0.10
0.08
0.06
0.04
0.02
0.00
Vehicle Control
+0.1ug/ul P4
+1ug/ul P4
+10ug/ul P4
Bovine Endometrial Epithelial Cells

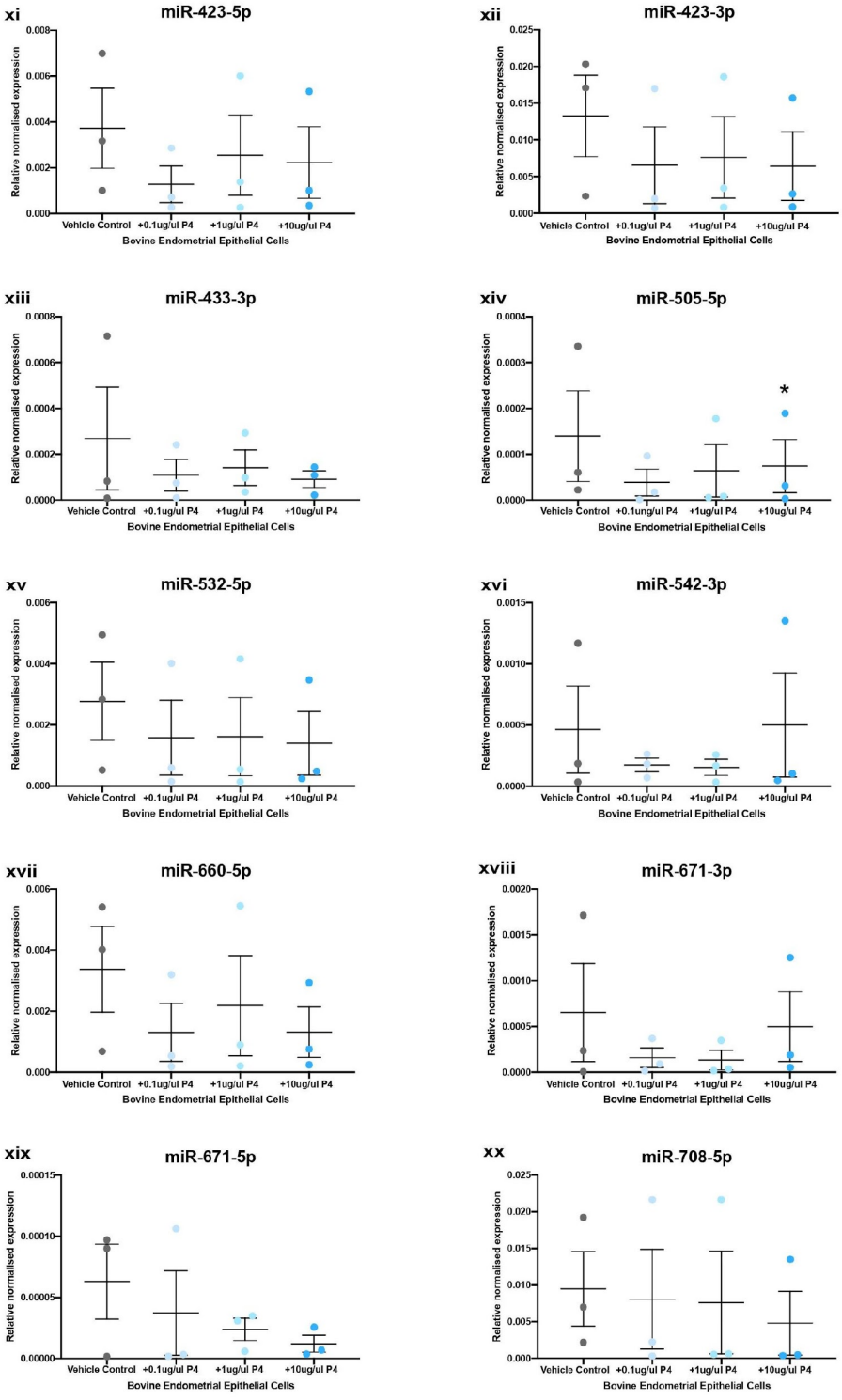


Figure 4.9 Expression of stem lineage miRNAs in bovine endometrial epithelial cells treated with P4. *Expression of stem lineage miRNA (i) miR-28-3p, (ii) miR-127-3p, (iii) miR-151a-3p, (iv) miR-151a-5p, (v) miR-185-5p, (vi) miR-188-5p, (vii) miR-324-5p, (viii) miR-331-3p, (ix) miR-340-5p, (x) miR-378a-5p, (xi) miR-423-3p, (xii) miR-423-5p, (xiii) miR-433-3p, (xiv) miR-505-5p, (xv) miR-532-5p, (xvi) miR-542-3p, (xvii) miR-660-5p, (xviii) miR-671-3p, (xix) miR-671-5p and (xx) miR-708-5p in bovine endometrial epithelial cells treated with vehicle control (grey circle), 0.1µg/mL (light blue), 1.0µg/mL (medium blue) or 10µg/mL P4 (dark blue circle) for 24 hours. Significant differences in miRNA expression values determined when p≤0.05 are depicted by an asterisk (*).*

4.4.10 Reactome pathway results for predicted targets of miR-505-5p, a stem lineage miRNA significantly downregulated in bovine endometrial epithelial cells treated with P4

Of the 25 most significant pathways for predicted targets of miR-505-5p (ranked by p-value), 16 were statistically significant (p-value ≤ 0.05), shown in bold in Table 4.5.

Table 4.5 Reactome pathway analysis results for miRNAs biologically regulated by P4 in bovine endometrial epithelial cells.

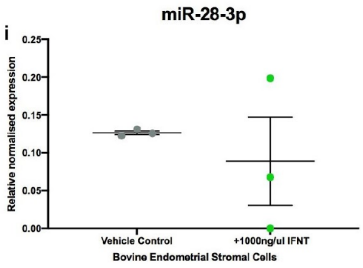
Pathway name	Entities				Reactions	
	found	ratio	p-value	FDR*	found	ratio
SUMOylation of DNA methylation proteins	12/16	0.001	0.009	0.859	3/3	2.40e-04
Semaphorin interactions	36/70	0.005	0.01	0.859	38/40	0.003
Potassium Channels	49/103	0.007	0.011	0.859	17/19	0.002
Formation of a pool of free 40S subunits	50/106	0.007	0.012	0.859	2/2	1.60e-04
VEGF ligand-receptor interactions	7/8	5.54e-04	0.02	0.859	4/4	3.20e-04
VEGF binds to VEGFR leading to receptor dimerization	7/8	5.54e-04	0.02	0.859	3/3	2.40e-04
CRMPs in Sema3A signaling	12/18	0.001	0.021	0.859	5/5	4.00e-04
GTP hydrolysis and joining of the 60S ribosomal subunit	53/120	0.008	0.03	0.859	3/3	2.40e-04
Sema4D induced cell migration and growth-cone collapse	14/24	0.002	0.035	0.859	6/6	4.80e-04
SUMO E3 ligases SUMOylate target proteins	76/183	0.013	0.038	0.859	130/130	0.01
SUMOylation of DNA damage response and repair proteins	37/81	0.006	0.041	0.859	24/24	0.002
L13a-mediated translational silencing of Ceruloplasmin expression	52/120	0.008	0.041	0.859	3/3	2.40e-04
SUMOylation	79/192	0.013	0.041	0.859	139/139	0.011
SUMOylation of transcription cofactors	22/44	0.003	0.046	0.859	27/27	0.002
Voltage gated Potassium channels	22/44	0.003	0.046	0.859	1/1	8.01e-05
Neurophilin interactions with VEGF and VEGFR	4/4	2.77e-04	0.047	0.859	2/2	1.60e-04
Eukaryotic Translation Initiation	55/130	0.009	0.052	0.859	18/21	0.002
Cap-dependent Translation Initiation	55/130	0.009	0.052	0.859	15/18	0.001
Adherens junctions interactions	18/35	0.002	0.053	0.859	15/16	0.001
Peptide chain elongation	42/97	0.007	0.061	0.859	4/5	4.00e-04
Nonsense-Mediated Decay (NMD)	52/124	0.009	0.065	0.859	6/6	4.80e-04
Nonsense Mediated Decay (NMD) enhanced by the Exon Junction Complex (EJC)	52/124	0.009	0.065	0.859	5/5	4.00e-04
Regulation of MECP2 expression and activity	19/39	0.003	0.073	0.859	13/14	0.001
RUNX1 interacts with co-factors whose precise effect on RUNX1 targets is not known	19/39	0.003	0.073	0.859	4/5	4.00e-04
Response of EIF2AK4 (GCN2) to amino acid deficiency	48 /115	0.008	0.078	0.859	10/16	0.001

Reactome pathway results miR-505-5p, a stem lineage miRNA significantly downregulated in bovine endometrial epithelial cells treated with 10µg/mL P4 for 24 hours in comparison to 0.1µg/mL P4. 'Entities - found' refers to the number of curated molecules within the named pathway found in the submitted dataset. 'Entities – ratio' refers to the number of curated molecules represented in the pathway compared to the entire database. 'Reactions – found' refers to the number of reactions within the named pathway represented by molecules in the submitted dataset. 'Reactions – ratio' refers to the number of Reactome reactions represented by the reactions in this pathway. False Discovery Rate (FDR) is calculated using the Benjamini-Hochberg method.

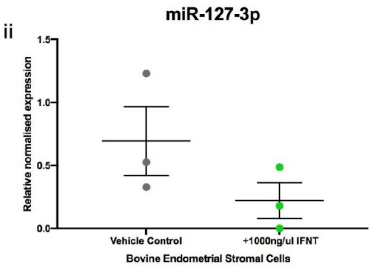
4.4.11 Expression analysis of bovine endometrial stromal cells treated with oIFNT

No significant differences were found in stem lineage miRNA expression in bovine endometrial stromal cells treated with oIFNT compared to vehicle control. miRNA expression levels are depicted in Figure 4.10.

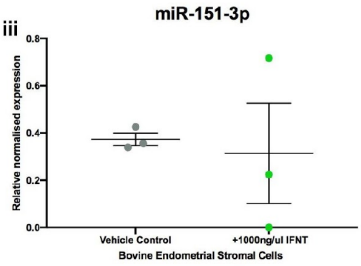
i



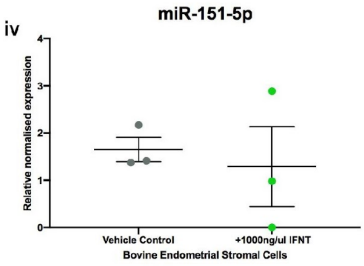
ii



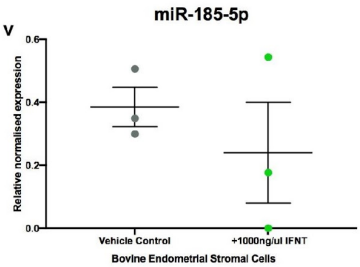
iii



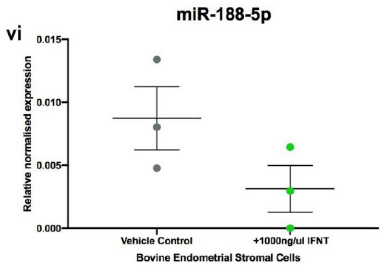
iv



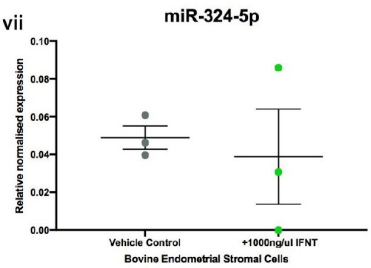
v



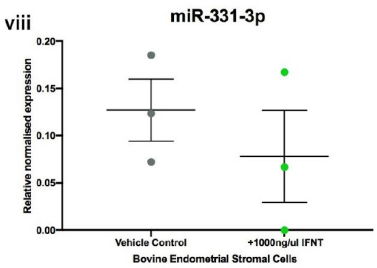
vi



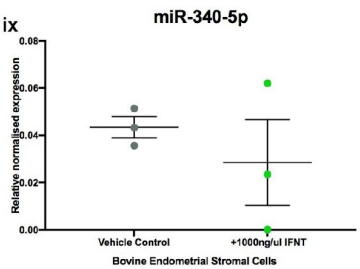
vii



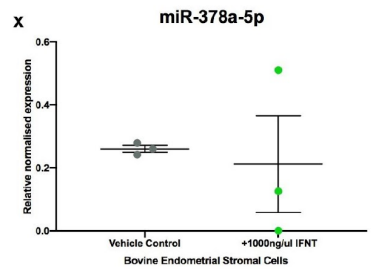
viii



ix



x



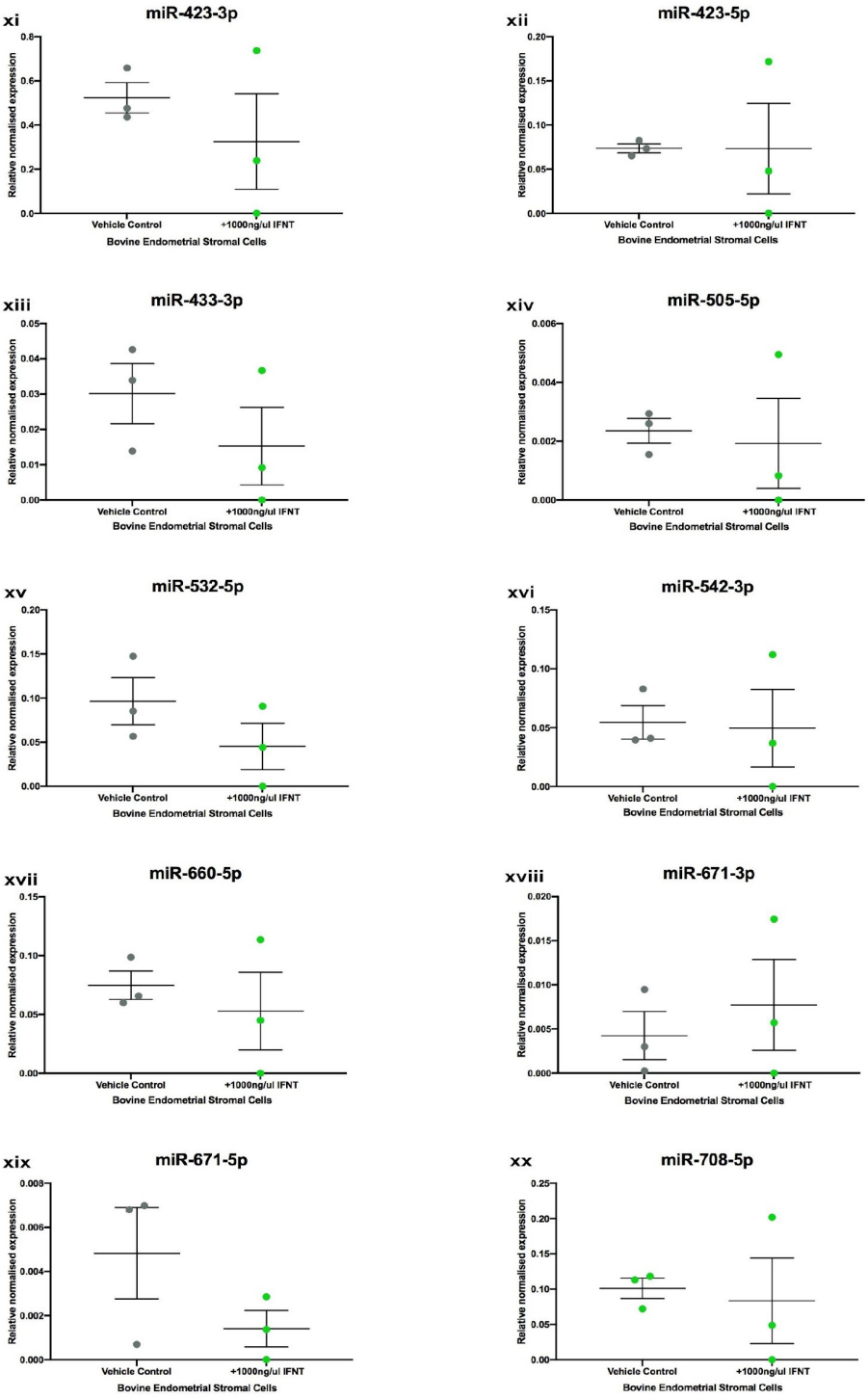
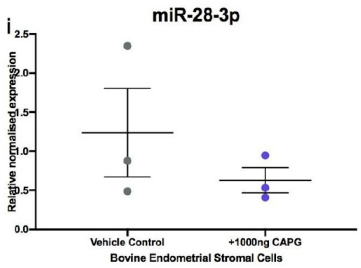


Figure 4.10 ***Expression of stem lineage miRNAs in bovine endometrial stromal cells treated with oIFNT.*** *Expression of stem lineage miRNA (i) miR-28-3p, (ii) miR-127-3p, (iii) miR-151a-3p, (iv) miR-151a-5p, (v) miR-185-5p, (vi) miR-188-5p, (vii) miR-324-5p, (viii) miR-331-3p, (ix) miR-340-5p, (x) miR-378a-5p, (xi) miR-423-3p, (xii) miR-423-5p, (xiii) miR-433-3p, (xiv) miR-505-5p, (xv) miR-532-5p, (xvi) miR-542-3p, (xvii) miR-660-5p, (xviii) miR-671-3p, (xix) miR-671-5p and (xx) miR-708-5p in bovine endometrial stromal cells treated with vehicle control (grey circle), or 1000ng/µl oIFNT (green circle) for 24 hours. No significant difference in miRNA expression was found*

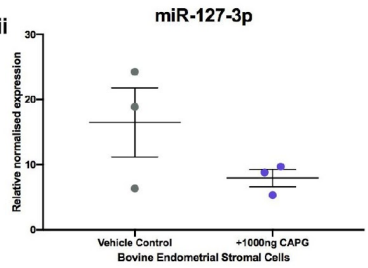
4.4.12 Expression analysis of bovine endometrial stromal cells treated with bCAPG

miR-660-5p, a paralogue of miR-188 originating on the Eutherian stem lineage, showed a significant decrease in expression in bovine endometrial stromal cells treated with 1000ng/µl bCAPG compared to vehicle control. miRNA expression levels are shown in Figure 4.11.

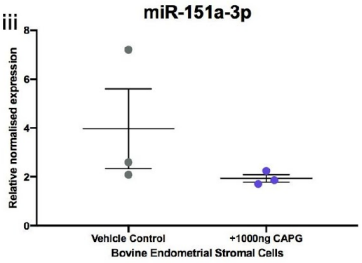
i



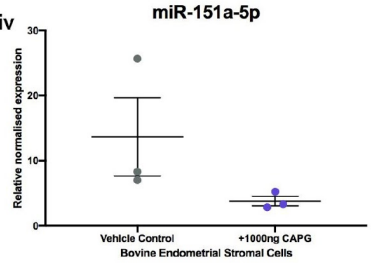
ii



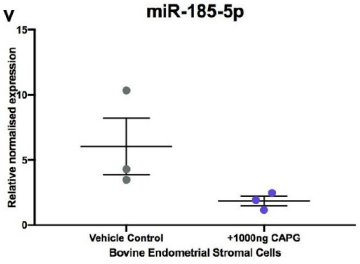
iii



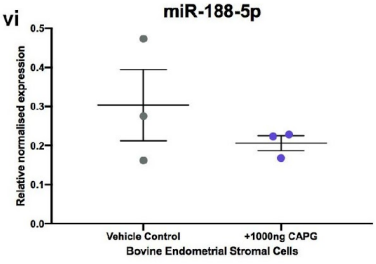
iv



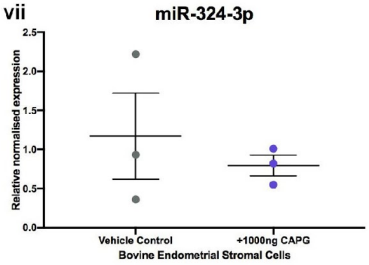
v



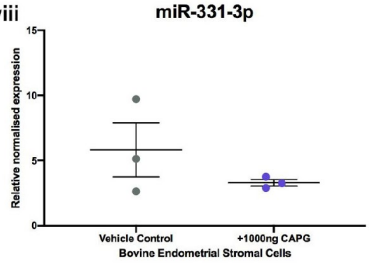
vi



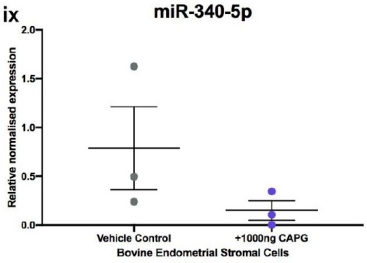
vii



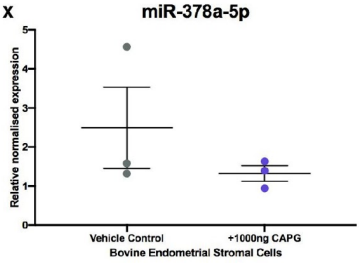
viii



ix



x



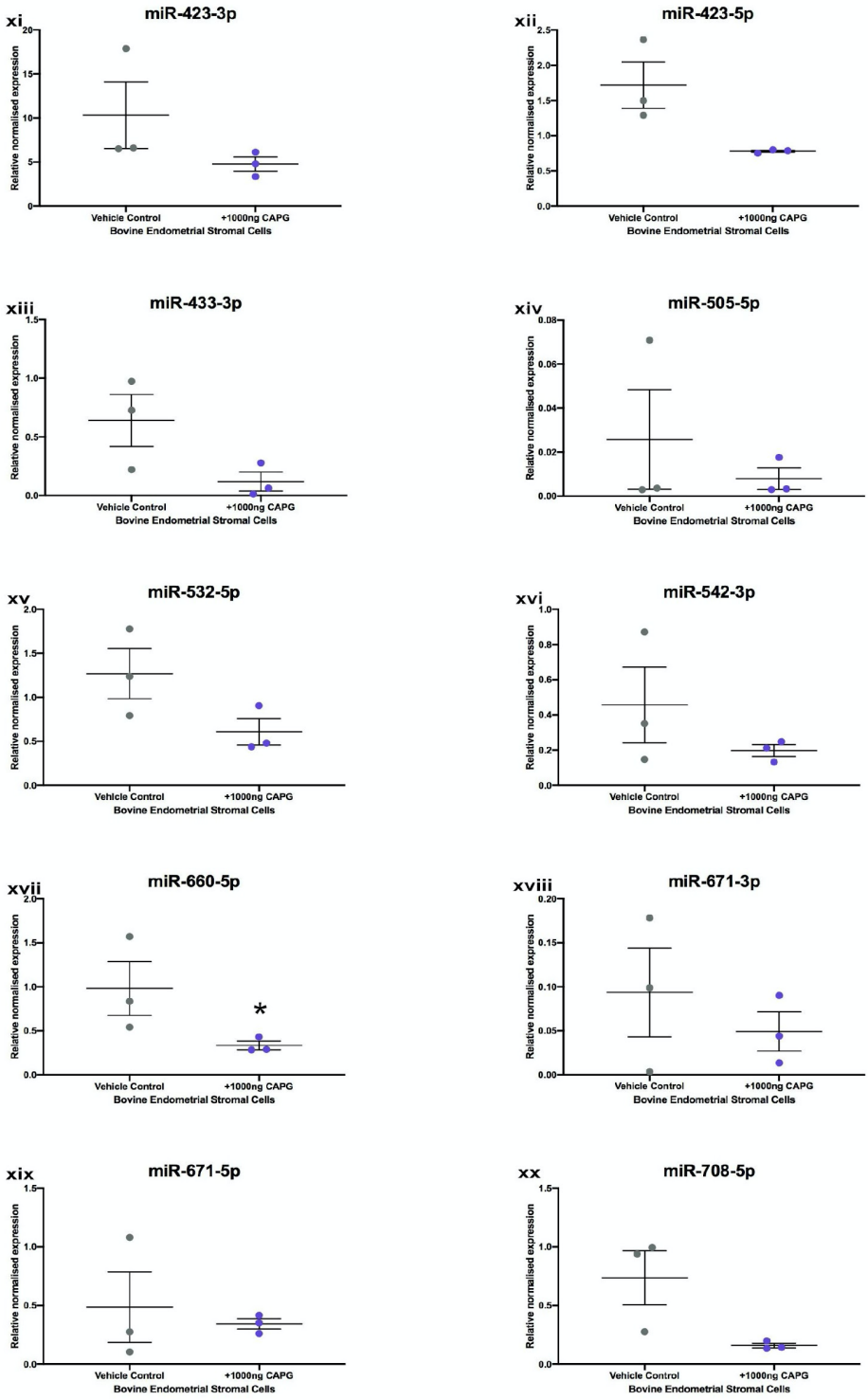


Figure 4.11 ***Expression of stem lineage miRNAs in bovine endometrial stromal cells treated with bCAPG.*** *Expression of stem lineage miRNA (i) miR-28-3p, (ii) miR-127-3p, (iii) miR-151a-3p, (iv) miR-151a-5p, (v) miR-185-5p, (vi) miR-188-5p, (vii) miR-324-5p, (viii) miR-331-3p, (ix) miR-340-5p, (x) miR-378a-5p, (xi) miR-423-3p, (xii) miR-423-5p, (xiii) miR-433-3p, (xiv) miR-505-5p, (xv) miR-532-5p, (xvi) miR-542-3p, (xvii) miR-660-5p, (xviii) miR-671-3p, (xix) miR-671-5p and (xx) miR-708-5p in bovine endometrial stromal cells treated with vehicle control (grey circle), or 1000ng/µl bCAPG (purple circle) for 24 hours. Significant differences in miRNA expression values determined when p≤0.05 are depicted by an asterisk (*).*

4.4.13 Reactome pathway results for predicted targets of miR-660-5p, a stem lineage miRNA significantly downregulated in bovine endometrial stromal cells treated with bCAPG

Of the 25 most significant pathways for predicted targets of miR-660-5p (ranked by p-value), 10 were statistically significant (p-value ≤ 0.05), shown in bold in Table 4.6.

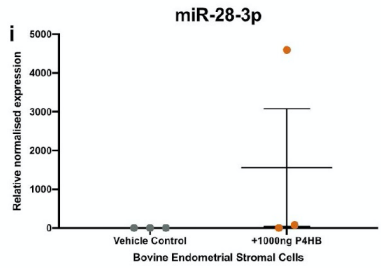
Table 4.6 Reactome pathway analysis results for miRNAs biologically regulated by bCAPG in bovine endometrial stromal cells.

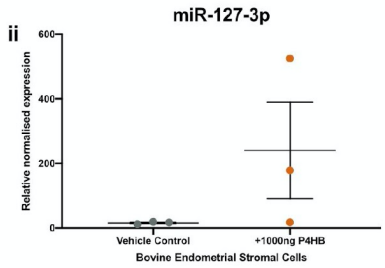
Pathway name	Entities				Reactions	
	found	ratio	p-value	FDR*	found	ratio
Intra-Golgi and retrograde Golgi-to-ER traffic	71/218	0.015	0.01	0.789	44/48	0.004
trans-Golgi Network Vesicle Budding	30/80	0.006	0.016	0.789	17/19	0.002
MECP2 regulates transcription of genes involved in GABA signaling	5/6	4.16e-04	0.017	0.789	4/4	3.20e-04
Interaction between L1 and Ankyrins	15/33	0.002	0.018	0.789	4/4	3.20e-04
Downregulation of ERBB2:ERBB3 signaling	9/16	0.001	0.018	0.789	8/8	6.41e-04
Elastic fibre formation	19/46	0.003	0.021	0.789	13/17	0.001
Golgi Associated Vesicle Biogenesis	23/61	0.004	0.029	0.789	6/7	5.61e-04
Downregulation of ERBB2 signaling	15/36	0.002	0.034	0.789	14/14	0.001
Diseases of Mismatch Repair (MMR)	4/5	3.46e-04	0.035	0.789	6/6	4.80e-04
MECP2 regulates transcription factors	6/10	6.93e-04	0.038	0.789	8/8	6.41e-04
Long-term potentiation	13/31	0.002	0.044	0.789	7/7	5.61e-04
Downregulation of TGF-beta receptor signaling	12/28	0.002	0.045	0.789	22/28	0.002
ERBB2 Regulates Cell Motility	9/19	0.001	0.046	0.789	2/2	1.60e-04
Molecules associated with elastic fibres	15/38	0.003	0.05	0.789	7/10	8.01e-04
Membrane Trafficking	183/665	0.046	0.05	0.789	172/218	0.017
Intra-Golgi traffic	18/48	0.003	0.051	0.789	10/10	8.01e-04
Retrograde transport at the Trans-Golgi-Network	20/55	0.004	0.054	0.789	17/20	0.002
MECP2 regulates neuronal receptors and channels	13/32	0.002	0.054	0.789	26/26	0.002
Downregulation of ERBB4 signaling	6/11	7.62e-04	0.055	0.789	5/5	4.00e-04
Nef mediated downregulation of MHC class I complex cell surface expression	6/11	7.62e-04	0.055	0.789	4/4	3.20e-04
Estrogen-dependent nuclear events downstream of ESR-membrane signaling	12/29	0.002	0.056	0.789	11/12	9.61e-04
Class I peroxisomal membrane protein import	9/20	0.001	0.06	0.789	6/6	4.80e-04
Activation of NOXA and translocation to mitochondria	4/6	4.16e-04	0.061	0.789	5/5	4.00e-04
GRB7 events in ERBB2 signaling	4/6	4.16e-04	0.061	0.789	1/1	8.01e-05
L1CAM interactions	41/130	0.009	0.061	0.789	38/54	0.004

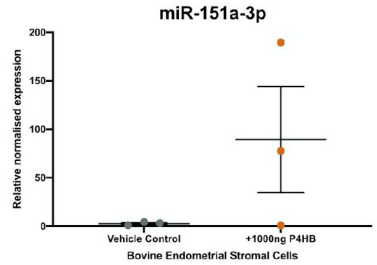
Reactome pathway results for miR-660-5p, a stem lineage miRNA significantly downregulated in bovine endometrial stromall cells treated with 1000ng/µl bCAPG for 24 hours. 'Entities - found' refers to the number of curated molecules within the named pathway found in the submitted dataset. 'Entities – ratio' refers to the number of curated molecules represented in the pathway compared to the entire database. 'Reactions – found' refers to the number of reactions within the named pathway represented by molecules in the submitted dataset. 'Reactions – ratio' refers to the number of Reactome reactions represented by the reactions in this pathway. False Discovery Rate (FDR) is calculated using the Benjamini-Hochberg method.

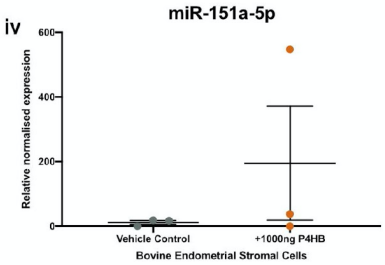
4.4.14 Expression analysis of bovine endometrial stromal cells treated with bP4HB

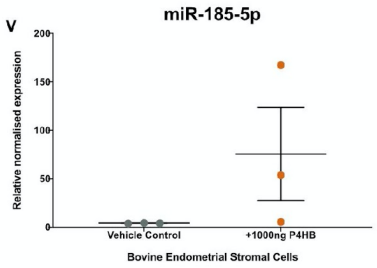
miR-188-5p showed a significant increase in expression in bovine endometrial stromal cells treated with 1000ng/µl bP4HB compared to vehicle control. miRNA expression levels are shown in Figure 4.12.

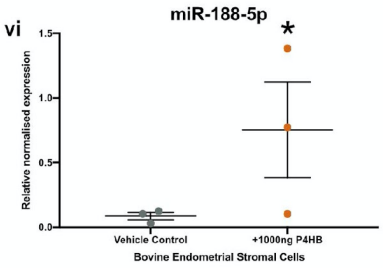

i
miR-28-3p
Relative normalised expression
5000
4000
3000
2000
1000
0
Vehicle Control
+1000ng P4HB
Bovine Endometrial Stromal Cells

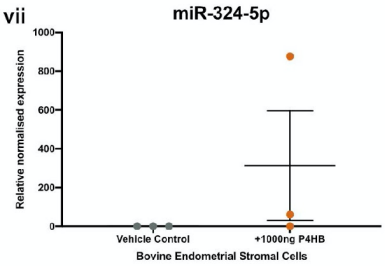

ii
miR-127-3p
Relative normalised expression
600
400
200
0
Vehicle Control
+1000ng P4HB
Bovine Endometrial Stromal Cells

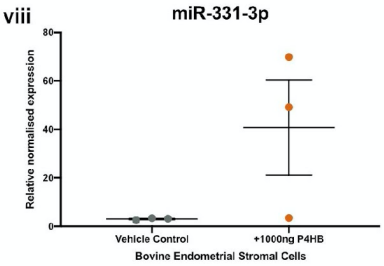

miR-151a-3p
Relative normalised expression
200
150
100
50
0
Vehicle Control
+1000ng P4HB
Bovine Endometrial Stromal Cells

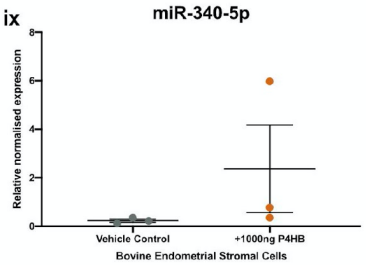

iv
miR-151a-5p
Relative normalised expression
600
400
200
0
Vehicle Control
+1000ng P4HB
Bovine Endometrial Stromal Cells

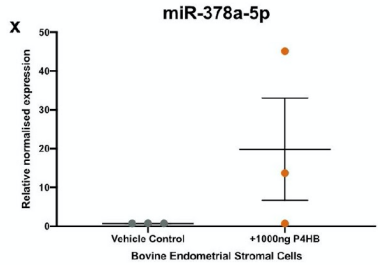

v
miR-185-5p
Relative normalised expression
200
150
100
50
0
Vehicle Control
+1000ng P4HB
Bovine Endometrial Stromal Cells


vi
miR-188-5p
*
Relative normalised expression
1.5
1.0
0.5
0.0
Vehicle Control
+1000ng P4HB
Bovine Endometrial Stromal Cells


vii
miR-324-5p
Relative normalised expression
1000
800
600
400
200
0
Vehicle Control
+1000ng P4HB
Bovine Endometrial Stromal Cells


viii
miR-331-3p
Relative normalised expression
80
60
40
20
0
Vehicle Control
+1000ng P4HB
Bovine Endometrial Stromal Cells


ix
miR-340-5p
Relative normalised expression
8
6
4
2
0
Vehicle Control
+1000ng P4HB
Bovine Endometrial Stromal Cells


x
miR-378a-5p
Relative normalised expression
50
40
30
20
10
0
Vehicle Control
+1000ng P4HB
Bovine Endometrial Stromal Cells

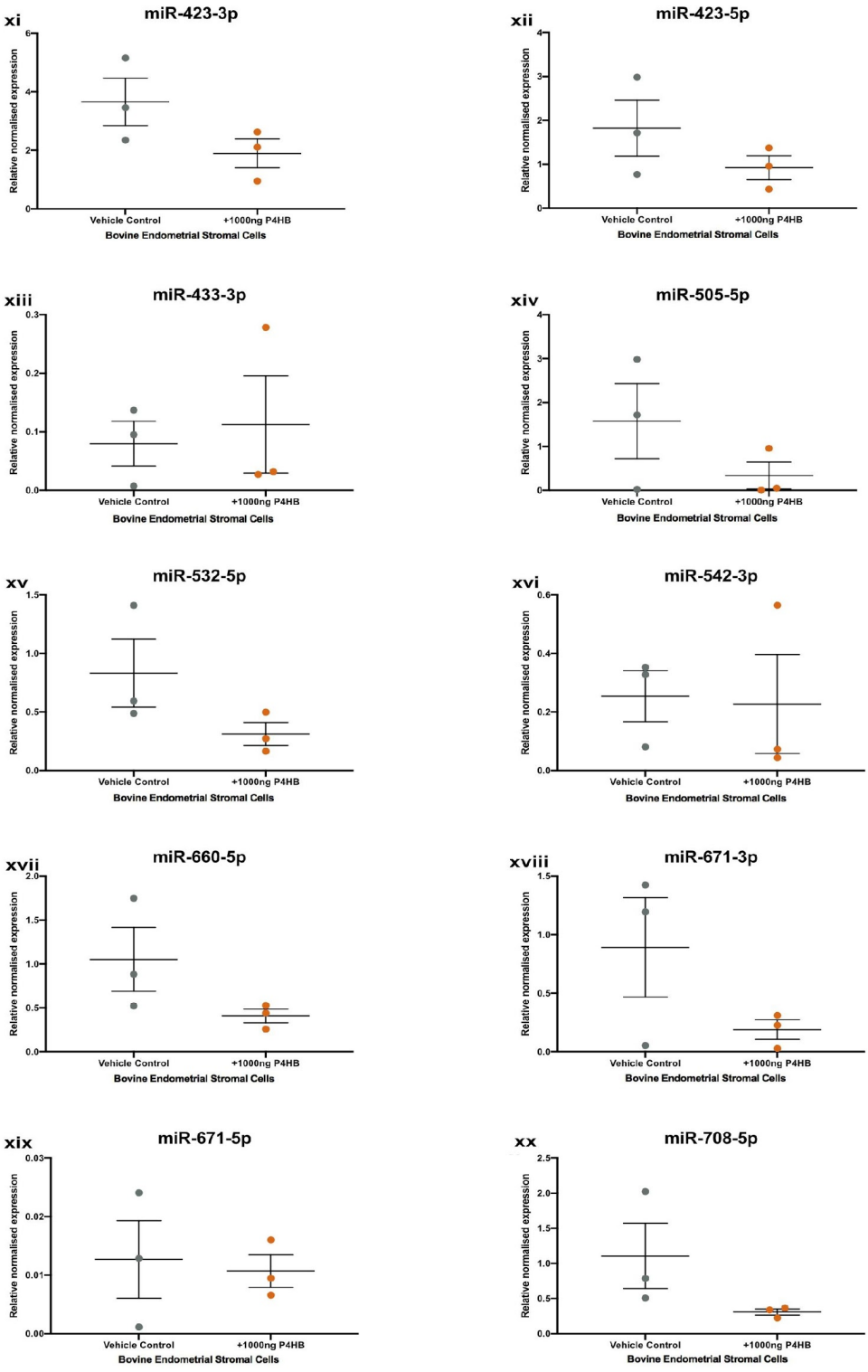


Figure 4.12 Expression of stem lineage miRNAs in bovine endometrial stromal cells treated with bP4HB. *Expression of stem lineage miRNA (i) miR-28-3p, (ii) miR-127-3p, (iii) miR-151a-3p, (iv) miR-151a-5p, (v) miR-185-5p, (vi) miR-188-5p, (vii) miR-324-5p, (viii) miR-331-3p, (ix) miR-340-5p, (x) miR-378a-5p, (xi) miR-423-3p, (xii) miR-423-5p, (xiii) miR-433-3p, (xiv) miR-505-5p, (xv) miR-532-5p, (xvi) miR-542-3p, (xvii) miR-660-5p, (xviii) miR-671-3p, (xix) miR-671-5p and (xx) miR-708-5p in bovine endometrial stromal cells treated with vehicle control (grey circle), or 1000ng/µl bP4HB (purple circle) for 24 hours. Significant differences in miRNA expression values determined when p≤0.05 are depicted by an asterisk (*).*

4.4.15 Reactome pathway results for predicted targets of miR-188-5p, a stem lineage miRNA significantly upregulated in bovine endometrial stromal cells treated with bP4HB

Of the 25 most significant pathways for predicted targets of miR-188-5p (ranked by p-value), 19 were statistically significant (p-value ≤ 0.05), shown in bold in Table 4.7.

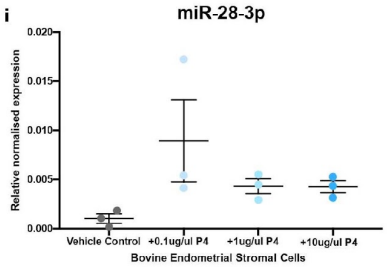
Table 4.7 Reactome pathway analysis results for miRNAs biologically regulated by bP4HB in bovine endometrial stromal cells.

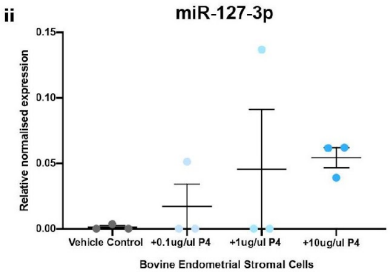
Pathway name	Entities				Reactions	
	found	ratio	p-value	FDR*	found	ratio
Phase 4 - resting membrane potential	12/20	0.001	0.012	0.717	2/2	1.60e-04
Loss of MECP2 binding ability to 5mC-DNA	5/5	3.46e-04	0.014	0.717	2/2	1.60e-04
NrCAM interactions	6/7	4.85e-04	0.015	0.717	4/4	3.20e-04
SMAD2/SMAD3:SMAD4 heterotrimer regulates transcription	19/39	0.003	0.016	0.717	17/17	0.001
Activation of NOXA and translocation to mitochondria	5/6	4.16e-04	0.028	0.717	5/5	4.00e-04
Downregulation of ERBB2 signaling	17/36	0.002	0.029	0.717	14/14	0.001
Long-term potentiation	15/31	0.002	0.032	0.717	7/7	5.61e-04
Inhibition of replication initiation of damaged DNA by RB1/E2F1	8/13	9.01e-04	0.033	0.717	3/3	2.40e-04
ERBB2 Activates PTK6 Signaling	10/18	0.001	0.033	0.717	2/2	1.60e-04
Intra-Golgi traffic	21/48	0.003	0.034	0.717	10/10	8.01e-04
Signaling by TGF-beta family members	43/114	0.008	0.035	0.717	93/107	0.009
Post NMDA receptor activation events	37/96	0.007	0.036	0.717	28/39	0.003
GRB2 events in ERBB2 signaling	11/21	0.001	0.038	0.717	4/4	3.20e-04
Potassium Channels	39/103	0.007	0.041	0.717	13/19	0.002
Retrograde transport at the Trans-Golgi-Network	23/55	0.004	0.042	0.717	20/20	0.002
Signaling by TGF-beta receptor complex	33/85	0.006	0.042	0.717	71/82	0.007
ERBB2 Regulates cell motility	10/19	0.001	0.045	0.717	2/2	1.60e-04
Voltage gated Potassium channels	19/44	0.003	0.046	0.717	1/1	8.01e-05
E2F mediated regulation of DNA replication	11/22	0.002	0.05	0.717	4/4	3.20e-04
Regulation of MECP2 expression and activity	17/39	0.003	0.053	0.717	14/14	0.001
SMAD4 MH2 Domain Mutants in cancer	3/3	2.08e-04	0.053	0.717	1/1	8.01e-05
SMAD2/3 MH2 Domain Mutants in cancer	3/3	2.08e-04	0.053	0.717	1/1	8.01e-05
Loss of Function of SMAD4 in Cancer	3/3	2.08e-04	0.053	0.717	1/1	8.01e-05
TFAP2 (AP-2) family regulates transcription of other transcription factors	4/5	3.46e-04	0.054	0.717	2/2	1.60e-04
MASTL Facilitates Mitotic Progression	7/12	8.31e-04	0.055	0.717	4/4	3.20e-04

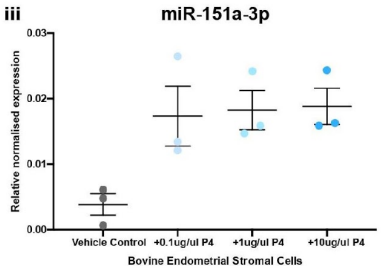
Reactome pathway results for miR-188-5p, a stem lineage miRNA significantly upregulated in bovine endometrial stromall cells treated with 1000ng/µl bP4HB for 24 hours. 'Entities - found' refers to the number of curated molecules within the named pathway found in the submitted dataset. 'Entities – ratio' refers to the number of curated molecules represented in the pathway compared to the entire database. 'Reactions – found' refers to the number of reactions within the named pathway represented by molecules in the submitted dataset. 'Reactions – ratio' refers to the number of Reactome reactions represented by the reactions in this pathway. False Discovery Rate (FDR) is calculated using the Benjamini-Hochberg method.

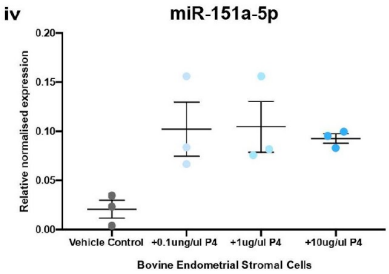
4.4.16 Expression analysis of bovine endometrial stromal cells treated with P4

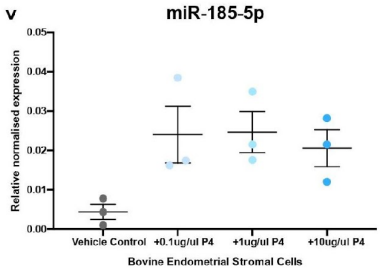
miR-505-5p showed a significant increase in expression in bovine endometrial epithelial cells treated with 10µg/mL P4 in comparison to vehicle control. miRNA expression levels are shown in Figure 4.13.

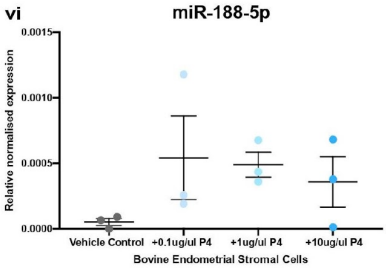

i
miR-28-3p
Relative normalised expression
0.020
0.015
0.010
0.005
0.000
Vehicle Control
+0.1ug/ul P4
+1ug/ul P4
+10ug/ul P4
Bovine Endometrial Stromal Cells

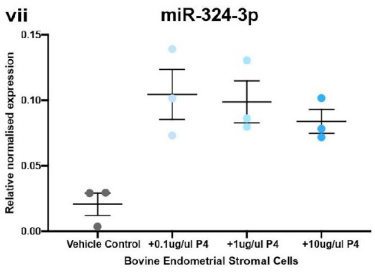

ii
miR-127-3p
Relative normalised expression
0.15
0.10
0.05
0.00
Vehicle Control
+0.1ug/ul P4
+1ug/ul P4
+10ug/ul P4
Bovine Endometrial Stromal Cells

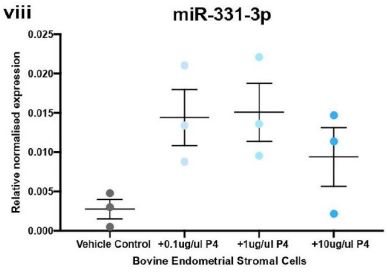

iii
miR-151a-3p
Relative normalised expression
0.03
0.02
0.01
0.00
Vehicle Control
+0.1ug/ul P4
+1ug/ul P4
+10ug/ul P4
Bovine Endometrial Stromal Cells

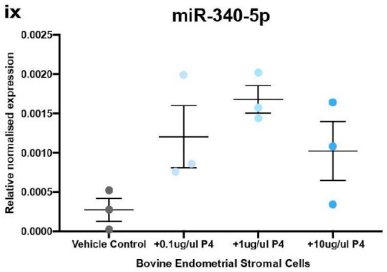

iv
miR-151a-5p
Relative normalised expression
0.20
0.15
0.10
0.05
0.00
Vehicle Control
+0.1ung/ul P4
+1ug/ul P4
+10ug/ul P4
Bovine Endometrial Stromal Cells

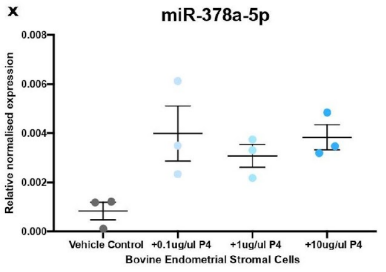

v
miR-185-5p
Relative normalised expression
0.05
0.04
0.03
0.02
0.01
0.00
Vehicle Control
+0.1ug/ul P4
+1ug/ul P4
+10ug/ul P4
Bovine Endometrial Stromal Cells


vi
miR-188-5p
Relative normalised expression
0.0015
0.0010
0.0005
0.0000
Vehicle Control
+0.1ug/ul P4
+1ug/ul P4
+10ug/ul P4
Bovine Endometrial Stromal Cells


vii
miR-324-3p
Relative normalised expression
0.15
0.10
0.05
0.00
Vehicle Control
+0.1ug/ul P4
+1ug/ul P4
+10ug/ul P4
Bovine Endometrial Stromal Cells


viii
miR-331-3p
Relative normalised expression
0.025
0.020
0.015
0.010
0.005
0.000
Vehicle Control
+0.1ug/ul P4
+1ug/ul P4
+10ug/ul P4
Bovine Endometrial Stromal Cells


ix
miR-340-5p
Relative normalised expression
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
Vehicle Control
+0.1ug/ul P4
+1ug/ul P4
+10ug/ul P4
Bovine Endometrial Stromal Cells


x
miR-378a-5p
Relative normalised expression
0.008
0.006
0.004
0.002
0.000
Vehicle Control
+0.1ug/ul P4
+1ug/ul P4
+10ug/ul P4
Bovine Endometrial Stromal Cells

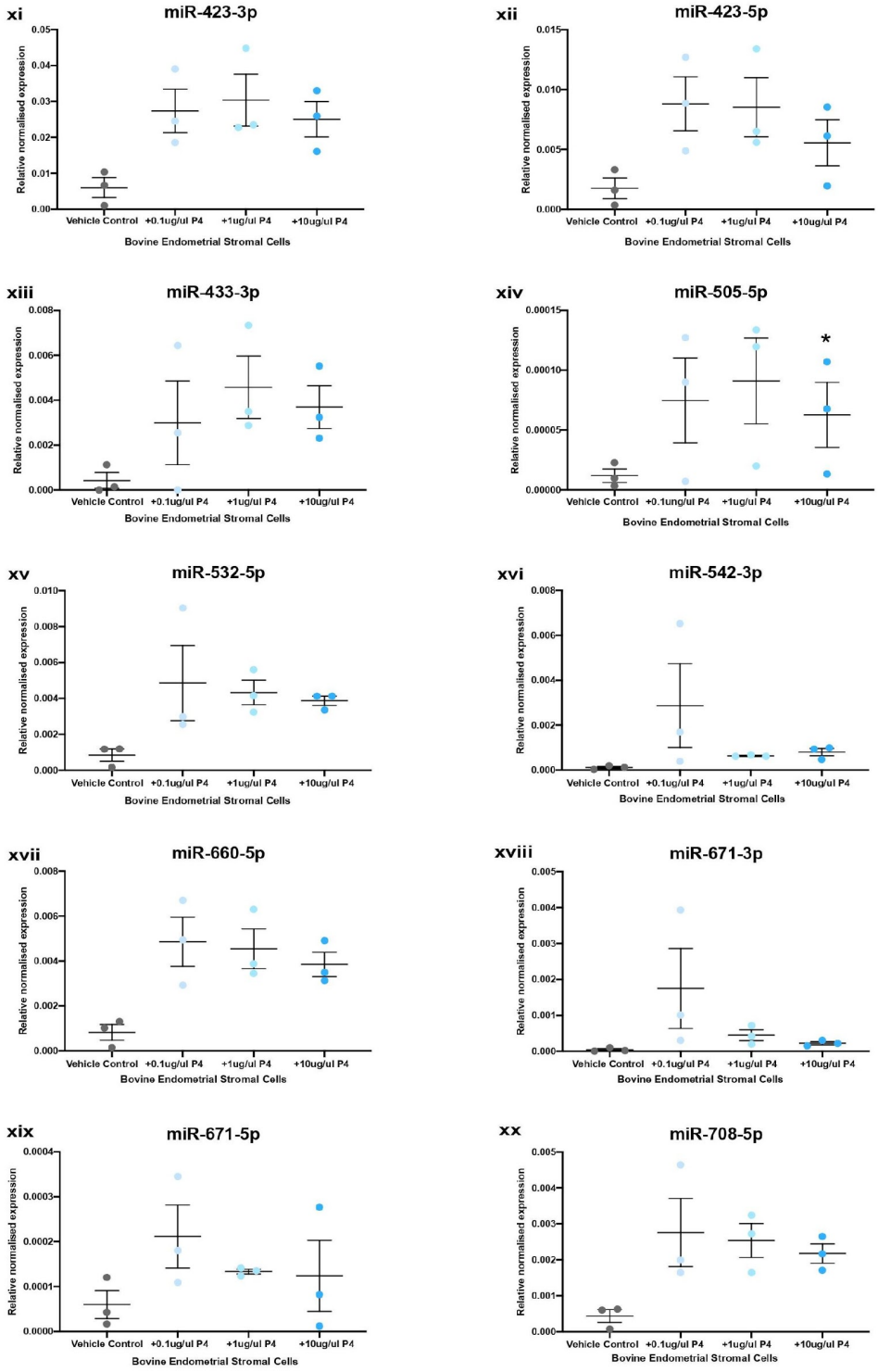


Figure 4.13 ***Expression of stem lineage miRNAs in bovine endometrial stromal cells treated with P4.*** *Expression of stem lineage miRNA (i) miR-28-3p, (ii) miR-127-3p, (iii) miR-151a-3p, (iv) miR-151a-5p, (v) miR-185-5p, (vi) miR-188-5p, (vii) miR-324-5p, (viii) miR-331-3p, (ix) miR-340-5p, (x) miR-378a-5p, (xi) miR-423-3p, (xii) miR-423-5p, (xiii) miR-433-3p, (xiv) miR-505-5p, (xv) miR-532-5p, (xvi) miR-542-3p, (xvii) miR-660-5p, (xviii) miR-671-3p, (xix) miR-671-5p and (xx) miR-708-5p in bovine endometrial stromal cells treated with vehicle control (grey circle), 0.1µg/mL (light blue), 1.0µg/mL (medium blue) or 10µg/mL P4 (dark blue circle) for 24 hours. Significant differences in miRNA expression values determined when p≤0.05 are depicted by an asterisk (*).*

4.4.17 Reactome pathway results for predicted targets of miR-505-5p, a stem lineage miRNA significantly downregulated in bovine endometrial stromal cells treated with P4

Of the 25 most significant pathways for predicted targets of miR-505-5p (ranked by p-value), 16 were statistically significant (p-value ≤ 0.05), shown in bold in Table 4.8.

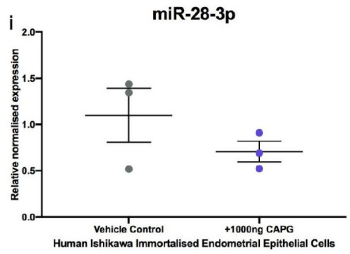
Table 4.8 Reactome pathway analysis results for miRNAs biologically regulated by P4 in bovine endometrial stromal cells.

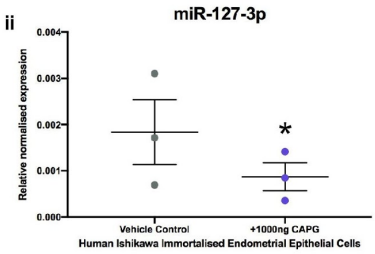
Pathway name	Entities				Reactions	
	found	ratio	p-value	FDR*	found	ratio
SUMOylation of DNA methylation proteins	12/16	0.001	0.009	0.859	3/3	2.40e-04
Semaphorin interactions	36/70	0.005	0.01	0.859	38/40	0.003
Potassium Channels	49/103	0.007	0.011	0.859	17/19	0.002
Formation of a pool of free 40S subunits	50/106	0.007	0.012	0.859	2/2	1.60e-04
VEGF ligand-receptor interactions	7/8	5.54e-04	0.02	0.859	4/4	3.20e-04
VEGF binds to VEGFR leading to receptor dimerization	7/8	5.54e-04	0.02	0.859	3/3	2.40e-04
CRMPs in Sema3A signaling	12/18	0.001	0.021	0.859	5/5	4.00e-04
GTP hydrolysis and joining of the 60S ribosomal subunit	53/120	0.008	0.03	0.859	3/3	2.40e-04
Sema4D induced cell migration and growth-cone collapse	14/24	0.002	0.035	0.859	6/6	4.80e-04
SUMO E3 ligases SUMOylate target proteins	76/183	0.013	0.038	0.859	130/130	0.01
SUMOylation of DNA damage response and repair proteins	37/81	0.006	0.041	0.859	24/24	0.002
L13a-mediated translational silencing of Ceruloplasmin expression	52/120	0.008	0.041	0.859	3/3	2.40e-04
SUMOylation	79/192	0.013	0.041	0.859	139/139	0.011
SUMOylation of transcription cofactors	22/44	0.003	0.046	0.859	27/27	0.002
Voltage gated Potassium channels	22/44	0.003	0.046	0.859	1/1	8.01e-05
Neurophilin interactions with VEGF and VEGFR	4/4	2.77e-04	0.047	0.859	2/2	1.60e-04
Eukaryotic Translation Initiation	55/130	0.009	0.052	0.859	18/21	0.002
Cap-dependent Translation Initiation	55/130	0.009	0.052	0.859	15/18	0.001
Adherens junctions interactions	18/35	0.002	0.053	0.859	15/16	0.001
Peptide chain elongation	42/97	0.007	0.061	0.859	4/5	4.00e-04
Nonsense-Mediated Decay (NMD)	52/124	0.009	0.065	0.859	6/6	4.80e-04
Nonsense Mediated Decay (NMD) enhanced by the Exon Junction Complex (EJC)	52/124	0.009	0.065	0.859	5/5	4.00e-04
Regulation of MECP2 expression and activity	19/39	0.003	0.073	0.859	13/14	0.001
RUNX1 interacts with co-factors whose precise effect on RUNX1 targets is not known	19/39	0.003	0.073	0.859	4/5	4.00e-04
Response of EIF2AK4 (GCN2) to amino acid deficiency	48 /115	0.008	0.078	0.859	10/16	0.001

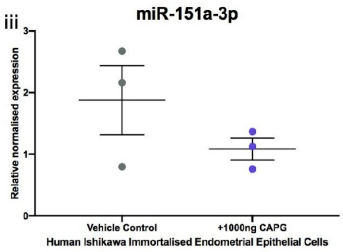
Reactome pathway results miR-505-5p, a stem lineage miRNA significantly downregulated in bovine endometrial stromal cells treated with 10µg/mL P4 for 24 hours in comparison to vehicle control. 'Entities - found' refers to the number of curated molecules within the named pathway found in the submitted dataset. 'Entities – ratio' refers to the number of curated molecules represented in the pathway compared to the entire database. 'Reactions – found' refers to the number of reactions within the named pathway represented by molecules in the submitted dataset. 'Reactions – ratio' refers to the number of Reactome reactions represented by the reactions in this pathway. False Discovery Rate (FDR) is calculated using the Benjamini-Hochberg method.

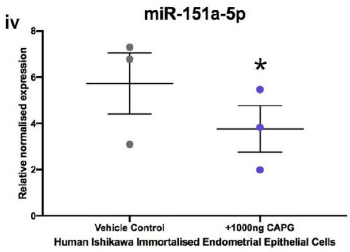
4.4.18 Expression analysis of human Ishikawa immortalised endometrial epithelial cells treated with bCAPG

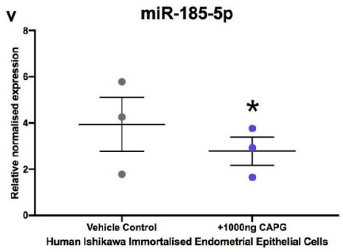
miR-127-3p, miR-151a-3p, a paralogue of miR-28 originating on the Eutherian stem lineage, and miR-188-5p showed a significant decrease in expression in human Ishikawa immortalised endometrial epithelial cells treated with 1000ng/µl bCAPG compared to vehicle control. miRNA expression levels are shown in Figure 4.14.

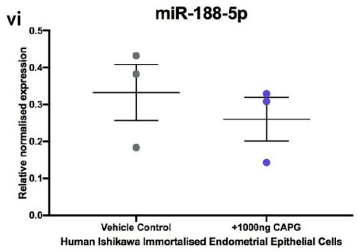

i
miR-28-3p
Relative normalised expression
2.0
1.5
1.0
0.5
0.0
Vehicle Control
+1000ng CAPG
Human Ishikawa Immortalised Endometrial Epithelial Cells

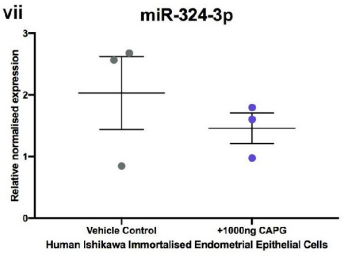

ii
miR-127-3p
Relative normalised expression
0.004
0.003
0.002
0.001
0.000
*
Vehicle Control
+1000ng CAPG
Human Ishikawa Immortalised Endometrial Epithelial Cells

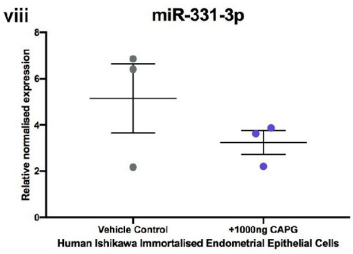

iii
miR-151a-3p
Relative normalised expression
3
2
1
0
Vehicle Control
+1000ng CAPG
Human Ishikawa Immortalised Endometrial Epithelial Cells

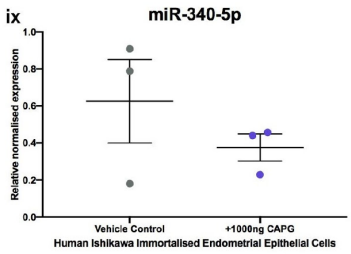

iv
miR-151a-5p
Relative normalised expression
8
6
4
2
0
*
Vehicle Control
+1000ng CAPG
Human Ishikawa Immortalised Endometrial Epithelial Cells

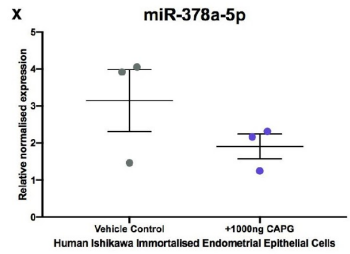

v
miR-185-5p
Relative normalised expression
8
6
4
2
0
*
Vehicle Control
+1000ng CAPG
Human Ishikawa Immortalised Endometrial Epithelial Cells


vi
miR-188-5p
Relative normalised expression
0.5
0.4
0.3
0.2
0.1
0.0
Vehicle Control
+1000ng CAPG
Human Ishikawa Immortalised Endometrial Epithelial Cells


vii
miR-324-3p
Relative normalised expression
3
2
1
0
Vehicle Control
+1000ng CAPG
Human Ishikawa Immortalised Endometrial Epithelial Cells


viii
miR-331-3p
Relative normalised expression
8
6
4
2
0
Vehicle Control
+1000ng CAPG
Human Ishikawa Immortalised Endometrial Epithelial Cells


ix
miR-340-5p
Relative normalised expression
1.0
0.8
0.6
0.4
0.2
0.0
Vehicle Control
+1000ng CAPG
Human Ishikawa Immortalised Endometrial Epithelial Cells


x
miR-378a-5p
Relative normalised expression
5
4
3
2
1
0
Vehicle Control
+1000ng CAPG
Human Ishikawa Immortalised Endometrial Epithelial Cells

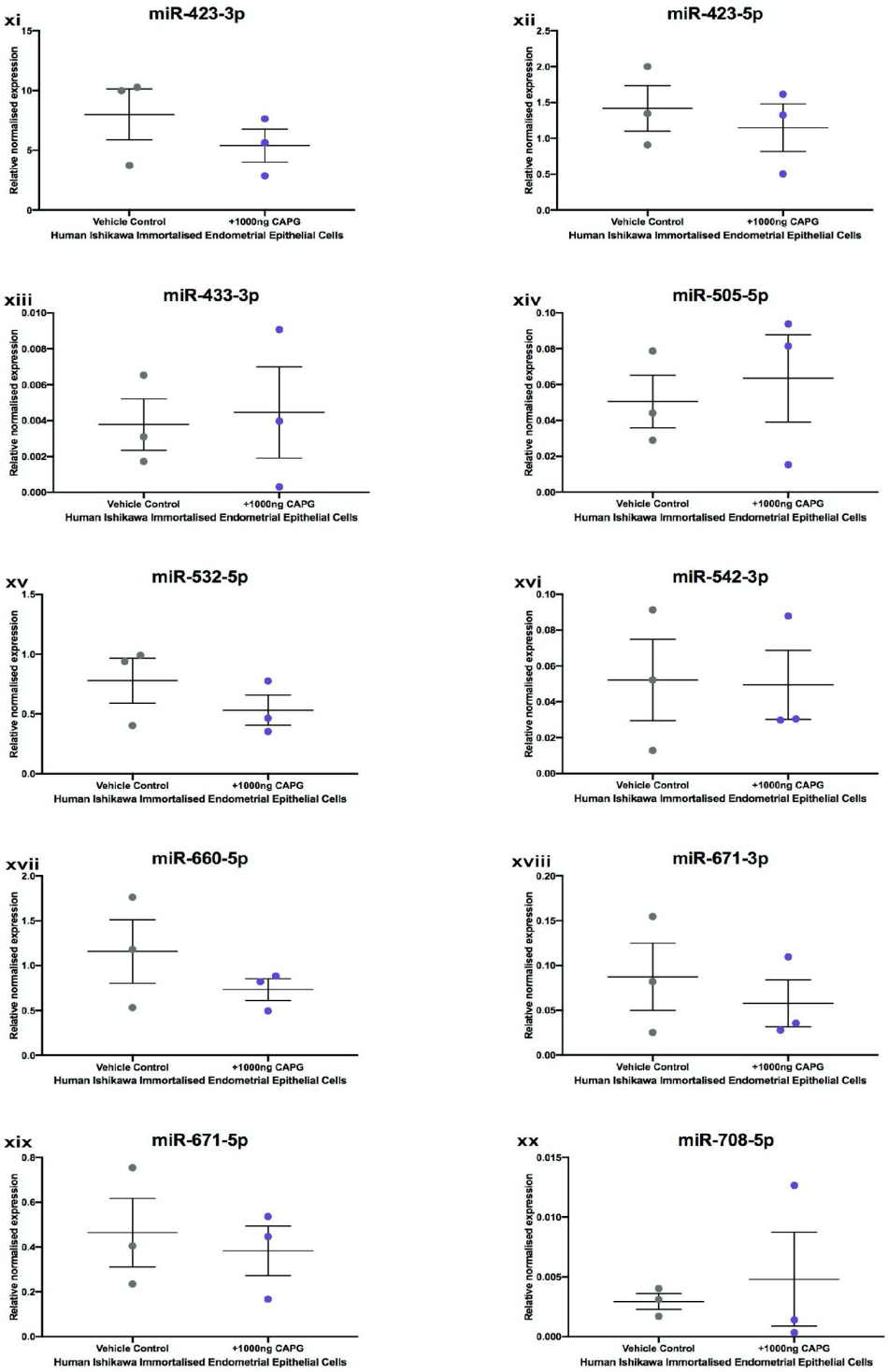


Figure 4.14 Expression of stem lineage miRNAs in human Ishikawa immortalised endometrial epithelial cells treated with bCAPG. *Expression of stem lineage miRNA (i) miR-28-3p, (ii) miR-127-3p, (iii) miR-151a-3p, (iv) miR-151a-5p, (v) miR-185-5p, (vi) miR-188-5p, (vii) miR-324-5p, (viii) miR-331-3p, (ix) miR-340-5p, (x) miR-378a-5p, (xi) miR-423-3p, (xii) miR-423-5p, (xiii) miR-433-3p, (xiv) miR-505-5p, (xv) miR-532-5p, (xvi) miR-542-3p, (xvii) miR-660-5p, (xviii) miR-671-3p, (xix) miR-671-5p and (xx) miR-708-5p in human Ishikawa immortalized endometrial epithelial cells treated with vehicle control (grey circle), or 1000ng/µl bCAPG (purple circle) for 24 hours. Significant differences in miRNA expression values determined when p≤0.05 are depicted by an asterisk (*).*

4.4.19 Reactome pathway results for predicted targets of miR-127-3p, miR-151a-3p and miR-188-5p, stem lineage miRNAs significantly downregulated in human Ishikawa immortalised endometrial epithelial cells treated with bCAPG

Of the 25 most significant pathways for predicted targets of stem lineage miRNAs significantly upregulated in human Ishikawa immortalised endometrial epithelial cells treated with bCAPG (ranked by p-value), 13 were statistically significant (p-value ≤ 0.05), shown in bold in Table 4.9.

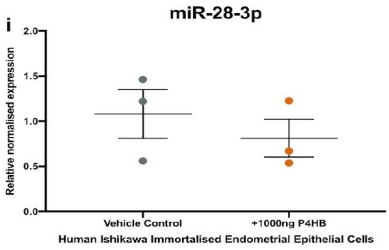
Table 4.9 Reactome pathway analysis results for miRNAs biologically regulated by bCAPG in human Ishikawa immortalised endometrial epithelial cells.

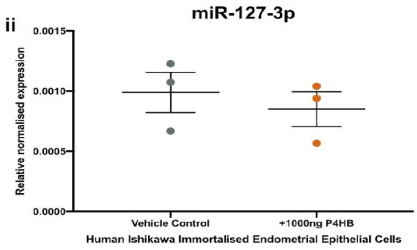
Pathway name	Entities				Reactions	
	found	ratio	p-value	FDR*	found	ratio
Voltage gated Potassium channels	25/44	0.003	0.02	0.794	1/1	8.01e-05
Intra-Golgi traffic	26/48	0.003	0.03	0.794	10/10	8.01e-04
Diseases of Mismatch Repair (MMR)	5/5	3.46e-04	0.036	0.794	6/6	4.80e-04
Loss of MECP2 binding ability to 5mC-DNA	5/5	3.46e-04	0.036	0.794	2/2	1.60e-04
NrCAM interactions	6/7	4.85e-04	0.043	0.794	4/4	3.20e-04
Post-transcriptional silencing by small RNAs	6/7	4.85e-04	0.043	0.794	3/3	2.40e-04
MECP2 regulates neuronal receptors and channels	18/32	0.002	0.046	0.794	26/26	0.002
Small interfering RNA (siRNA) biogenesis	7/9	6.24e-04	0.047	0.794	5/5	4.00e-04
Nef mediated downregulation of MHC class I complex cell surface expression	8/11	7.62e-04	0.049	0.794	4/4	3.20e-04
SMAD2/SMAD3:SMAD4 heterotrimer regulates transcription	21/39	0.003	0.049	0.794	17/17	0.001
Regulation of MECP2 expression and activity	21/39	0.003	0.049	0.794	14/14	0.001
MECP2 regulates transcription of neuronal ligands	9/13	9.01e-04	0.049	0.794	8/8	6.41e-04
Insulin-like Growth Factor-2 mRNA Binding Proteins (IGF2BPs/IMPs/VICKZs) bind RNA	9/13	9.01e-04	0.049	0.794	3/3	2.40e-04
TP53 Regulates Transcription of DNA Repair Genes	42/89	0.006	0.052	0.794	12/17	0.001
Neddylation	102/241	0.017	0.059	0.794	25/31	0.002
Phase 4 - resting membrane potential	12/20	0.001	0.063	0.794	2/2	1.60e-04
Potassium Channels	47/103	0.007	0.065	0.794	13/19	0.002
Nuclear Events (kinase and transcription factor activation)	16/29	0.002	0.066	0.794	10/11	8.81e-04
Downregulation of ERBB2 signaling	19/36	0.002	0.068	0.794	14/14	0.001
Activation of BH3-only proteins	19/36	0.002	0.068	0.794	17/19	0.002
Activation of NOXA and translocation to mitochondria	5/6	4.16e-04	0.068	0.794	5/5	4.00e-04
MASTL Facilitates Mitotic Progression	8/12	8.31e-04	0.073	0.794	4/4	3.20e-04
Transcriptional Regulation by MECP2	45/100	0.007	0.082	0.794	77/77	0.006
Membrane Trafficking	261/665	0.046	0.084	0.794	195/218	0.017
ERK/MAPK targets	14/26	0.002	0.094	0.794	7/7	5.61e-04

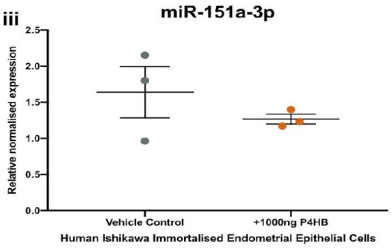
Reactome pathway results for stem lineage miRNAs significantly downregulated in human Ishikawa immortalised endometrial epithelial cells treated with 1000ng/µl bCAPG for 24 hours in comparison to vehicle control. 'Entities - found' refers to the number of curated molecules within the named pathway found in the submitted dataset. 'Entities – ratio' refers to the number of curated molecules represented in the pathway compared to the entire database. 'Reactions – found' refers to the number of reactions within the named pathway represented by molecules in the submitted dataset. 'Reactions – ratio' refers to the number of Reactome reactions represented by the reactions in this pathway. False Discovery Rate (FDR) is calculated using the Benjamini-Hochberg method.

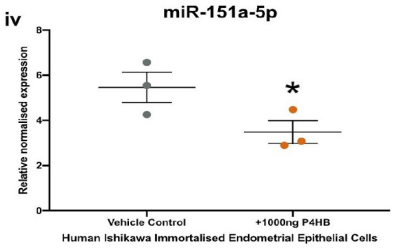
4.4.20 Expression analysis of human Ishikawa immortalised endometrial epithelial cells treated with bP4HB

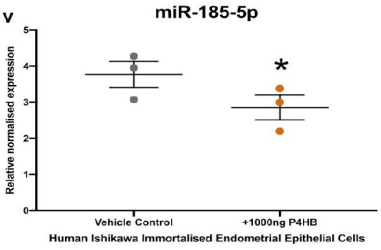
miR-151a-5p, a paralogue of miR-28 originating on the Eutherian stem lineage, miR-185-5p, miR-378a-3p and miR-532-5p, a paralogue of miR-188 originating on the Eutherian stem lineage, showed a significant decrease in expression in human Ishikawa immortalised endometrial epithelial cells treated with 1000ng/µl bP4HB compared to vehicle control. miRNA expression levels are shown in Figure 4.15.

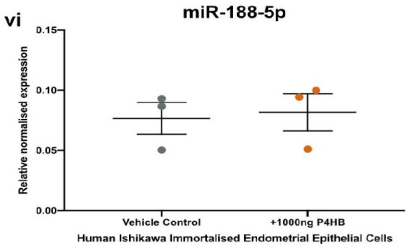

i
miR-28-3p
Relative normalised expression
2.0
1.5
1.0
0.5
0.0
Vehicle Control
+1000ng P4HB
Human Ishikawa Immortalised Endometrial Epithelial Cells

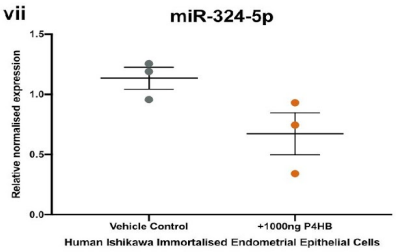

ii
miR-127-3p
Relative normalised expression
0.0015
0.0010
0.0005
0.0000
Vehicle Control
+1000ng P4HB
Human Ishikawa Immortalised Endometrial Epithelial Cells

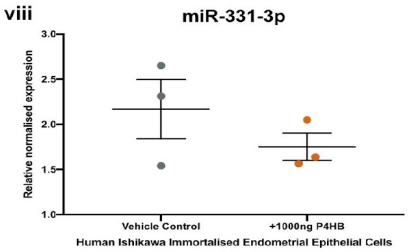

iii
miR-151a-3p
Relative normalised expression
2.5
2.0
1.5
1.0
0.5
0.0
Vehicle Control
+1000ng P4HB
Human Ishikawa Immortalised Endometrial Epithelial Cells

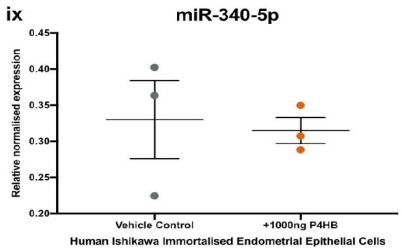

iv
miR-151a-5p
Relative normalised expression
8
6
4
2
0
*
Vehicle Control
+1000ng P4HB
Human Ishikawa Immortalised Endometrial Epithelial Cells

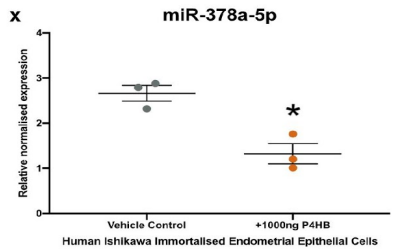

v
miR-185-5p
Relative normalised expression
5
4
3
2
1
0
*
Vehicle Control
+1000ng P4HB
Human Ishikawa Immortalised Endometrial Epithelial Cells


vi
miR-188-5p
Relative normalised expression
0.15
0.10
0.05
0.00
Vehicle Control
+1000ng P4HB
Human Ishikawa Immortalised Endometrial Epithelial Cells


vii
miR-324-5p
Relative normalised expression
1.5
1.0
0.5
0.0
Vehicle Control
+1000ng P4HB
Human Ishikawa Immortalised Endometrial Epithelial Cells


viii
miR-331-3p
Relative normalised expression
3.0
2.5
2.0
1.5
1.0
Vehicle Control
+1000ng P4HB
Human Ishikawa Immortalised Endometrial Epithelial Cells


ix
miR-340-5p
Relative normalised expression
0.45
0.40
0.35
0.30
0.25
0.20
Vehicle Control
+1000ng P4HB
Human Ishikawa Immortalised Endometrial Epithelial Cells


x
miR-378a-5p
Relative normalised expression
4
3
2
1
0
*
Vehicle Control
+1000ng P4HB
Human Ishikawa Immortalised Endometrial Epithelial Cells

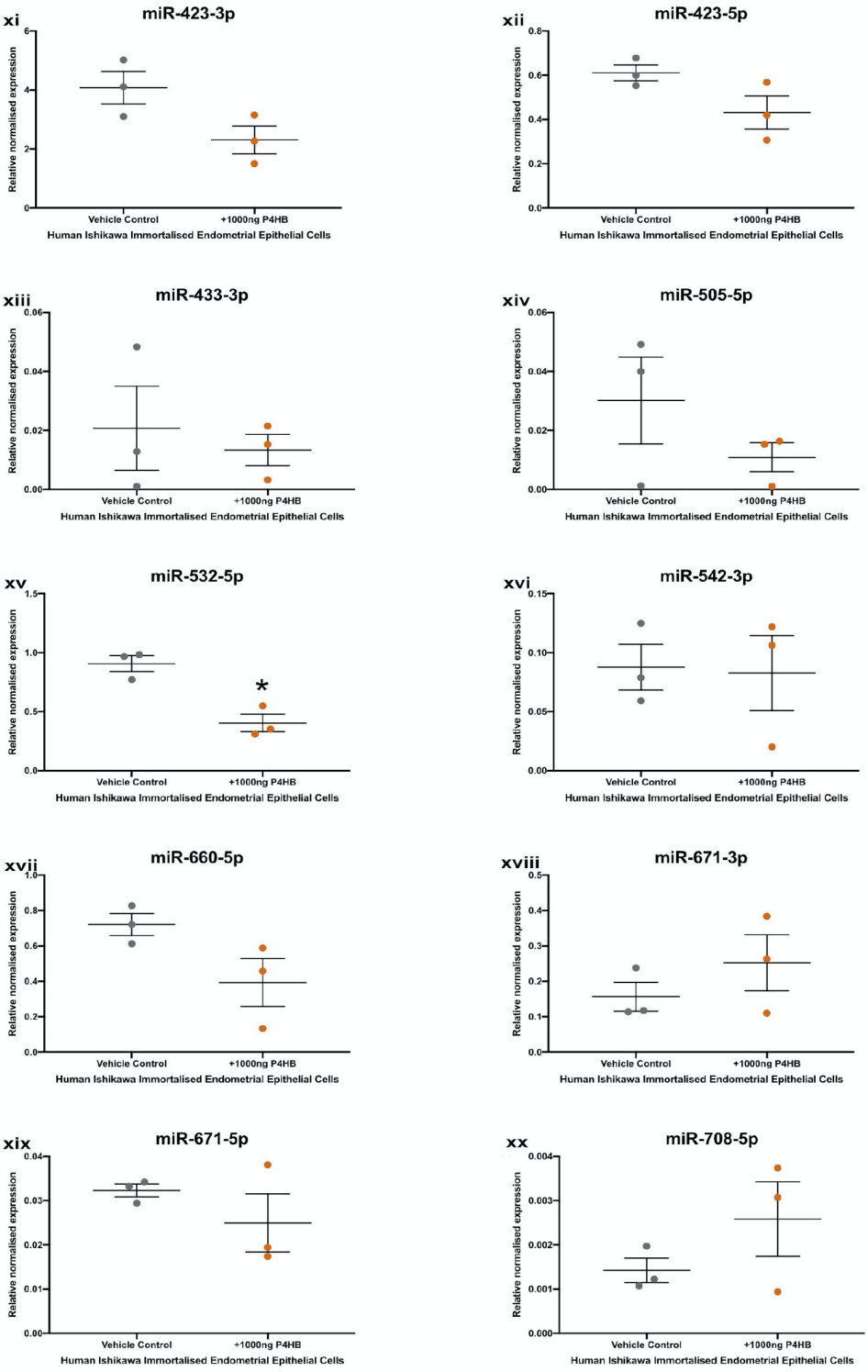


Figure 4.15 ***Expression of stem lineage miRNAs in human Ishikawa immortalised endometrial epithelial cells treated with bP4HB.*** *Expression of stem lineage miRNA (i) miR-28-3p, (ii) miR-127-3p, (iii) miR-151a-3p, (iv) miR-151a-5p, (v) miR-185-5p, (vi) miR-188-5p, (vii) miR-324-5p, (viii) miR-331-3p, (ix) miR-340-5p, (x) miR-378a-5p, (xi) miR-423-3p, (xii) miR-423-5p, (xiii) miR-433-3p, (xiv) miR-505-5p, (xv) miR-532-5p, (xvi) miR-542-3p, (xvii) miR-660-5p, (xviii) miR-671-3p, (xix) miR-671-5p and (xx) miR-708-5p in human Ishikawa immortalized endometrial epithelial cells treated with vehicle control (grey circle), or 1000ng/µl bP4HB (orange circle) for 24 hours. Significant differences in miRNA expression values determined when p≤0.05 are depicted by an asterisk (*).*

4.4.21 Reactome pathway results for predicted targets of miR-151a-5p, miR-185-5p, miR-378a-3p and miR-532-5p, stem lineage miRNAs significantly downregulated in human Ishikawa immortalised endometrial epithelial cells treated with bP4HB

Of the 25 most significant pathways for predicted targets of stem lineage miRNAs significantly upregulated in human Ishikawa immortalised endometrial epithelial cells treated with bP4HB (ranked by p-value), none were statistically significant (p-value ≤ 0.05), shown in Table 4.10.

Table 4.10 Reactome pathway analysis results for miRNAs biologically regulated by bP4HB in human Ishikawa immortalised endometrial epithelial cells.

Pathway name	Entities				Reactions	
	found	ratio	p-value	FDR*	found	ratio
Deactivation of the beta-catenin transactivating complex	38/44	0.003	0.119	0.911	14/14	0.001
RUNX3 regulates WNT signaling	10/10	6.93e-04	0.172	0.911	5/5	4.00e-04
Neurexins and neuroligins	48/60	0.004	0.202	0.911	19/19	0.002
Intra-Golgi traffic	39/48	0.003	0.202	0.911	10/10	8.01e-04
RUNX3 regulates CDKN1A transcription	8/8	5.54e-04	0.205	0.911	6/6	4.80e-04
Protein-protein interactions at synapses	72/93	0.006	0.219	0.911	33/33	0.003
Nectin/Necl trans heterodimerization	7/7	4.85e-04	0.226	0.911	8/8	6.41e-04
Post-transcriptional silencing by small RNAs	7/7	4.85e-04	0.226	0.911	3/3	2.40e-04
Loss of Function of TGFBR1 in Cancer	7/7	4.85e-04	0.226	0.911	2/2	1.60e-04
SUMOylation of DNA methylation proteins	14/16	0.001	0.241	0.911	3/3	2.40e-04
RAB GEFs exchange GTP for GDP on RABs	72/94	0.007	0.246	0.911	21/21	0.002
Transcriptional activation of p53 responsive genes	6/6	4.16e-04	0.249	0.911	5/5	4.00e-04
Transcriptional activation of cell cycle inhibitor p21	6/6	4.16e-04	0.249	0.911	5/5	4.00e-04
RUNX3 regulates BCL2L11 (BIM) transcription	6/6	4.16e-04	0.249	0.911	2/2	1.60e-04
TGFBR1 KD Mutants in Cancer	6/6	4.16e-04	0.249	0.911	1/1	8.01e-05
GRB7 events in ERBB2 signaling	6/6	4.16e-04	0.249	0.911	1/1	8.01e-05
Downregulation of ERBB4 signaling	10/11	7.62e-04	0.25	0.911	5/5	4.00e-04
CDC6 association with the ORC:origin complex	10/11	7.62e-04	0.25	0.911	2/2	1.60e-04
WNT ligand biogenesis and trafficking	23/28	0.002	0.254	0.911	12/12	9.61e-04
Adherens junctions interactions	28/35	0.002	0.271	0.911	16/16	0.001
Diseases of Mismatch Repair (MMR)	5/5	3.46e-04	0.277	0.911	6/6	4.80e-04
TP53 Regulates Transcription of Death Receptors and Ligands	15/18	0.001	0.289	0.911	7/7	5.61e-04
Insulin processing	24/30	0.002	0.289	0.911	13/16	0.001
E2F mediated regulation of DNA replication	18/22	0.002	0.291	0.911	4 /4	3.20e-04
Signaling by TGF-beta Receptor Complex in Cancer	8/9	6.24e-04	0.301	0.911	7/7	5.61e-04

Reactome pathway results for stem lineage miRNAs significantly downregulated in human Ishikawa endometrial epithelial cells treated with 1000ng/µl bP4HB for 24 hours in comparison to vehicle control. 'Entities - found' refers to the number of curated molecules within the named pathway found in the submitted dataset. 'Entities – ratio' refers to the number of curated molecules represented in the pathway compared to the entire database. 'Reactions – found' refers to the number of reactions within the named pathway represented by molecules in the submitted dataset. 'Reactions – ratio' refers to the number of Reactome reactions represented by the reactions in this pathway. False Discovery Rate (FDR) is calculated using the Benjamini-Hochberg method.

4.5 Discussion

Expression of miRNA families computationally predicted to originate on the Eutherian stem lineage in female reproductive tissue was confirmed through RT-qPCR of human miRNA primers against human, bovine, porcine and murine samples. Attempts to investigate the expression of two miRNAs (miR-340 and miR-671) computationally predicted to originate on the Therian stem lineage using Opossum uterine and mammary gland tissue samples failed to provide results. The high Ct values (indicating low miRNA expression) of miR-340 and miR-671 in opossum tissue may, however, be due to mutations in miRNA seed sequences seen on the eutherian stem lineage (shown in Table 4.11). Alternatively, these miRNAs may not be expressed in the opossum uterus or mammary gland, evolving uterine expression on the eutherian stem lineage.

Table 4.11 Seed sequences of miR-340 and miR-671 in mammalian species used to investigate phylogenetic placement.

Species	miR-340 Seed Sequence	miR-671 Seed Sequence
Homo sapiens	UAUAAAG	GGAAGCC
Mus musculus	UAUAAAG	GGAAGCC
Sus domesticus	Not available	Not available
Bos taurus taurus	UAUAAAG	CCGGUUC
Monodelphus domestica	AGUAAUG	Not available

Seed sequences for miR-340 and miR-671 were obtained from miRGeneDB2.0 (Fromm *et al.*, 2018) *for H. sapiens, M. muscules, S. domesticus, B. taurus taurus and M. domestica. Seed sequences were compared to identify mutations in miRNA seed sequence that occurred in eutherian members of the miR-340 and miR-671 gene families.*

However, both miR-340 and miR-671 have been confirmed to have originated on the Therian root using small RNA-sequencing (Fromm *et al.*, 2018). The low Ct value of miR-28-3p in Opossum tissue may be explained by erroneous binding to mdo-miR-7286, which has 89% homology to the mature hsa-miR-28-3p sequence. This RT-qPCR analysis therefore served to confirm the presence of these miRNA families across the Eutherian tree of life. This analysis of stem lineage miRNA expression in mammalian endometrial tissue and cell samples also confirmed that these miRNAs are expressed in the maternal contribution to the placenta, allowing further investigations into their role in early pregnancy events.

Results from RT-qPCR of miRNA expression in human Ishikawa endometrial epithelial cells and bovine endometrial epithelial and stromal cells show variation in Ct values obtained between samples subject to the same treatment – including when exposed to vehicle control (e.g. bovine endometrial epithelial cells treated with oIFNT). This variation may be due to animal-animal biological differences, in the case of bovine primary samples. This variation could also be explained by differences in cell response to treatment in cell culture. The large differences between miRNA expression may be obscuring other significant results. Therefore, if these experiments were repeated or subject to further analysis, a larger group of replicates may solve this issue.

miR-28-3p was found to be significantly upregulated in bovine endometrial explants treated with 1000ng/µl oIFNT. The lack of differential expression of miR-28-3p, or any other stem lineage miRNA analysed, in bovine endometrial epithelial and stromal primary cells suggests an extracellular function of miR-28-3p in bovine maternal recognition of pregnancy, given that differential expression was only observed in tissue samples, not in isolated endometrial cells. The presence of miRNAs in circulating extracellular vesicles (EVs) has been well documented (Moldovan *et al.*, 2013) with EVs now hypothesized to be an important method of cell-cell communication. miRNA's encapsulated in EVs have been linked to trophoblast differentiation in placental development (Gillet *et al.*, 2019), embryo-endometrium cross talk during early implantation (Bidarimath *et al.*, 2017) and over-proliferation of endometrial tissue during endometriosis (Khalaj *et al.*, 2019). Furthermore, miRNAs in circulating EVs have been posited as a possible biomarker of successful pregnancy in cattle (Pohler *et al.*, 2017), with 3 differentially expressed miRNAs found in serum from cattle with successful pregnancies when compared to cattle affected with early pregnancy loss or lack of pregnancy.

The extracellular nature and specific response of miR-28-3p to oIFNT in *in-vitro* models of early pregnancy suggests a role of miR-28-3p in the transcriptional changes and altered cell communication occurring after maternal recognition of pregnancy. This hypothesis was supported by a Reactome Pathway analysis (Jassal *et al.*, 2019) of the predicted targets of miR-28-3p. Among the overrepresented pathways, targets of miR-28-3p are implicated in Grb2 events

in the epidermal growth factor (EGF) family. Grb2 signalling has been associated with successful pregnancy in murine models, whereby Grb2-/- mice were found to display early embryonic mortality due to impaired endoderm differentiation (Cheng *et al.*, 1998). A second mouse study investigating the role of EGF receptors in early pregnancy found that siRNA-mediated *Egfr* knockout mice experience pregnancy loss post-implantation due to inhibited cell differentiation, cell proliferation and issues with post-implantation decidualisation (Large *et al.*, 2014). Targets of miR-28-3p were also associated with SMAD signalling, with SMAD transcription factors acting as transducers of transforming growth factor-β (TGF-β) signalling. The TGF-β family has been extensively linked to female reproduction, including ovulation, through the growth differentiation factor 9 (GDF9) gene; decidualisation, through NODAL signalling; and implantation, through activin receptor-like kinase-2 (ALK2)-mediated signalling (Li, 2014). Predicted targets of miR-28-3p were also implicated in phosphoinositide 3-kinases (PI3K) events in EGF family signalling. PI3K enzymes phosphorylate the inositol ring of phosphatidylinositol lipids and assist the action of intracellular effector molecules (Vanhaesebroeck, Stephens and Hawkins, 2012). Previously, PI3K action was found to regulate the transcription of ISGs, with ISG expression significantly downregulated in the presence of PI3K inhibitors (Sheikh, Hooda and Dang, 2018). Pathways related to PI3K-mediated action have been observed to be over-expressed in genes modified by IFNT secretion during MRP (Forde *et al.*, 2012). Taken together, this may suggest that miR-28-3p has a role in regulating molecules involved in PI3K signalling as they induce the expression of ISGs, thereby assisting IFNT in modifying the maternal endometrium.

This downstream analysis of the targets of miR-28-3p provide further evidence of the role of miR-28-3p in altering the uterine transcriptome post-MRP, targeting genes involved in cell proliferation, implantation and mediating the effects of IFNT on ISGs. Further analysis of miR-28-3p in *in-vitro* models of pregnancy recognition, specifically in tissue explants and isolated EVs from endometrium of successful and unsuccessful bovine pregnancies may elucidate a more specific role of miR-28-3p in successful pregnancy establishment, implantation and maintenance.

Human Ishikawa endometrial epithelial cells showed significant downregulation of miR-127-3p, miR-151a-5p and miR-188-5p in response to 1000ng/µl bCAPG, a protein found to be secreted by the bovine conceptus pre-implantation during an investigation of conceptus-mediated alterations to the maternal endometrium independently of MRP (Forde *et al.*, 2015) that shows high sequence conservation across the eutherian clade, with >90% sequence homology between human and bovine CAPG sequences, indicating a degree of conservation of bCAPG over eutherian evolution (Tinning *et al.*, 2020). miRNA expression profiling of bovine epithelial and stromal cells treated with 1000ng/µl bCAPG showed little overlap with miRNAs differentially expressed in human Ishikawa cells. Bovine endometrial epithelial cells treated with 1000ng/µl bCAPG showed significant upregulation of miR-324-5p, miR-331-3p and miR-505-5p. In contrast, bovine endometrial stromal cells treated with 1000ng/µl bCAPG showed a significant decrease in miR-660-5p expression, a paralogue of miR-188-5p originating on the stem Eutherian lineage.

The lack of overlap of miRNA differential expression between human and bovine endometrial epithelial cells treated in the same manner suggest a species-specific response to a conserved implantation-related protein. The lack of a conserved response, in context of the high sequence homology between human and bovine CAPG, suggests the adaptation of conserved pathways to facilitate morphologically different implantation strategies and pregnancies. Marked morphological differences exist between human and bovine early pregnancies. As previously discussed, the human embryo implants 6-10 days post-ovulation, with the endometrium only receptive to the embryo for this short window, before invasive implantation, with the fetal trophoblast cells embedding into the endometrium (McGowen *et al.*, 2014; Su and Fazleabas, 2015). The bovine conceptus, however, has a protracted period during which it free floats in the endometrium for a number of days, elongating from a spherical to a filamentous conceptus prior to implantation 19-20 days post-insemination, prior to adhesion to the maternal endometrium and less invasive implantation (Hue, Degrelle and Turenne, 2012). The longer implantation period and different conceptus morphology may contribute to these differences in miRNA differential expression in response to the same conserved treatment.

Downstream analysis of the targets of these differentially expressed miRNAs supports species-specific functional responses to bCAPG treatment. Targets of bCAPG-regulated miRNAs differentially expressed in human Ishikawa cells showed overrepresentation of genes related to intra-Golgi transport; SMAD-mediated transcription; and Insulin-like Growth Receptor binding proteins, previously found to be differentially expressed in maternal serum throughout pregnancy (GIUDICE *et al.*, 1990). In contrast, targets of bCAPG-regulated miRNAs differentially expressed in bovine endometrial epithelial cells were implicated in adherens interactions, EGF family signalling, and the formation of new blood vessels via vascular endothelial growth factor (VEGF) and VEGF receptor interactions. While the associated functions are important in early pregnancy, the markedly different Reactome Pathway reconstructions support the species-specific nature of bCAPG response and may point to a different transcriptomic environment required to facilitate successful implantation depending on the implantation strategy. The SMAD family of proteins transduce the signals of the transforming growth factor beta (TGF-β) family. The TGF-β family are well-known inhibitors of epithelial and endothelial cell proliferation by inhibiting cyclin-dependent kinases (CDKs), possibly by the activation of the p21 and p53 to inhibit the expression or counter-act the function of CDKs (Datto *et al.*, 1995). TGF-β also been observed to regulate human trophoblast invasion during the implantation process by negatively regulating metalloproteases in concert with Interleukin 1 beta (IL-1β), an upregulator of trophoblast proteases, leading to the theory that TGF-β and IL-1β act to fine-tune the invasiveness of the trophoblast cells in order to tightly control the implantation process (Karmakar and Das, 2002). That this pathway was significantly overexpressed among the targets of stem lineage miRNAs regulated by bCAPG in human Ishikawa immortalised endometrial epithelial cells indicates that these miRNAs may work to modulate the action of TGF-β during the invasive implantation process seen in human pregnancy. In contrast, the significant overexpression of functions related to adherens junctions may be due to the greater focus on adhesion versus invasion seen in the bovine implantation strategy.

The downregulation of miR-660-5p in bovine endometrial stromal cells in response to bCAPG provides evidence of a cell-specific, as well as a species-specific response to conserved implantation-related proteins. While bovine stromal cells do not decidualise (Kin *et al.*, 2016),

they play necessary roles during bovine pregnancy, with the majority of receptors for progesterone and estradiol being located in the stromal cell population (Kurita *et al.*, 1998). The predicted pathways associated with targets of miR-660-5p include events related to ERBB2, a member of the EGF family, with the EGF family also implicated as possibly regulated by miR-28-3p in response to oIFNT. Targets of miR-660-5p are overexpressed in pathways related to the downregulation of TGF-β signalling. While TGF-β is an inhibitor of epithelial and endothelial cell proliferation, it is a promotor of fibroblast cell proliferation (Kubiczkova *et al.*, 2012), indicating that miR-660-5p may act to prevent any proliferative effect of TGF-β on stromal cells during the implantation period.

These species- and cell-specific transcriptional responses were also found in human Ishikawa and bovine endometrial epithelial and stromal cells treated with 1000ng/µl bP4HB. Human Ishikawa endometrial epithelial cells showed significant downregulation of miR-151a-5p, miR-185-5p, miR-378a-3p and miR-532-5p (a paralogue of miR-188 originating on the *Eutherian* lineage). While miR-185-5p was also significantly downregulated in bovine endometrial epithelial cells treated with 1000ng/µl bP4HB (discussed below), this was the only significantly differentially expressed miRNA, in contrast to the multiple miRNAs displaying a significant response to bP4HB in human Ishikawa cells. The downstream targets of miRNAs differentially expressed only in human Ishikawa cells treated with bP4HB showed no significant overexpression of any specific pathway. However, TGF-β related functions, such as activation of p53 and p21 cell cycle inhibitors are associated with predicted targets of stem lineage miRNAs differentially expressed in response to bP4HB in human Ishikawa endometrial epithelial cells. That these pathways were implicated, similarly to miRNAs regulated by bCAPG in human Ishikawa endometrial epithelial cells, but are not significantly overexpressed, suggest that the main modulation of TGF-β function is carried out by miRNAs regulated by bCAPG but may be supported by bP4HB regulated miRNAs in humans

Similarly to the response of stromal cells treated with bCAPG, bovine endometrial stromal cells treated with 1000ng/µl bP4HB showed significant upregulation of miR-188-5p, with predicted associated pathways including TGF-β signalling, SMAD regulation of transcription and EGF family activity, similarly to targets of miR-660-5p. This result again demonstrates a separate stromal cell

response to conserved implantation related proteins. While miR-660-5p, downregulated in the presence of 1000ng/µl bCAPG, is a paralogue of miR-188-5p, miR-188-5p and miR-660-5p display different responses to these two proteins implicated in implantation. These different responses by two miRNA genes within the same family indicates the evolution of separate roles after the evolution of the miR-660-5p paralogue. Reactome Pathway analysis of miR-188-5p and miR-660-5p implicates the targets of these miRNAs in a number of common pathways, including SMAD-mediated transcription; downregulation of ERBB2 (a member of the EGF receptor family) signalling; and intra-Golgi traffic, indicating alterations in protein and lipid secretion. The crossover in significantly overexpressed pathways, when considered along with the different expression patterns of miR-188-5p and its paralogue miR-660-5p in response to bP4HB and bCAPG respectively, may suggest a fine-tuning of sorts, with miR-188-5p and miR-660-5p acting in concert to control the action of EGF proteins and the proliferative action of TGF-β during implantation. However, a number of differences in implicated pathways are present. While the targets of miR-188-5p and miR-660-5p are implicated in ERBB2 signalling, targets of miR-660-5p are also implicated in the downregulation of ERBB4 signalling, suggesting multiple pathways of epidermal growth alterations. The targets of miR-660-5p are also overrepresented in Estrogen Receptor (ESR)-mediated nuclear events, which include the activation of MAPK signalling (Driggers and Segars, 2002), previously identified as an important mediator of endothelial cell proliferation in angiogenesis (Hu and Zhang, 2017).

miR-185-5p was found to be significantly downregulated in both human Ishikawa and bovine endometrial epithelial cells in response to 1000ng/µl bP4HB, indicating a level of conservation in endometrial epithelial cell response to implantation related proteins in tandem with the species-specific alterations discussed above. The conserved response of miR-185-5p indicates some conservation within the morphologically different human and bovine implantation strategies, with some section of the implantation-related transcriptional response induced by bP4HB regulated in the same manner across different implantation morphologies. The Reactome Pathway reconstruction of the predicted targets of miR-185-5p include some previously discussed pathways, including Grb family signalling, WNT signalling, EGF and EGF receptor interactions and signalling, along with loss of function of the TGF-β receptor during cancer. Genes

involved in WNT signalling have previously been observed to be downregulated at Day 16 of bovine pregnancy (Forde *et al.*, 2011). While this downregulation may have been induced by IFNT secretion by the conceptus, bP4HB may have a role in the modulation of WNT signalling by downregulating miR-185-5p. Whether this serves to remove a repressive action of miR-185-5p on genes involved with Wnt signalling, or prevents miR-185-5p-mediated upregulation of Wnt signalling in order to continue the downregulation after Day 16 of pregnancy is unclear. However, members of the Wnt signalling pathway have previously been observed to promote changes to the endometrium to facilitate implantation in both humans and ruminants (Tepekoy, Akkoyunlu and Demir, 2015), providing evidence of a role of bP4HB in returning Wnt signalling pathway members to normal expression levels. Two adhesion-related pathways; adherens junction interactions (also overrepresented in miRNAs regulated by bCAPG in bovine endometrial epithelial cells) and leucine repeat-rich adhesion molecules, also implicated in synaptic adhesion (not overrepresented in any experiment in this study). The overrepresentation of two separate adhesion pathways suggests an increased role of miR-185-5p in mediating successful adhesion. This, along with the conserved response to bP4HB observed in both human and bovine endometrial epithelial cells mark miR-185-5p as a target for future investigations into successful conceptus adhesion and implantation into the endometrium across the *Eutherian* clade.

miR-505-5p, a stem lineage miRNA originating on the eutherian stem lineage, was found to be significantly upregulated in bovine endometrial epithelial cells treated with 10µg/mL P4 in comparison to bovine endometrial epithelial cells treated with 0.1µg/mL P4; and in bovine endometrial stromal cells treated with 10µg/mL in comparison to vehicle control. During early pregnancy, P4 serves to suppress proliferation by downregulating estradiol expression, and by extension its own receptor, and induce differentiation of epithelial cells to line the endometrium (Radi, Marusak and Morris, 2009). The induction of differential expression of miR-505-5p only occurred at P4 concentrations of 10µg/mL. During early bovine pregnancy, P4 concentrations were found to be >0.001µg/mL at Day 24, reaching 0.008µg/mL on Day 31. These endogenous P4 concentrations, when compared to the concentration at which miR-505-5p suggest that miR-505-5p pregnancy is either downregulated or expressed at normal levels during early pregnancy. Among the significantly overexpressed pathways related to predicted targets of miR-505-5p are

a number of functions related to VEGF-VEGFR interactions, as in the predicted associated pathways of targets regulated by miRNAs differentially expressed in response to bCAPG treatment in bovine endometrial epithelial cells; and SUMOylation of DNA methylation proteins, DNA damage response and repair proteins, transcription factors; and the action of SUMO E3 ligases. SUMOylation, an post-translational modification carried out by small ubiquitin-like modifier (SUMO) proteins, in which a member of the SUMO family is ligated to a lysine residue on the target protein, is a process involved in the regulation of cell growth; response to DNA damage; transcriptional regulation; and protein-protein interactions (Yang *et al.*, 2017). SUMOylation may also act to prevent the binding of a protein to a target via the ligation of a SUMO protein to a lysine residue within the target binding site, or alter the function of a target gene by changing its conformation (Wilkinson and Henley, 2010). SUMOylation has been observed on to influence the action of both isoforms of the progesterone receptor (PR-A and PR-B) and has been extensively studied in the context of progesterone-receptor positive breast cancer, with deSUMOylated or SUMO-deficient progesterone receptors found to positively regulate EGF family mediated cell proliferation (Knutson *et al.*, 2012), with SUMOylation of PRs found to downregulate the PR-dependent gene transcription (Abdel-Hafiz and Horwitz, 2012). A study of the effects of SUMOylated vs deSUMOylated (or 'normal') PR receptors found that, when PR-A was SUMOylated, PR-A lost its ability to upregulate genes related to PR-induced cell proliferation, the cell cycle and cellular assembly and organisation, indicating that SUMOylation of the PR-A receptor may be a method to abrogate its proliferative effects. As mentioned above, progesterone secretion during early pregnancy serves to repress cell proliferation via the downregulation of its own receptor and preventing PR-mediated cell proliferation. It is possible that SUMOylation may work in concert with progesterone to downregulate proliferation and prevent PR-mediated gene transcription. That the predicted targets of miR-505-5p, a miRNA upregulated at high progesterone concentrations but unaffected at concentrations associated with early pregnancy, are significantly associated with SUMOylation activities may indicate that the effects of miR-505-5p upregulation on the SUMOylation process are either un-needed or detrimental during early pregnancy.

4.6 Conclusion

The expression of of miRNAs predicted to have emerged on the Eutherian stem lineage (Chapter 3) in eutherian female reproductive tissues was confirmed, using five mammalian species. While the RT-qPCR based investigation the expression of stem lineage miRNAs in therian female reproductive tissue was unsuccessful, these miRNAs were confirmed to be expressed in Eutherian uterine and endometrial samples. Evidence of the involvement of these stem lineage miRNAs in early pregnancy events was discovered, with the expression of 10 miRNAs found to be significantly altered in response to bovine MRP, bovine progesterone treatment and conceptus-secreted implantation-related proteins in both human and bovine cells. The species- and cell-specific miRNA response observed in bovine endometrial epithelial and stromal cells treated with 1000ng/µl bCAPG provides novel candidates for research into bovine conceptus implantation. That the miRNA response to bCAPG treatment is different in human and bovine cells may indicate a requirement for different transcriptional environments in the endometria of animals with different implantation strategies. This is supported by the association of TGF-β action, a mechanism which controls metalloprotease efficiency during invasive implantation, with the targets of miRNAs regulated by bCAPG in human cells, in contrast to the association of adherens proteins with the targets of miRNAs regulated by bCAPG in bovine cells. The conserved response to bP4HB in both human and bovine endometrial epithelial cells suggests a conserved mechanism by which the implantation strategy is regulated, perhaps by inducing the transcription of genes required for both types of implantation strategies. Finally, while these 12 miRNAs emerged at the same point in mammalian evolution and remain fixed in all extant mammals, the small overlap between differentially expressed miRNAs in bovine and human systems analysed suggests that these miRNAs may have evolved to regulate early pregnancy in a species-specific manner.

Chapter 5: Expression Analysis of Conserved Stem Lineage miRNAs

5.1 Aim of this chapter

1. Use *in vitro* models of early pregnancy to investigate whether positively selected mRNA targets of miRNAs that are modified in response to early pregnancy cues have coordinated changes to their expression patterns.

5.2 Introduction

As discussed in Chapter 1, the successful establishment of pregnancy and the initiation of placentation in eutherian mammals is a complex process, requiring extensive remodelling of the endometrium and communication between the receptive mother and the conceptus. Secretion of the steroid hormone progesterone from the corpus luteum facilitates transcriptional changes by binding to its nuclear receptor, which exists in two isoforms, progesterone receptor A (PGRA) and progesterone receptor B (PGRB) (Graham and Clarke, 1997). Communication between the conceptus and the maternal uterus is mediated by the secretion of signalling molecules, including the maternal recognition of pregnancy (MRP) signal. The process of MRP is carried out by human chorionic gonadotropin (hCG) at Day 8-10 of pregnancy in humans (Fishel, Edwards and Evans, 1984) and by Interferon Tau (IFNT) (Godkin *et al.*, 1982) at Day 16 of bovine pregnancy, acting to prevent degradation of the corpus luteum and continue progesterone production during early pregnancy (Bazer *et al.*, 2011). Along with the MRP signal, the conceptus has been found to secrete other molecules which may function to support implantation (Forde *et al.*, 2015), with these molecules being conserved across *Eutheria*, leading to the hypothesis that conceptus secreted factors (aside from MRP) may serve to support the establishment of early pregnancy (Tinning *et al.*, 2020).

Studies of these molecules and their role in early pregnancy in *Eutheria* have led to the observation of transcriptional changes induced by these molecules, which may enable the extensive alterations to the uterine environment required for early pregnancy and implantation.

These transcriptional alterations are detailed for both invasive and non-invasive implantation in Table 1.2, repeated below as Table 5.1.

Table 5.1 Molecules associated with early pregnancy and the transcriptional alterations they induce in species representative of invasive (human) and non-invasive implantation and placental types.

Implantation Strategy	Signal	Species	Number of differentially expressed genes (DEGs)	Study
Invasive	P4	*Homo sapiens*	653 (454↑, 199↓)	(Chi *et al.*, 2019)
	MRP (hCG)	*Homo sapiens*	173 (130↑, 43↓)	(Li *et al.*, 2020)
	Conceptus produced signals	*Homo sapiens*	340 (184↑, 156↓)	(Altmäe *et al.*, 2012)
Non-invasive	P4	*Bos taurus*	4421 (2313↑, 2108↓)	(Forde *et al.*, 2012)
	MRP (IFNT)	*Bos taurus*	212 (173↑, 39↓)	(Bauersachs *et al.*, 2012)
	Conceptus produced signals	*Bos taurus*	764 (514↑, 250↓)	(Forde *et al.*, 2011)

The establishment of eutherian pregnancy requires extensive communication between the maternal endometrium and the conceptus. This communication is facilitated by molecules such as (i) progesterone (P4) secreted by the corpus luteum after ovulation; (ii) maternal recognition of pregnancy, a signal emitted by the conceptus to signal its presence to the maternal uterus, which is hCG in humans and IFNT in ruminants; and (iii) Other conceptus produced signals which may aid in promoting uterine receptivity during implantation. These signals induce specific transcriptional changes, with the number of differentially expressed genes (DEGs) associated with each signal outlined.

It is clear that successful establishment and maintenance of pregnancy, along with the initiation of the mechanisms required to begin placentation, require extensive transcriptional changes and remodelling of the maternal endometrium. As found in Chapter 4, molecules implicated in early pregnancy induce differential expression of conserved miRNAs that emerged on the therian and

eutherian stem lineages. It may therefore be hypothesized that these differentially expressed miRNAs may help to modulate transcriptional changes in early pregnancy.

miRNAs have been extensively studied in the context of pregnancy (Kuokkanen *et al.*, 2010; Ponsuksili *et al.*, 2014) and established as important regulatory molecules. The targets of these miRNAs and the dynamics of the regulatory interaction are necessary elements in the analysis of miRNA regulation. Specific miRNA/target interactions have been identified. miR-199b-5p was found to be upregulated in the decidua of aborted pregnancies in humans. A predicted regulatory interaction between miR-199b-5p and serum/glucocorticoid regulated kinase 1 (SGK1). Using qPCR, immunohistochemistry and miRNA inhibitors, SGK1 was found to be expressed in endometrial glandular epithelial cells, with decreased expression in the decidua of aborted pregnancies. The interaction between miR-199b-5p and SGK1 was investigated using miRNA mimics and inhibitors, finding a significant inverse relationship between miR-199b-5p and SGK1 expression, indicating that miR-199b-5p may regulate SGK1 expression and modulate its role in the maintenance of pregnancy (Wang *et al.*, 2015). miRNA/target interactions have also been investigated in the context of pregnancies affected by fetal growth restriction. Using RT-qPCR, the regulatory interaction of miR-141 and zinc finger pleomorphic adenoma gene 1 (PLAG1) was investigated in the placentas of FGR pregnancies, with miR-141 being significantly upregulated in FGR pregnancies and PLAG1 downregulated. Furthermore, PLAG1 was found to positively regulate insulin-like growth factor 1 (IGF-1), indicating aberrant miR-141 expression contributes to the development of FGR (Tang *et al.*, 2013). miRNA/target interactions are also implicated in preeclampsia, where elevated levels of miR-519d-3p are observed. Using RT-qPCR and luciferase assays, miR-519d-3p was found to negatively regulate matrix metalloproteinase-2 (MMP-2), inhibiting trophoblast cell invasion (Ding *et al.*, 2015). Within the set of miRNAs that evolved on the Therian and Eutherian stem lineages, miR-378a-5p is experimentally confirmed to target Nodal, a member of the transforming growth factor beta (TGF-β) family of proteins, which suppresses trophoblast cell proliferation and induces placental apoptosis. Using luciferase reporter assays and RT-qPCR, miR-378a-5p was found to negatively regulate Nodal, suppressing its expression in human trophoblast cells (Luo *et al.*, 2012).

With miRNA/target interactions found to play necessary roles in pregnancy, evaluating targets that may be implicated in the evolution of Eutherian pregnancy could reveal further candidates for future analyses. The evolution of placentation and Eutherian pregnancy is hypothesized to have been facilitated by the development of novel regulatory networks (Wagner, 2015). Using phylogeny-based methods, a conserved regulatory network of miRNAs that emerged on the stem Therian and Eutherian lineages; and genes that underwent selective pressure on the stem Eutherian lineage; was identified, with the implicated positively selected genes being enriched for stem lineage miRNA binding sites (Chapter 3). These positively selected genes, therefore, may be candidates in the analysis of miRNA/target interactions in the establishment and maintenance of pregnancy.

5.3 Materials and Methods

Bovine and human endometrial cells were chosen to assess transcriptional changes in conserved positively selected genes induced by stem lineage miRNAs regulated by molecules related to early pregnancy events (Table 5.2). These two species were chosen as they display differing implantation and placentation morphologies; and have been extensively studied in the context of early pregnancy and placentation. In addition, their differing morphology may allow us to investigate conserved or species-specific actions of the conserved genes that underwent positive selection on the stem Eutherian lineage. Unless otherwise stated, all chemicals were sourced from Sigma Aldrich.

Table 5.2 Stem lineage miRNAs and the in vitro models of early pregnancy in (i) human immortalized Ishikawa endometrial epithelial cells, (ii) bovine endometrial explants, (iii) bovine endometrial epithelial cells and (iv) bovine endometrial stromal cells.

Species	Early pregnancy model	Differentially expressed stem lineage miRNAs
Human	bCAPG	miR-127-3p, miR-151a-3p, miR-188-5p
	bP4HB	miR-151a-5p, miR-185-5p, miR-378a-3p, miR-532-5p
	P4	miR-340-5p, miR-542-3p, miR-671-5p
Bovine	IFNT in endometrial explants	miR-28-3p
	bCAPG in endometrial epithelial cells	miR-324-5p, miR-331-3p, miR-505-5p
	bCAPG in endometrial stromal cells	miR-660-5p
	bP4HB in endometrial epithelial cells	miR-185-5p
	bP4HB in endometrial stromal cells	miR-188-5p
	P4 in endometrial epithelial cells	miR-505-5p
	P4 in endometrial stromal cells	miR-505-5p

In Chapter 4, the expression of conserved miRNAs that originated on the stem Therian and Eutherian lineages were investigated in in vitro models of molecules associated with early pregnancy in human Ishikawa immortalised endometrial epithelial cells and in bovine explants and endometrial epithelial and stromal cells. miRNAs were determined to be significantly differentially expressed when $p \leq 0.05$.

In order to investigate the emergence of a conserved regulatory network on the stem Eutherian lineage, TargetScan (Agarwal *et al.*, 2015) was used to predicted targets for conserved miRNAs that emerged on the therian and eutherian stem lineages, using multiple sequence alignments (MSAs) of 3'UTR regions of eutherian genomes and seed sequences from stem lineage miRNAs. TargetScan output was filtered by the level of complementary binding between the miRNA seed region and target site. Targets with ≥7mer complementary binding (where the target transcript 3'UTR matches positions 2-8 of the mature miRNA seed region; matches positions 2-7 of the seed region, followed by an 'A'; or matches positions 2-8 of the seed region, followed by an 'A') to

the seed region were selected, with the high number of consecutive matches indicative of a efficient binding site (Bartel, 2009).

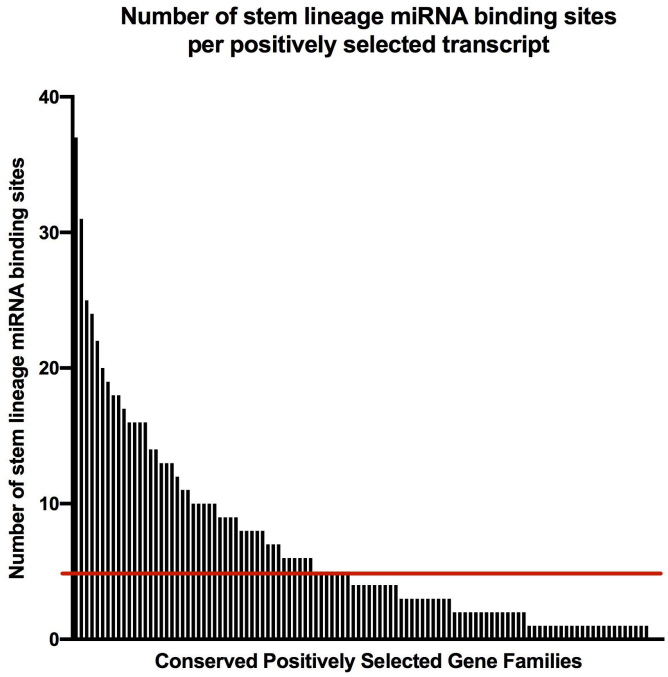


Figure 5.1 Number of stem lineage miRNA binding sites per positively selected transcript. *Targets of stem lineage miRNAs regulated by human and bovine in vitro models of early pregnancy events were predicted using TargetScan ADDIN CSL_CITATION {"citationItems":[{"id":"ITEM-1","itemData":{"DOI":"10.7554/eLife.05005","ISBN":"2050-084X","ISSN":"2050084X","PMID":"26267216","abstract":"MicroRNA targets are often recognized through pairing between the miRNA seed region and complementary sites within target mRNAs, but not all of these canonical sites are equally effective, and both computational and in vivo UV-crosslinking approaches suggest that many mRNAs are targeted through non-canonical interactions. Here, we show that recently reported non-canonical sites do not mediate repression despite binding the miRNA, which indicates that the vast majority of functional sites are canonical. Accordingly, we developed an improved quantitative model of canonical targeting, using a compendium of experimental datasets that we pre-processed to minimize confounding biases. This model, which considers site type and another 14 features to predict the most effectively targeted mRNAs, performed significantly better than existing models and was as informative as the best high-throughput in vivo crosslinking approaches. It drives the latest version of*

These 36 genes were then filtered using human (Hume *et al.*, unpublished) and bovine RNA-sequencing data (Tinning *et al.*, 2020) to identify their baseMean, or the mean of the normalised counts of all samples, determined using DESeq2 (Love, Huber and Anders, 2014). Primers were designed for genes with i) ≥30 baseMean, ii) TargetScan predicted interactions with miRNAs regulated in early pregnancy events (see Chapter 4), and iii) a RefSeq ID (O'Leary *et al.*, 2016). The baseMean minimum was set at ≥30, with 30 being the lowest baseMean with a successful primer validation determined during primer validation tests prior to this analysis. The RefSeq ID was used in order to ensure a well annotated sequence for primer design. The primers validated using these parameters are detailed in Tables 5.3 and 5.4.

Table 5.3 Candidate human genes for primer design.

Gene	Transcript	baseMean >30	RefSeq ID	Targeted	Designed
COX15	ENST00000016171	Yes	Yes	Yes	Yes
TCFL5	ENST00000217162	Yes	Yes	Yes	Yes
RP2	ENST00000218340	Yes	Yes	Yes	Yes
FA2H	ENST00000219368	No	Yes	Yes	
EEPD1	ENST00000242108	Yes	Yes	Yes	Yes
USPL1	ENST00000255304	Yes	Yes	Yes	Yes
MYDGF	ENST00000262947	Yes	Yes	Yes	Yes
EXPH5	ENST00000265843	Yes	Yes	Yes	Yes
NOP9	ENST00000267425	Yes	Yes	Yes	Yes
N4BP3	ENST00000274605	Yes	Yes	Yes	Yes
MLIP	ENST00000274897	No	Yes	Yes	
HIRIP3	ENST00000279392	Yes	Yes	Yes	Yes
TRH	ENST00000302649	Yes	Yes	Yes	Yes
TERT	ENST00000310581	Yes	Yes	Yes	Yes
FJX1	ENST00000317811	Yes	Yes	Yes	Yes
CLSPN	ENST00000318121	Yes	Yes	Yes	Yes
C7orf26	ENST00000344417	Yes	Yes	No	Yes
NSMCE1	ENST00000361439	Yes	Yes	Yes	Yes
DPH5	ENST00000370109	Yes	Yes	Yes	Yes
BSND	ENST00000371265	No	Yes	Yes	
JCAD	ENST00000375377	Yes	Yes	Yes	Yes
C19orf54	ENST00000378313	No	Yes	No	
MBTPS2	ENST00000379484	Yes	Yes	Yes	Yes
CCDC171	ENST00000380701	Yes	Yes	Yes	Yes
C5orf51	ENST00000381647	Yes	Yes	Yes	Yes
AKAP5	ENST00000394718	Yes	Yes	Yes	Yes
ZBTB40	ENST00000404138	Yes	Yes	Yes	Yes
NOLC1	ENST00000405356	Yes	Yes	Yes	Yes
MOGS	ENST00000409065	Yes	Yes	Yes	Yes
MDM1	ENST00000430606	Yes	Yes	Yes	Yes
TMEM213	ENST00000442682	Yes	Yes	Yes	Yes
NFRKB	ENST00000446488	Yes	Yes	Yes	Yes
KIAA0408	ENST00000483725	Yes	Yes	Yes	Yes
NUP88	ENST00000573584	Yes	Yes	Yes	Yes
C17orf75	ENST00000577809	Yes	Yes	Yes	Yes
RABAC1	ENST00000601891	Yes	Yes	Yes	Yes

Using RNA-sequencing data, candidate positively selected human genes were investigated for their baseMean. Primers were designed for transcripts with a baseMean >30, were predicted to interact with biologically regulated stem lineage miRNAs (indicated under Targeted field) and with a RefSeq ID. Using serially diluted cDNA, efficiency of primers was investigated. Primers were selected if they had (i) efficiency between 90%-

120% and (ii) dissociation curve showing a single, pure amplicon. If the primer was designed, it is stated under "Designed". If the primer did not fulfil one or more of the criteria, it was not tested and the corresponding field "Designed" is filled in in black.

Table 5.4 Candidate bovine genes for primer design.

Gene	Transcript	baseMean >30	RefSeqID	Targeted	Efficiency
COX15	ENSBTAT00000064176	Yes	No	Yes	Yes
TCFL5	ENSBTAT00000063242	Yes	No	Yes	
RP2	ENSBTAT00000049336	Yes	Yes	Yes	Yes
FA2H	ENSBTAT00000016212	No	Yes	Yes	
EEPD1	ENSBTAT00000033712	Yes	Yes	Yes	Yes
USPL1	ENSBTAT00000001562	Yes	No	Yes	Yes
MYDGF	ENSBTAT00000024823	Yes	Yes	Yes	Yes
EXPH5	ENSBTAT00000010864	Yes	No	Yes	
NOP9	ENSBTAT00000047185	Yes	Yes	Yes	Yes
N4BP3	ENSBTAT00000008355	Yes	Yes	Yes	Yes
MLIP	ENSBTAT00000077761	No	Yes	Yes	
HIRIP3	ENSBTAT00000006619	Yes	Yes	Yes	Yes
TRH	ENSBTAT00000006982	No	Yes	Yes	
TERT	ENSBTAT00000016685	No	Yes	Yes	
FJX1	ENSBTAT00000065157	No	No	Yes	
CLSPN	ENSBTAT00000074281	Yes	Yes	Yes	Yes
C7orf26	ENSBTAT00000028313	Yes	Yes	No	
NSMCE1	ENSBTAT00000055230	Yes	Yes	Yes	Yes
DPH5	ENSBTAT00000025327	Yes	Yes	Yes	Yes
BSND	ENSBTAT00000017615	No	Yes	Yes	
JCAD	ENSBTAT00000001595	No	Yes	Yes	
C19orf54	ENSBTAT00000043989	No	No	Yes	
MBTPS2	ENSBTAT00000086968	Yes	Yes	Yes	Yes
CCDC171	ENSBTAT00000065456	No	No	Yes	
C5orf51	ENSBTAT00000006486	No	No	Yes	
AKAP5	ENSBTAT00000012705	No	Yes	Yes	
ZBTB40	ENSBTAT00000021869	Yes	Yes	Yes	Yes
NOLC1	ENSBTAT00000009775	Yes	Yes	No	
MOGS	ENSBTAT00000077351	Yes	No	Yes	
MDM1	ENSBTAT00000026305	Yes	No	Yes	Yes
TMEM213	ENSBTAT00000001999	No	Yes	Yes	
NFRKB	ENSBTAT00000077454	Yes	No	Yes	
KIAA0408	ENSBTAT00000005116	No	No	Yes	
NUP88	ENSBTAT00000001625	Yes	Yes	Yes	Yes
C17orf75	ENSBTAT00000027955	No	No	Yes	
RABAC1	ENSBTAT00000024793	Yes	Yes	Yes	Yes

Using RNA-sequencing data, candidate positively selected bovine genes were investigated for their baseMean. Primers were designed for transcripts with a baseMean >30, were predicted to interact with biologically regulated stem lineage miRNAs (indicated under Targeted field) and with a RefSeq ID. Using serially diluted cDNA, efficiency of primers was investigated. Primers were selected if they had (i) efficiency between 90%-120% and (ii) dissociation curve showing a single, pure amplicon. If the primer was designed, it is stated under "Designed". If the primer did not fulfil one or more of the criteria, it was not tested and the corresponding field "Designed" is filled in in black.

5.3.1 Primer Design

Coding sequences for 40 conserved positively selected genes targeted by stem lineage miRNAs were accessed from Ensembl 96 (Zerbino *et al.*, 2017). SnapGene® software (from GSL Biotech; available at snapgene.com) was used to find putative primer sequences of appropriate length and melting temperature (T_m). Primer-BLASTv.3 (Ye *et al.*, 2012) was used to investigate self-complementarity, 3' self-complementarity, GC content and non-specific amplicons (measures detailed in Table 5.5).

Table 5.5 Parameters for primer design and selection of optimal available primer for a gene.

Parameter	Parameter Range
Product size	80-150bp
Melting temperature	Minimum=58, Optimal=61, Maximum=64
Exon junction span	Must span an exon-exon junction
Splice variants	Allowed
Primer length	18-22bp
GC content	40-60%
Dinucleotide repeats	≤4
Number of single baseruns	≤4
Low self-complementarity	Primer pair closest to 0
Low 3' self-complementarity	Primer pair closest to 0

Coding sequences for each of the 40 conserved positively selected miRNAs in both human and cow were accessed from Ensembl 96 (Zerbino *et al.*, 2017)*. The coding sequences were then viewed using SnapGene® in order to find candidate primer sequences of the appropriate length and T_m. Putative primer sequences were analysed using Primer-BLASTv.3* (Ye *et al.*, 2012) *to obtain information on self-complementarity, 3' self-complementarity, GC content and any non-specific amplicons. Primer pairs with values within those stated in the table above were selected for validation.*

Desalted primers were ordered from Integrated DNA Technologies, UK. Primers were validated using pooled i) human Ishikawa immortalised endometrial epithelial cells, or ii) bovine endometrial stromal cells. Converted cDNA was serially diluted and gene expression evaluated in cDNA concentrations of 10ng/µl, 5ng/µl, 2.5ng/µl, 1.25ng/µl and 0.625ng/µl. Primers were considered validated when efficiency was >90% and <120% and the dissociation curve showed a single, pure amplicon (example melt curve shown in Figure 5.2).

5.3.2 Primary Cell and Explant Collection

5.3.2.1 Bovine Endometrial Explant Culture

Bovine endometrial explants were isolated from late luteal tracts as per the protocol outlined in Chapter 4.

5.3.2.2 Bovine Epithelial and Stromal Cell Isolation

Bovine epithelial and stromal cells were isolated from early and late luteal tracts as per the protocol outlined in Chapter 4.

5.3.3 Cell Culture

Ishikawa cells (ECACC: #99040201, passage 18) were incubated at 37°C and 5% CO_2 in a T75 flask. Ishikawa cells (n=3) were maintained in Gibco Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (DMEM/F-12) (Thermo Fisher Scientific, USA) with the addition of 10% (v/v) Gibco One-shot EV-depleted FBS (Thermo Fisher Scientific, USA) (charcoal stripped using 1g Sigma-Aldrich dextran-coated charcoal) and 1% Gibco GSP. Cells were passaged every 3 days, until they reached 70% confluency.

5.3.3.1 Treatment of Bovine Endometrial Explants

Bovine endometrial explants (n=3 biological replicates) were incubated in 6-well plates, in 37°C and 5% CO_2. After 4 hours, wells were replenished with 2mL Gibco RPMI 1640 media with the addition of 10% (v/v) PAA FBS Gold and 1% Sigma-Aldrich ABAM. Explants were then treated with

iv) Control.

v) vehicle control (200⌠l PBS).

vi) 100⌠l 1000ng/mL ovine Interferon Tau (oIFNT).

Treated explants were incubated for 24 hours at 37°C and 5% CO_2. After 24 hours, explants were removed from wells using sterile tweezers, then placed in sterile Eppendorfs and snap-frozen in liquid nitrogen. Frozen explants were stored at -80°C.

5.3.3.2 Treatment of Primary Bovine Epithelial and Stromal Cells with oIFNT, bCAPG and bP4HB

Cultured epithelial and stromal cells (n=3 biological replicates) were maintained until cells had reached 70% confluency. Live cells were counted using Trypan Blue and a haemocytometer. Epithelial cells were diluted to 100,000 cells/mL and stromal cells were diluted to 25,000 cells/mL using Gibco RPMI 1640 media with 10% (v/v) PAA FBS Gold and 1% Sigma-Aldrich ABAM. Cells were plated at 2mL/well in 6 well plates and treated with

vi) Control.

vii) Vehicle control (20⌠l PBS).

viii) 100⌠l 1000ng/mL oIFNT.

ix) 100⌠l 1000ng/mL bCAPG.

x) 100⌠l 1000ng/mL bP4HB.

Treated cells were incubated at 37°C and 5% CO_2. After 24 hours cells were trypsinized, pelleted and lysed using 700µl QIAzol Lysis Reagent. Lysed cells were transferred to a sterile labelled Eppendorf and were snap-frozen in liquid nitrogen. Frozen cells were then stored at -80°C.

5.3.3.3 Treatment of Primary Bovine Epithelial and Stromal Cells with P4

Cultured early luteal epithelial and stromal cells (n=3 biological replicates) were maintained until cells had reached 70% confluency. Live cells were counted using Trypan Blue and a haemocytometer. Epithelial cells were diluted to 100,000 cells/mL and stromal cells were diluted to 25,000 cells/mL using Gibco RPMI 1640 media with 10% (v/v) PAA FBS Gold and 1% Sigma-Aldrich ABAM. Cells were plated at 2mL/well in 6 well plates and treated with

vi) Control.

vii) Vehicle control (10⌠l ethanol).

viii) 10⌠l 0.1⌠g/mL P4.

ix) 100⌠l 1.0⌠g/mL P4

x) 100⌠l 10⌠g/mL P4.

Treated cells were incubated at 37°C and 5% CO_2. After 24 hours cells were trypsinized, pelleted and lysed using 700µl QIAzol Lysis Reagent. Lysed cells were transferred to a sterile labelled Eppendorf and were snap-frozen in liquid nitrogen. Frozen cells were then stored at -80°C.

5.3.3.4 Human Ishikawa Cell Treated with bCAPG and bP4HB

Human Ishikawa cells (n=3 technical replicates) were maintained until cells had reached 70% confluency. Live cells were counted using Trypan Blue and a haemocytometer. Ishikawa cells were diluted to 100,000 cells/mL using DMEM/F-12 media with the addition of 10% (v/v) Gibco One-shot EV-depleted FBS (charcoal stripped using 1g Sigma-Aldrich dextran-coated charcoal) and 1% Gibco GSP. Cells were plated at 2mL cells/well in a 6 well plate and treated with

v) Control.
vi) Vehicle control (20⌠l PBS).
vii) 20⌠l 10ng/mL bCAPG.
viii) 20⌠l 100ng/mL bCAPG.
ix) 20⌠l 1000ng/mL bCAPG.
x) 20⌠l 10ng/mL bP4HB.
xi) 20⌠l 100ng/mL bP4HB.
xii) 20⌠l 1000ng/mL bP4HB.

Treated cells were incubated at 37°C and 5% CO_2. After 24 hours cells were trypsinized, pelleted and lysed using 700µl QIAzol Lysis Reagent. Lysed cells were transferred to a sterile labelled Eppendorf and were snap-frozen in liquid nitrogen. Frozen cells were then stored at -80°C.

5.3.3.5 Human Ishikawa Cells Treated with P4

Human Ishikawa cells (n=3 technical replicates) were maintained until cells had reached 70% confluency. Live cells were counted using Trypan Blue and a haemocytometer. Ishikawa cells were diluted to 100,000 cells/mL using DMEM/F-12 media with the addition of 10% (v/v) Gibco One-shot EV-depleted FBS (charcoal stripped using 1g Sigma-Aldrich dextran-coated charcoal) and 1% Gibco GSP. Cells were plated at 2mL cells/well in a 6 well plate and treated with

i) Control.
ii) Vehicle control (10⌠l ethanol).
iii) 10⌠l 0.1⌠g/mL P4.
iv) 100⌠l 1.0⌠g/mL P4
v) 100⌠l 10⌠g/mL P4.

Treated cells were incubated at 37°C and 5% CO_2. After 24 hours cells were trypsinized, pelleted and lysed using 700µl QIAzol Lysis Reagent. Lysed cells were transferred to a sterile labelled Eppendorf and were snap-frozen in liquid nitrogen. Frozen cells were then stored at -80°C.

5.3.4 RNA Extractions

Total RNA was extracted from bovine endometrial explants treated with oIFNT (9 total samples); bovine endometrial epithelial cells treated with oIFNT, bCAPG and bP4HB (15 total samples); bovine endometrial epithelial cells treated with P4 in a dose-dependent manner (15 total samples); bovine endometrial stromal cells treated with oIFNT, bCAPG and bP4HB (15 total samples); bovine endometrial stromal cells treated with P4 dose-dependent manner (15 total samples); human Ishikawa immortalised endometrial epithelial cells treated with bCAPG and bP4HB in a dose-dependent manner (24 total samples); and human Ishikawa immortalised endometrial epithelial cells treated with P4 in a dose-dependent manner (15 total samples); using the miRNeasy Mini RNA Extraction kit (Qiagen, UK). Prior to bovine endometrial explant extraction, 50mg snap-frozen bovine endometrial explants were mixed with 700µl QIAzol Lysis Reagent and homogenised using a mechanical homogenizer for 10-15 seconds. Following homogenization, RNA extraction was performed under the same conditions as RNA extraction from bovine and human endometrial cells, as follows:

In the fume hood, 140µl chloroform was added to the homogenised cells. Samples were shaken for 15sec and left to rest on the bench for 2-3 minutes at room temperature. Samples were centrifuged for 15min at 12,000xg at 4°C. The upper aqueous phase was transferred to a 2ml collection tube and 1.5 volumes of ethanol was added. The sample was mixed by pipetting and the sample was added to a RNeasy Mini spin column (700µl at a time, centrifuging at 8,000xg for 15sec at room temperature, adding any remaining mixture to the column and centrifuging again). A DNase digestion was then performed by adding 350µl Buffer RWT to the spin column and centrifuging for 15sec at room temperature. DNase stock was made by adding 10µl DNase I stock solution to 70µl Buffer RDD, mixing by inverting. DNase stock was added by pipetting 80µl stock solution onto each spin column, leaving the sample to rest for 15min at room temperature. The spin column was washed by adding 350µl Buffer RWT and centrifuging for 15sec at 8,000xg at

room temperature. Flow-through was discarded and the column washed again with 500µl Buffer RPE and centrifuged for 15sec at 8,000xg at room temperature, discarding the flowthrough. RNA was eluted into a fresh collection tube using 30µl RNase-free water, centrifuging for 1min at 8,000xg at room temperature. Eluted RNA was collected and RNA content immediately quantified using the NanoDrop N1000 (Thermo Fisher Scientific, USA). Extracted RNA was stored at -80°C.

5.3.5 Reverse Transcription of Extracted RNA

Reverse transcription was performed on extracted total RNA from bovine primary epithelial and stromal cells, and human Ishikawa cells using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, UK), according to the manufacturer's instructions.

Extracted DNA was diluted to 200ng/µl using DNase/RNase free water. 10µl diluted RNA was added to 2µl 10X RT Buffer, 0.8µl 25x dNTP Mix (100mM), 2µl 10x RT Random Primers, 1µl MultiScribe™ Reverse Transcriptase, and 4.2µl DNase/RNase water, for a 20µl total volume per sample. A -RT control was prepared, substituting 1µl DNase/RNase water for MultiScribe™ Reverse Transcriptase. A +RT control was prepared, substituting 10µl DNase/RNase H_2O for RNA. Solutions were incubated at 25°C for 10 minutes to activate the reagents. The reaction was then incubated at 37°C for 120 minutes to allow reverse transcription, then heated to 85°C for 5 minutes to inactivate the reaction. Resulting cDNA was stored at -20°C.

5.3.6 Expression profiling of positively selected genes

Quantification of positively selected genes targeted by biologically regulated miRNAs was performed using LightCycler® 480 SYBR Green I Master (Roche Life Science, UK). Expression of positively selected genes were analysed in the *in vitro* model that their corresponding miRNAs were found to be differentially expressed in. The specific genes tested in each experiment are detailed in Table 5.6. Stem lineage miRNA expression in human immortalized endometrial Ishikawa cells was determined in a study of progesterone regulation in the endometrium (Hume, *et al.*, unpublished).

cDNA samples were diluted to 2.5ng/µl using DNase/RNase free water. A reaction mix was prepared for each gene; adding 5µl LightCycler® 480 SYBR Green I Master, 2.5µl DNase/RNase free water, 0.25µl forward primer, and 0.25µl reverse primer to 2µl cDNA, with a final reaction volume of 10µl added to each well. Plates were then sealed, centrifuged at 1000xg for 30 seconds and placed into a Roche LightCycler ® 480 (Roche Life Science, UK), using a 384 block. The plate was first incubated at 95°C for 5 minutes to activate the FastStart Taq DNA Polymerase. cDNA strands were denatured by incubating at 95°C for 10 seconds, followed by incubating at 61°C (optimal primer T_m) for 5 seconds to allow primer annealing. The plate was then cooled to 72°C to allow cDNA strands to extend. This was repeated for 45 cycles, followed by a melting curve performed using a 95°C incubation for 5 seconds to denature the DNA, rapid cooling and incubation at 65°C for 1 minute, followed by slow ramping of the temperature back to 97°C, with 5 fluorescence readings performed at each 1°C temperature increment. At the end of the reaction, the raw Ct values and melt curves (example in Figure 5.2) were exported from the LightCycler 480/96 software. Ct values of 36 or above were determined to be negative, as per the Qiagen Ct analysis parameters. ΔCt values were obtained using GAPDH and ACTN normalisation genes (PPIA was used instead of ACTN in RT-qPCR experiments on cells treated with bCAPG, as ACTN was found to respond to bCAPG treatment) as follows: . The ΔΔCt for each mRNA was calculated as follows: (Livak and Schmittgen, 2001). Fold change was then obtained for the construction of graphs. One-way ANOVA with multiple comparisons was performed on dCt values in GraphPad PRISM, correcting for multiple comparisons using the Tukey test, comparing all samples to all samples (control, vehicle control, treatment samples), where mRNA was determined to be differentially expressed when $p \leq 0.05$.

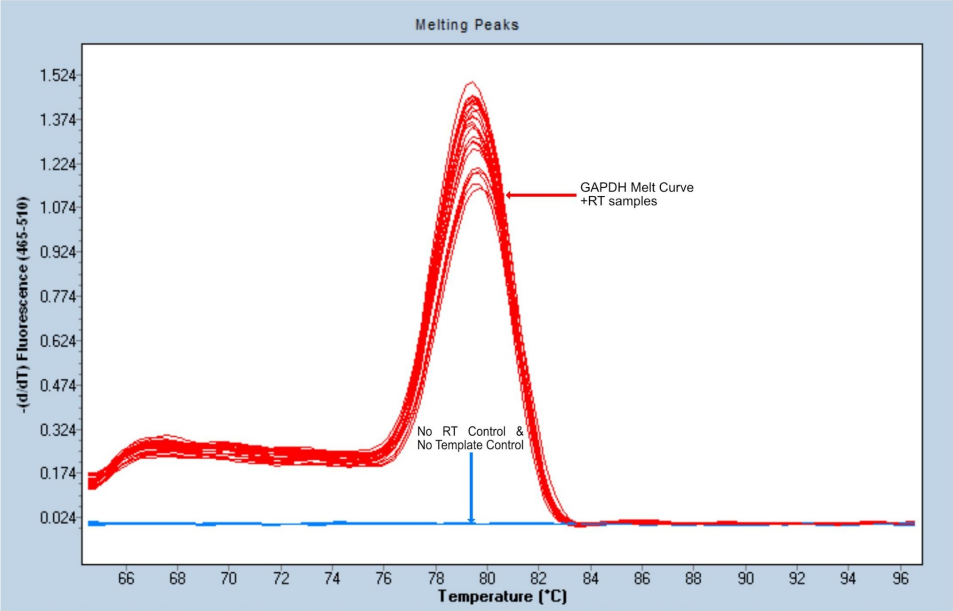


Figure 5.2 Dissociation curve for GAPDH in untreated bovine epithelial cell cDNA. *Total RNA was extracted from bovine endometrial epithelial cells and reverse transcribed into cDNA. Expression profile of GAPDH was investigated using LightCycler® SYBR Green I Master in a LightCycler® 480. The resulting amplification and dissociation curves of GAPDH in wells containing cDNA (in red). No RT controls (RNA with all reverse transcription reagents but MultiScribe™ Reverse Transcriptase) and no template controls (All reverse transcription reagents added, water substituted for RNA) are shown in blue, demonstrating that no genomic DNA contamination was present.*

Table 5.6 Breakdown of RT-qPCR experiments performed by (i) Species, (ii) *in vitro* early pregnancy model and (iii) genes analysed.

Species	Early pregnancy model	Genes tested
Human	CAPG	endonuclease/exonuclease/phosphatase family domain containing 1 (*EEPD1*), HIRA interacting protein 3 (*HIRIP3*), telomerase reverse transcriptase (*TERT*), claspin (*CLSPN*), chromosome 5 open reading frame 51 (*C5orf51*), nucleoporin 88 (*NUP88*).
	P4HB	myeloid derived growth factor (*MYDGF*), exophilin 5 (*EXPH5*), NOP9 nucleolar protein (*NOP9*), NEDD4 binding protein 3 (*N4BP3),* telomerase reverse transcriptase (*TERT*), Claspin (*CLSPN*), chromosome 7 open reading frame 26 (*C7orf26*), NSE1 homolog SMC5-SMC6 complex component (*NSMCE1*), junctional cadherin 5 associated (*JCAD*), membrane bound transcription factor peptidase site 2 (*MBTPS2*), chromosome 5 open reading frame 51 (*C5orf51*), mannosyl-oligosaccharide glucosidase (*MOGS*), rab acceptor 1 (*RABAC1*), ubiquitin specific peptidase like 1 (*USPL1*).
	P4	myeloid derived growth factor (*MYDGF*), exophilin 5 (*EXPH5*), chromosome 7 open reading frame 26 (*C7orf26*), mannosyl-oligosaccharide glucosidase (*MOGS*), claspin (*CLSPN*), nucleoporin 88 (*NUP88*), transcription factor like 5 (*TCFL5*), RP2 activator of ARL3 GTPase (*RP2*),), NOP9 nucleolar protein (*NOP9*), diphthamide biosynthesis 5 (*DPH5*), nuclear factor related to kappa-B-binding protein (*NFRKB*), chromosome 17 open reading frame 75 (*C17orf75*), NEDD4 binding protein 3 (*N4BP3*), HIRA interacting protein 3 (*HIRIP3*), telomerase reverse transcriptase (*TERT*), NSE1 homolog SMC5-SMC6 complex component (*NSMCE1*), junctional cadherin 5 associated (*JCAD*), membrane bound transcription factor peptidase site 2 (*MBTPS2*), chromosome 5 open reading frame 51 (*C5orf51*), Mdm1 nuclear protein (*MDM1*), Rab acceptor 1 (*RABAC1*).

Bovine	IFNT	No validated primers
	CAPG epithelial	endonuclease/exonuclease/phosphatase family domain containing 1 (*EEPD1*), diphthamide biosynthesis 5 (*DPH5*), claspin (*CLSPN*), HIRA interacting protein 3 (*HIRIP3*).
	CAPG stromal	No validated primers
	P4HB epithelial	endonuclease/exonuclease/phosphatase family domain containing 1 (*EEPD1*), ubiquitin specific peptidase like 1 (*USPL1*), myeloid derived growth factor (*MYDGF*), NOP9 nucleolar protein (*NOP9*), claspin (*CLSPN*), membrane bound transcription factor peptidase site 2 (*MBTPS2*), Rab acceptor 1 (*RABAC1*).
	P4HB stromal	endonuclease/exonuclease/phosphatase family domain containing 1 (*EEPD1*), HIRA interacting protein 3 (*HIRIP3*), nucleoporin 88 (*NUP88*).
	P4 epithelial	endonuclease/exonuclease/phosphatase family domain containing 1 (*EEPD1*), NOP9 nucleolar protein (*NOP9*), myeloid derived growth factor (*MYDGF*), NSE1 homolog SMC5-SMC6 complex component (*NSMCE1*), HIRA interacting protein 3 (*HIRIP3*).
	P4 stromal	endonuclease/exonuclease/phosphatase family domain containing 1 (*EEPD1*), NOP9 nucleolar protein (*NOP9*), myeloid derived growth factor (*MYDGF*), NSE1 homolog SMC5-SMC6 complex component (*NSMCE1*), HIRA interacting protein 3 (*HIRIP3*).

Using TargetScan70 (Agarwal *et al.*, 2015) *was used to predict targets for stem lineage miRNAs differentially expressed in in vitro models of early pregnancy. TargetScan results were filtered for miRNA seed sequence-target 3'UTR complementarity ≤7mer, which is indicative of stronger binding and may result in the prediction of 'biologically real' targets* (Bartel, 2009)*. Transcripts of genes that underwent positive selection on the stem Eutherian lineage were identified within the filtered TargetScan output and identified as putative primer sequences. . Primers with a baseMean >30 (determined from RNA sequencing data from Tinning et al., 2020, or Hume et al., 2020), a RefSeqID and were targeted by biologically regulated miRNAs were designed. Coding sequences were obtained from Ensembl 96* (Zerbino *et al.*, 2017) *and viewed in SnapGene® to select putative primer sequences of length 18-22bp with appropriate T_m (minimum 58°C, optimal 61°C, maximum 64°C). Using Primer-BLASTv.3* (Ye *et al.*, 2012)*, candidate sequences were assessed on GC content (optimal 40-60%), dinucleotide repeats (optimal ≤4), number of single base runs (optimal ≤4), self-self complementarity (primer pair with score closest to 0) and 3' self-complementarity (primer pair closest to 0). Selected desalted primers were ordered from Integrated DNA Technologies, UK. Primer efficiency was assessed using serially diluted cDNA from pooled untreated bovine endometrial stromal cells. Optimal primer efficiency was >90% and <120% when a single, pure amplicon was observed in the dissociation curve. Primers with a single pure amplicon and efficiency >90% and <120% were used to assess differential expression of positively selected genes targeted by stem lineage miRNAs regulated by molecules associated with early pregnancy. Genes targeted by differentially expressed miRNAs in each experiment are listed, along with official gene names.*

5.4 Results

5.4.1 Validation of candidate primers for human and bovine genes that underwent positive selection on the stem Eutherian lineage

TargetScan results for miRNAs that emerged on the stem Therian and Eutherian lineages and were found to be regulated by molecules associated with early pregnancy events in human and bovine *in vitro* models were searched for conserved genes that underwent a fixed positive selection event on the stem Eutherian lineage. 81 of these genes were found to be targeted by stem lineage miRNAs. These genes were prioritised by the number of binding sites for stem lineage miRNAs (as seen in Figure 5.1). 36 genes (with 40 transcripts) were found to have ≥5 stem lineage miRNA binding sites and were kept in the pipeline for further computational and expression analyses. This was based on the visible increase in binding sites at this point. Of the positively selected targets of stem lineage miRNAs, 36 genes (40 transcripts) had ≥5 binding sites for stem lineage miRNAs. Using RNA-sequencing data, these candidate genes were filtered for those with i) a baseMean greater than 30; ii) a RefSeq ID; and predicted interactions with miRNAs biologically regulated by molecules associated with early pregnancy in human and bovine *in vitro* models (Table 5.3 and Table 5.4). Primer validation was performed on 32 human genes and 16 bovine genes. Eighteen human genes had efficiency between 90% and 120% and showed a single, pure amplicon (example in Figure 5.5A). Nine bovine genes had efficiency between 90% and 120% and showed a single, pure amplicon (example in Figure 5.3A). Four human genes (*N4BP3, C7orf26, NSMCE1* and *JCAD*) and 1 bovine gene (*NOP9*) had efficiency between 85% and 90%, but showed amplification of a single, pure amplicon and were therefore included in further analyses. Two human genes (*COX15* and *USPL1*) showed primer dimers in the dissociation curve (example in Figure 5.3B). Eight human genes (*TRH, FJX1, CCDC171, AKAP5, ZBTB40, NOLC1, TMEM213* and *KIAA0408)* showed poor or non-specific amplification (example in Figure 5.3C). One bovine gene, *N4BP3*, had primer dimers (example in Figure 5.5B). Two bovine genes (*COX15* and *RP2*) showed poor or non-specific amplification (example in Figure 5.3C). Three bovine genes (*USPL1, ZBTB40* and *MDM1)* showed no amplification of the gene product (example in Figure 5.3D).

In total, primer pairs for 22 human genes and 10 bovine genes amplified a single, pure amplicon and had acceptable levels of efficiency. These primer pairs were used to analyse the expression of their target genes in *in vitro* models of early pregnancy in human Ishikawa immortalised endometrial epithelial cells; bovine endometrial explants; and bovine endometrial epithelial and stromal cells.

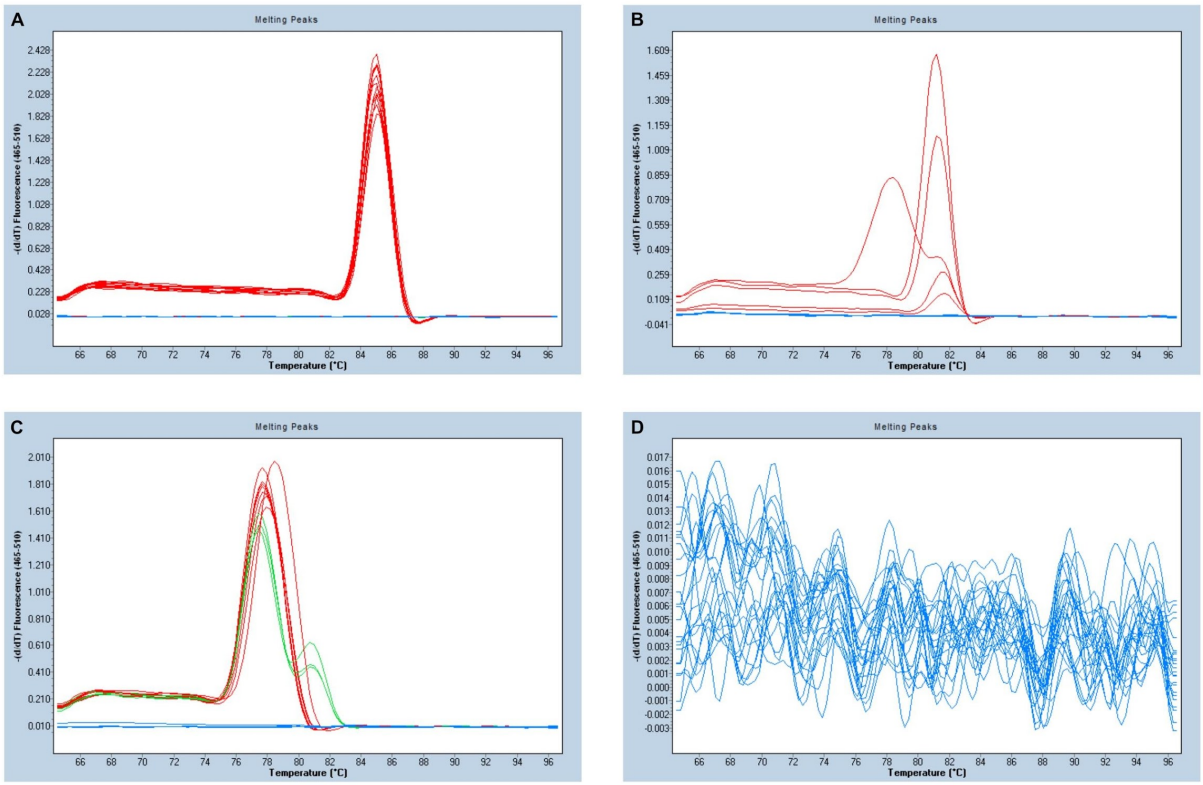


Figure 5.3 Sample dissociation curves for candidate human and bovine primers. *Total RNA was extracted from either human Ishikawa immortalised endometrial epithelial cells or bovine endometrial stromal cells and reverse transcribed into cDNA. Expression profile of candidate primers was investigated using LightCycler® SYBR Green I Master in a LightCycler® 480. Figure 5.5A shows a melt curve where a single, pure amplicon was obtained. Figure 5.5B shows a melt curve where primer dimers were formed, resulting in lower peaks and peaks formed at lower T_m. Figure 5.5C shows a melt curve with non-specific binding of primers, resulting in peaks with multiple shoulders (in green). Figure 5.5D shows a melt curve with no amplification of any gene products, with no peaks; instead low baseline fluorescence is seen in blue.*

Table 5.7 Primers for Positively Selected Human Genes

Gene	Transcript	RefSeq ID	Forward Primer	Reverse Primer	Product BP	Efficiency
COX15	ENST00000016171	NM_078470.5	CACACCGAATGTGGGGTCG	CAACAGACCCTGGAAGCAGAC	140	67%
TCFL5	ENST00000217162	XM_024451808.1	GCTGCCAAGCACCAGGATATT	GCTGCTTCTCCCACTTTAATCCA	140	116%
RP2	ENST00000218340	NM_006915.3	GACGGAAGGCTGACAAGGAG	TTTGGATCAACCTTCTCGCGCT	91	98.90%
EEPD1	ENST00000242108	NM_030636.3	TGACACTCCTGGAAAACAGCATC	CTGGTTTAGCTCCGTGCAGA	92	107%
USPL1	ENST00000255304	NM_005800.5	TCCTAGTGGAGACGCGAGTG	TTTTCAATGAACCCTGCCGC	146	77.50%
MYDGF	ENST00000262947	NM_019107.3	CGGGGCGCTGAGATTGAGTA	CCTGTGAGCCACTGCTGTTTT	117	102.50%
EXPH5	ENST00000265843	NM_015065.3	CCTCCGGCGTTTGATTTCAGT	CTGAAGTTTGCTGATCCTGTCCTT	117	118%
NOP9	ENST00000267425	NM_174913.3	CATGGCAGCTTCGTGGTCAG	CTTCTGTGCTTCAGATGATTGGG	105	93.00%
N4BP3	ENST00000274605	NM_015111.2	GAAGCCCGATGGGAGGTGT	TGCCCCGAAGTTGTGTCTTC	130	88%
HIRIP3	ENST00000279392	NM_138569.2	AAACTGACGGTCTCTGCTGG	TGGGTTGGTAGTCTTCTGACAGT	83	105%
TRH	ENST00000302649	NM_003609.5	GAAGTCAGCCCAGCCAAAGA	CTCTGCCGTTCCTCATCACTG	80	80%
TERT	ENST00000310581	NM_198253.2	TGGCACGGCTTTTGTTCAGAT	GGCATAGCTGGAGTAGTCGCT	106	98%
FJX1	ENST00000317811	NM_014344.4	CGAAGAAACTGGCTACTGGGG	TCAACGGCTCAAACGCAAGT	121	164%
CLSPN	ENST00000318121	NM_022111.4	ACCTGAAACCAACAGAGAACTGG	CTTCCTTTCCATCAGTGCCCA	137	118%
C7orf26	ENST00000344417	NM_024067.4	ACGCCTCCCAGGAGATTGAG	GGAGGAGGTTATTATGGGGCAG	138	88%
NSMCE1	ENST00000361439	NM_145080.4	CTTGGGCTCCCTGTTCGTTC	TCCCATTCCTCTAGCACGCC	119	85%
DPH5	ENST00000370109	NM_015958.3	ACGATTCTCTCAGCGGTCTCT	GGCCCTTGACTGTGATGTCC	97	93.80%
JCAD	ENST00000375377	NM_020848.4	CGGAGGCGGGGTTTTGTAAT	CTCCAGGTTTTCGTGCTCCC	134	87%
MBTPS2	ENST00000379484	NM_015884.4	CCGACTTGGTGCTGAAGTCAT	CGCCGTCCCCAACTGTAAAA	142	94%
CCDC171	ENST00000380701	NM_173550.4	ACCTCCCAGCTTCTCTGTTGT	CATGCCTTTTCCTCATTAGATGC	96	1.013E+16
C5orf51	ENST00000381647	XM_005248289.4	AGTGTGGTGAGACGAGTGGAA	AAGTGATCATCTTCACCGGCA	89	111%
AKAP5	ENST00000394718	NM_004857.3	CGACGCCCTACGTTGATCTTT	TGCCCAGTTTCTCTATGGCTTG	103	204%
ZBTB40	ENST00000404138	NM_001083621.1	ATCGGTGCTGAACAGCAAAAG	CCAGAAGAACCGACGAGAGGA	99	196%
NOLC1	ENST00000405356	NM_001284388.1	CCAAAGCGACAGGAGCTACA	TTGGCAGACTTGAGCCAGAAG	82	51%

MOGS	ENST00000409065	NM_001146158.1	ACTCTTCCAGCCCTACAGTGG	CACATCAGTCCGGTGAGGA	112	97%
MDM1	ENST00000430606	NM_020128.3	GCTTCAAGGGGCTGAGTGAA	CTTTCGTGATGCCTAATTGATCTG	135	104%
TMEM213	ENST00000442682	NM_001085429.2	TTGCTCGGCAGAAGCAAGCA	TGTGGGCAGAAGTCCACGTT	102	-6%
NFRKB	ENST00000446488	NM_001143835.1	TGGAAATAACGGGAGGGAGGA	CTCCACAGGTACTCACGCCA	82	100%
KIAA0408	ENST00000483725	NM_014702.5	GGCCTGACCTGAGGACTTCT	CTTCATCCTTCTATGGTTAGACCGC	146	128%
NUP88	ENST00000573584	NM_002532.6	AAACGTGGTCTTTGGCCTCG	AAGCAATCTCTGGTACTGGGAC	133	114%
C17orf75	ENST00000577809	NM_022344.4	CAATGCTCCCCTCTTTGCAGG	TGCCAATCTGCTTCCGCGAT	149	110%
RABAC1	ENST00000601891	NM_006423.3	CATCCTGTACTGTGTGGTGACG	GAGCATACTGATGCGCTGGG	146	111%

Primers with a baseMean >30 (determined from RNA sequencing data from Hume et al., unpublished), a RefSeqID and were targeted by biologically regulated miRNAs were designed. Coding sequences were obtained from Ensembl 96 (Zerbino *et al.*, 2017) *and viewed in SnapGene® to select putative primer sequences of length 18-22bp with appropriate T_m (minimum 58°C, optimal 61°C, maximum 64°C). Using Primer-BLASTv.3* (Ye *et al.*, 2012)*, candidate sequences were assessed on GC content (optimal 40-60%), dinucleotide repeats (optimal ≤4), number of single base runs (optimal ≤4), self-self complementarity (primer pair with score closest to 0) and 3' self-complementarity (primer pair closest to 0). Selected desalted primers were ordered from Integrated DNA Technologies, UK. Primer efficiency was assessed using serially diluted cDNA from pooled untreated human Ishikawa immortalised endometrial epithelial cells. Optimal primer efficiency was >90% and <120% when a single, pure amplicon was observed in the dissociation curve.*

Table 5.8 Primers for positively selected bovine genes.

Gene	Transcript	RefSeq ID	Forward Primer	Reverse Primer	Product BP	Efficiency
COX15	ENSBTAT00000064176	NM_001076861.1	TTTGGCATATACACAGGTGGGC	ATGAGCCAAAGGGCAACACT	120	0.00%
RP2	ENSBTAT00000049336	NM_001035403.1	ACACTTGGACAGAAAGGAATGAGG	AGTCTTTCAGATCAACTCATTGGCA	82	18%
EEPD1	ENSBTAT00000033712	NM_001034409.2	GATCGGCGGTTGTGGCTTT	CCGCAGTTGCGTTCATCCA	103	94%
USPL1	ENSBTAT00000001562	XM_002691811.5	CGAGGAGAACGTGGGTTGAA	TCTCAATGAACCCTGCCGC	114	0%
MYDGF	ENSBTAT00000024823	NM_001001164.1	GAGGGACCAATGAGAAGTGGC	CACCTCTGCCTTGAACTGCG	125	118%
NOP9	ENSBTAT00000047185	NM_001103265.1	GCTGTTGGAGGCATTCCACT	CTCCTCCGCCAGTCCATAGT	109	87%
N4BP3	ENSBTAT00000008355	NM_001205742.1	ACTACCAGGAACACCCTCGG	ATGGGCGAATGCGACACTTG	87	177%
HIRIP3	ENSBTAT00000006619	NM_001192296.1	GATGCAGGTGGATGAAGCGA	TTTCTCTCTGGGTCACGGCA	96	104%
CLSPN	ENSBTAT00000074281	NM_001192177.1	GTGAAGTGGTTTCCGAGGTTCA	CCCTGTCCACTGTCTGAAGG	91	120%
NSMCE1	ENSBTAT00000055230	NM_001035406.1	TACGCGCTGGTGAATCTTGC	CGCAAAGCCAGTGTCTGAGT	123	96%
DPH5	ENSBTAT00000025327	NM_001076821.1	GGAGAAGAACCAGCAGTCACC	TGCAATGGTCCTCCCAAGTC	130	102%
MBTPS2	ENSBTAT00000086968	NM_001075981.1	AGTTGCTGAGGACTCACCTG	TTGGGGCTCATAGGTGATGGT	136	97%
ZBTB40	ENSBTAT00000021869	NM_001206061.1	GCAGTGCAGCCAAAAGGGAG	CCAGGAACTCCAAAATGGGTCT	118	-99%
MDM1	ENSBTAT00000026305	XM_010805108.3	ATAAAGCTGGGGCTTGGACAC	CTTTAACCTCCGCATACCACAT	84	0%
NUP88	ENSBTAT00000001625	NM_001034222.1	GAGGAGATTCAGCGGAGGGT	TCCCGTAGACTTTTCCTCTCCT	98	98%
RABAC1	ENSBTAT00000024793	NM_001075575.1	CATCCTGTACTGTGTGGCGACG	CATACTGATGGGCAGGGCTCA	143	116%

Primers with a baseMean >30 (determined from RNA sequencing data from Tinning et al., 2020), a RefSeqID and were targeted by biologically regulated miRNAs were designed. Coding sequences were obtained from Ensembl 96 (Zerbino *et al.*, 2017) *and viewed in SnapGene® to select putative primer sequences of length 18-22bp with appropriate T_m (minimum 58°C, optimal 61°C, maximum 64°C). Using Primer-BLASTv.3* (Ye *et al.*, 2012)*, candidate sequences were assessed on GC content (optimal 40-60%), dinucleotide repeats (optimal ≤4), number of single base runs (optimal ≤4), self-self complementarity (primer pair with score closest to 0) and 3' self-complementarity (primer pair closest to 0). Selected desalted primers were ordered from Integrated DNA Technologies, UK. Primer efficiency was assessed using serially diluted cDNA from pooled untreated bovine endometrial stromal cells. Optimal primer efficiency was >90% and <120% when a single, pure amplicon was observed in the dissociation curve.*

5.4.2 Expression analysis of bovine endometrial explant treated with oIFNT

One candidate gene targeted by miR-28-3p (found to be significantly upregulated in bovine endometrial explants treated with oIFNT), *RP2*, had i) a baseMean ≥30 and ii) a RefSeq ID. When primer validation was attempted, RP2 showed low efficiency (18%) and poor amplification of the gene product. Therefore, no expression analysis could be attempted using this *in vitro* model.

5.4.3 Expression analysis of bovine endometrial epithelial cells treated with bCAPG

No significant differences were found in positively selected genes targeted by biologically regulated stem lineage miRNAs in bovine endometrial epithelial cells treated with bCAPG. mRNA expression levels are shown in Figure 5.4.

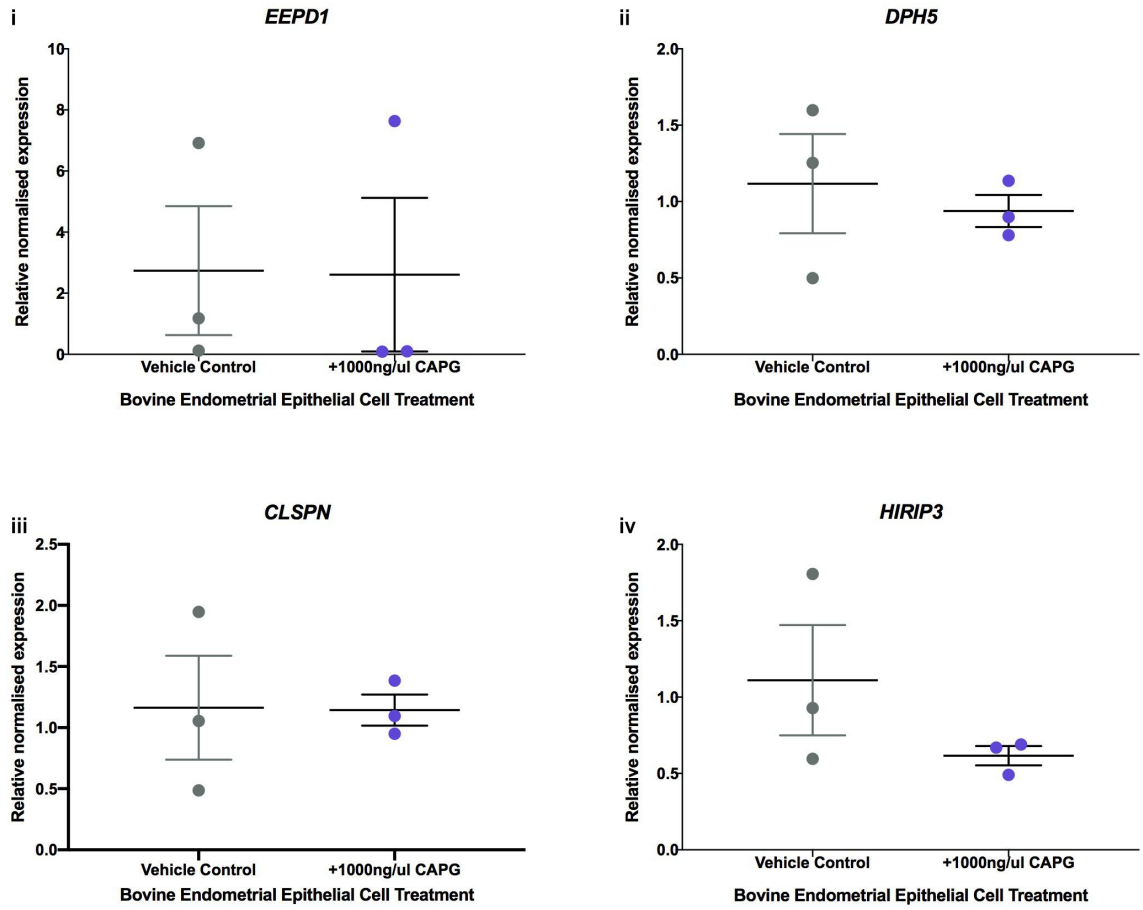


Figure 5.4 Expression of candidate positively selected genes in bovine endometrial epithelial cells treated with bCAPG. *Changes in expression of conserved positively selected genes* ***(i)*** *EEPD1,* ***(ii)*** *DPH5,* ***(iii)*** *CLSPN and* ***(iv)*** *HIRIP3 in primary bovine endometrial epithelial cells treated with* ***a)*** *vehicle control and* ***b)*** *1000ng/µl bCAPG. No significant (p≤0.05) differential expression was determined in response to 1000ng/µl bCAPG.*

5.4.4 Expression analysis of bovine endometrial epithelial cells treated with bP4HB

No significant differences were found in positively selected genes targeted by biologically regulated stem lineage miRNAs in bovine endometrial epithelial cells treated with bP4HB. mRNA expression levels are shown in Figure 5.5.

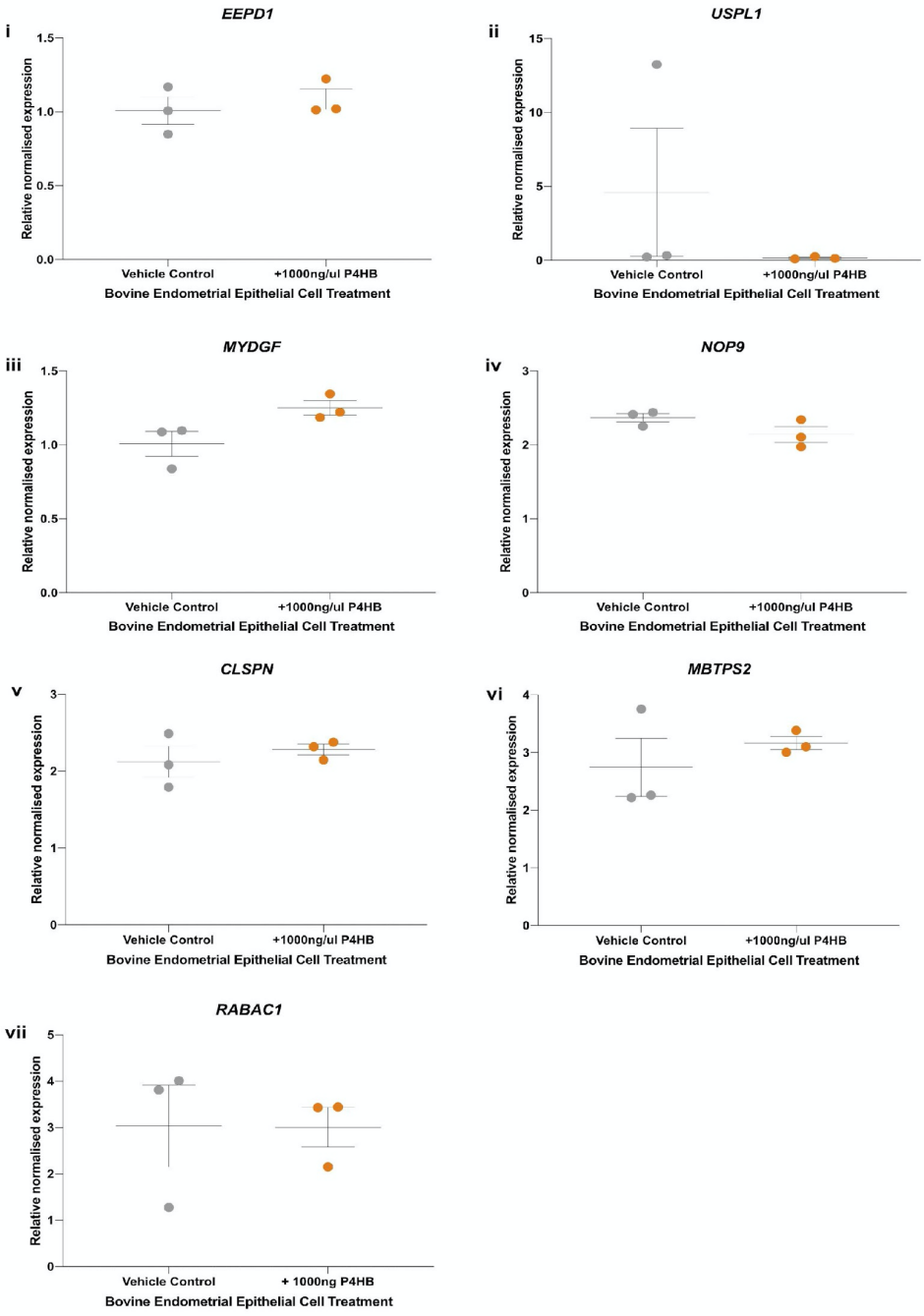


Figure 5.5 Expression of candidate positively selected genes in bovine endometrial epithelial cells treated with bP4HB. *Changes in expression of conserved positively selected genes* ***(i)*** *EEPD1,* ***(ii)*** *USPL1,* ***(iii)*** *MYDGF,* ***(iv)*** *NOP9,* ***(v)*** *CLSPN,* ***(vi)*** *MBTPS2 and* ***(vii)*** *RABAC1 in primary bovine endometrial epithelial cells treated with* ***a)*** *vehicle control and* ***b)*** *1000ng/µl bP4HB. No significant (p≤0.05) differential expression was determined in response to 1000ng/µl bP4HB.*

5.4.5 Expression analysis of bovine endometrial epithelial cells treated with P4

No significant differences were found in positively selected genes targeted by biologically regulated stem lineage miRNAs in bovine endometrial epithelial cells treated with P4. mRNA expression levels are shown in Figure 5.6.

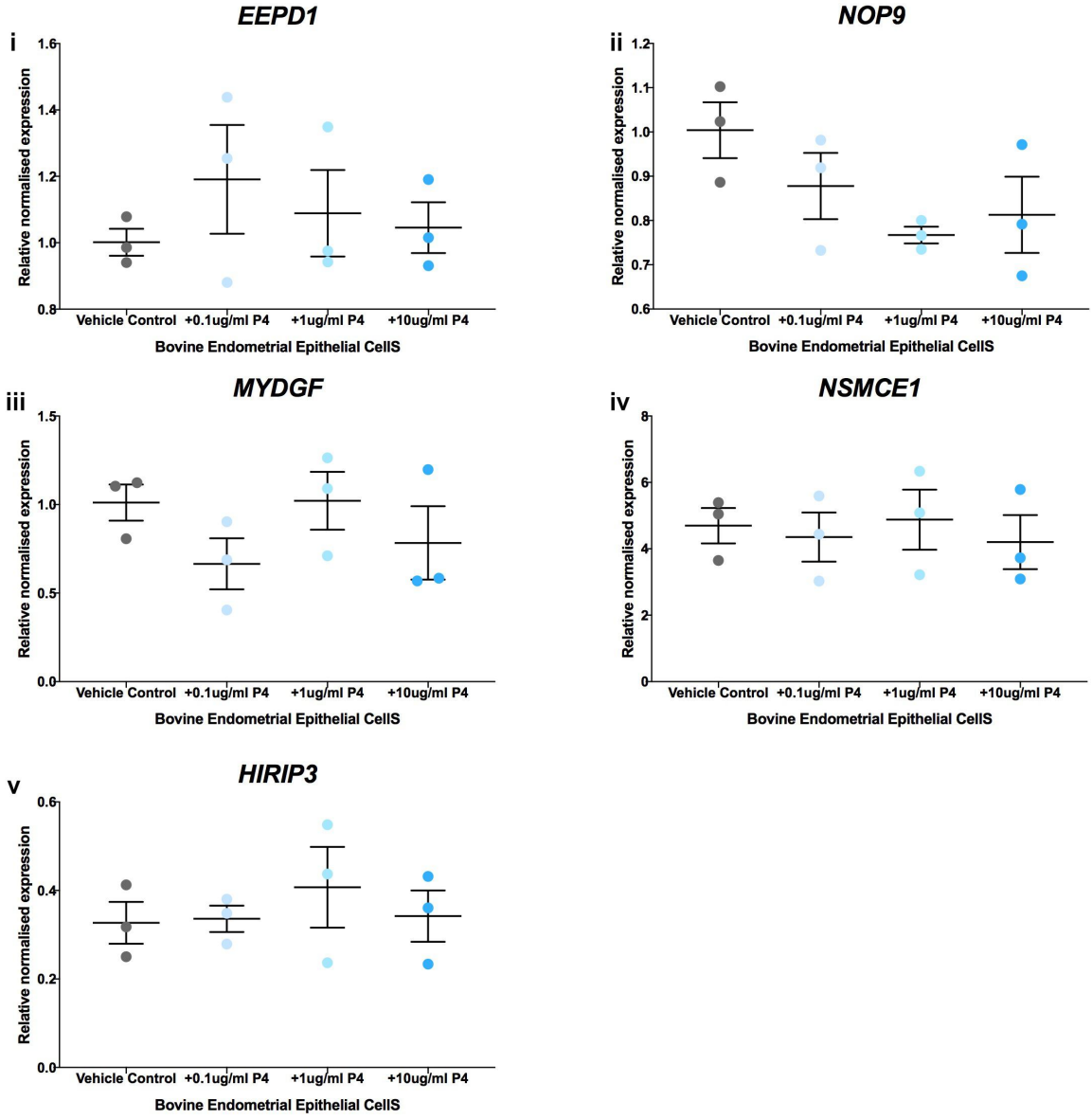


Figure 5.6 Expression of candidate positively selected genes in bovine endometrial epithelial cells treated with P4. *Changes in expression of conserved positively selected genes* ***(i)*** *EEPD1,* ***(ii)*** *NOP9,* ***(iii)*** *MYDGF,* ***(iv)*** *NSMCE1 and* ***(v)*** *HIRIP3, in primary bovine endometrial epithelial cells treated with* ***a)*** *vehicle control,* ***b)*** *0.1µg/ml,* ***c)1****µg/ml and* ***d)****10µg/ml P4. No significant (p≤0.05) differential expression was determined in response to dose-dependent treatment with P4.*

5.4.6 Expression analysis of bovine endometrial stromal cells treated with bCAPG

Two candidate genes targeted by miR-660-5p (found to be significantly upregulated in bovine endometrial stromal cells treated with bCAPG), RP2 and N4BP3, had i) a baseMean ≥30 and ii) a RefSeq ID. When primer validation was attempted, neither gene showed optimal levels of efficiency (N4BP3=177%, ZBTB40=-99%), or adequate amplification of the gene product. Therefore, no expression analysis could be attempted using this *in vitro* model.

5.4.7 Expression analysis of bovine endometrial stromal cells treated with bP4HB

No significant differences were found in positively selected genes targeted by biologically regulated stem lineage miRNAs in bovine endometrial epithelial cells treated with bP4HB. mRNA expression levels are shown in Figure 5.7.

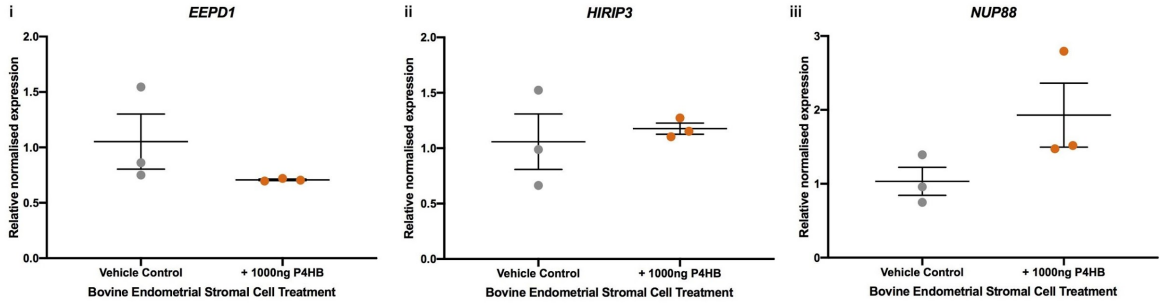


Figure 5.7 Expression of candidate positively selected genes in bovine endometrial stromal cells treated with bP4HB. *Changes in expression of conserved positively selected genes* ***(i)*** *EEPD1,* ***(ii)*** *HIRIP3 and* ***(iii)*** *NUP88 in primary bovine endometrial stromal cells treated with* ***a)*** *vehicle control and* ***b)*** *1000ng/µl bP4HB. No significant (p≤0.05) differential expression was determined in response to 1000ng/µl bP4HB.*

5.4.8 Expression analysis of bovine endometrial stromal cells treated with P4

One gene, non-structural maintenance of chromosomes element 1 (*NSMCE1*) was found to be significantly downregulated in bovine endometrial stromal cells treated with 1µg/mL P4 when compared to vehicle control and treatment with 0.1 1µg/mL P4. mRNA expression levels are shown in Figure 5.8.

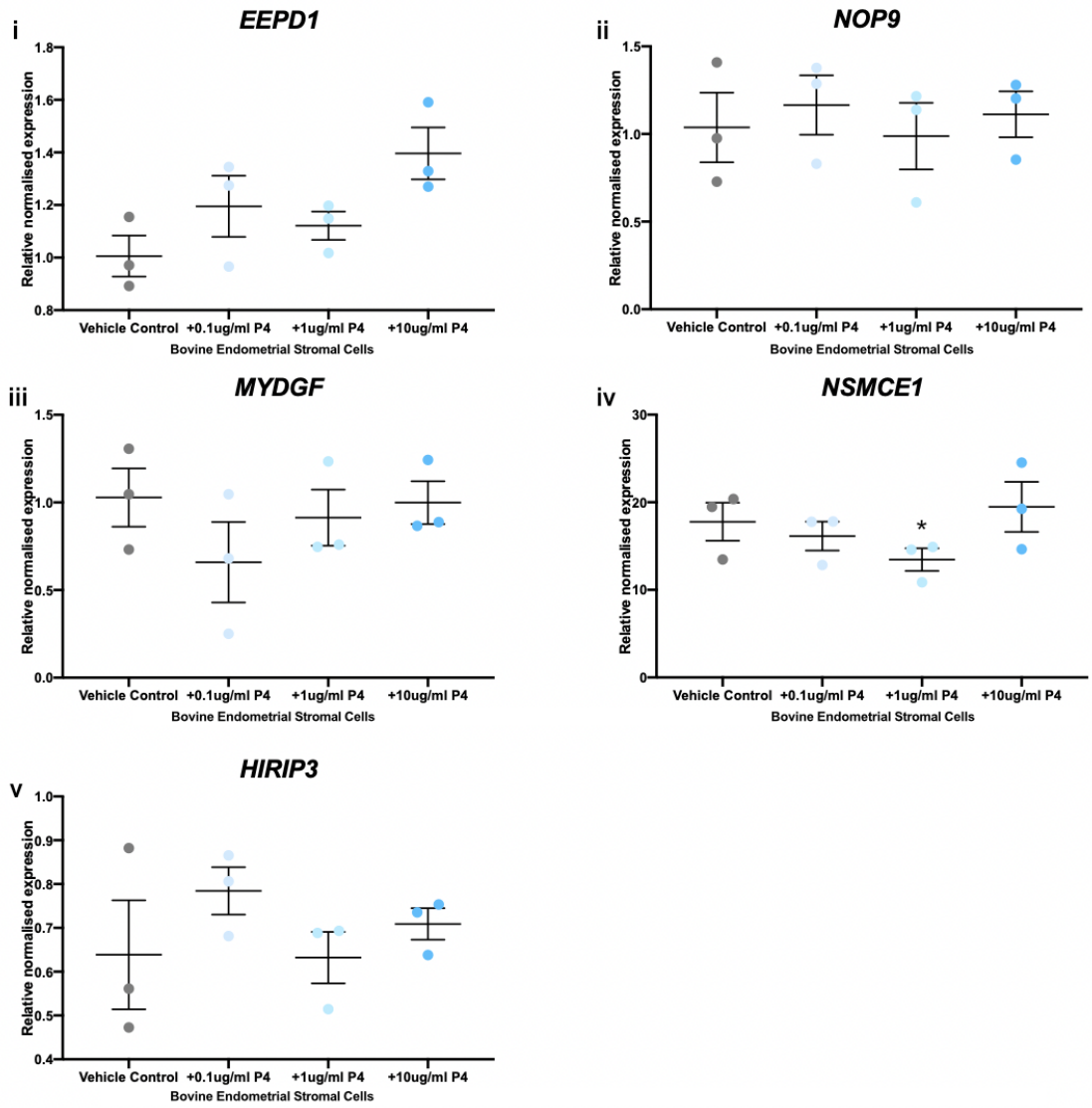


Figure 5.8 Expression of candidate positively selected genes in bovine endometrial stromal cells treated with P4. *Changes in expression of conserved positively selected genes* ***(i)*** *EEPD1,* ***(ii)*** *NOP9,* ***(iii)*** *MYDGF,* ***(iv)*** *NSMCE1 and* ***(v)*** *HIRIP3, in primary bovine endometrial stromal cells treated with* ***a)*** *vehicle control and* ***b)*** *1000ng/µl bP4HB. Significant differences in expression values determined when p≤0.05 are depicted by an asterisk (*).*

5.4.9 Expression analysis of human Ishikawa immortalised endometrial epithelial cells treated with bCAPG

No significant differences were found in positively selected genes targeted by biologically regulated stem lineage miRNAs in human Ishikawa immortalised endometrial epithelial cells treated with bCAPG. mRNA expression levels are shown in Figure 5.9.

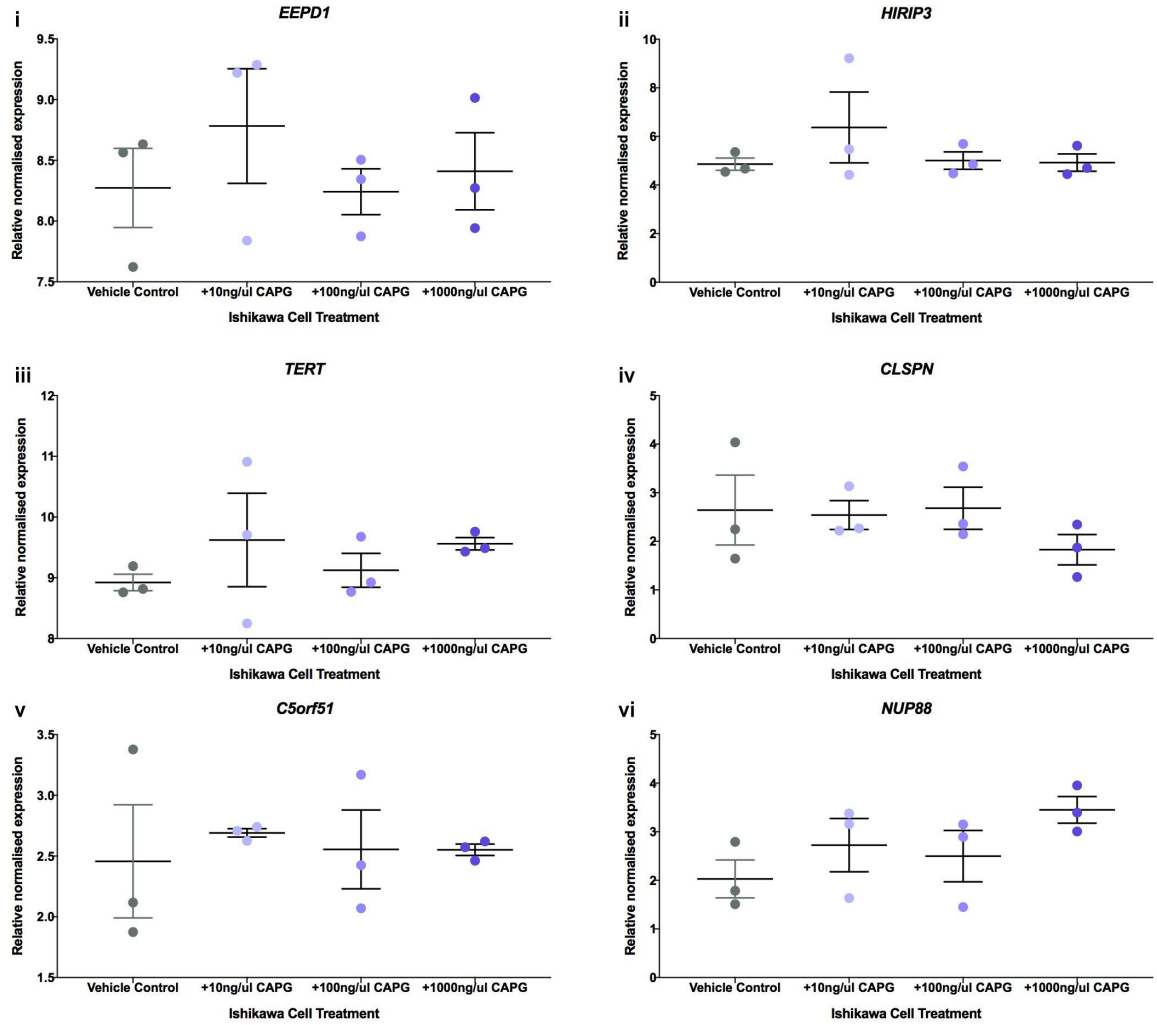


Figure 5.9 Expression of candidate positively selected genes in human Ishikawa immortalised endometrial epithelial cells treated with bCAPG. *Changes in expression of conserved positively selected genes* ***(i)*** *EEPD1,* ***(ii)*** *HIRIP3,* ***(iii)*** *TERT,* ***(iv)*** *CLSPN,* ***(v)*** *C5orf5, and* ***(vi)*** *NUP88 in human Ishikawa immortalised endometrial epithelial cells treated with* ***a)*** *vehicle control,* ***b)*** *10ng/µl bCAPG,* ***c)*** *100ng/µl bCAPG and* ***d)*** *1000ng/µl bCAPG.No significant (p≤0.05) differential expression was determined in response to dose-dependent treatment with bCAPG.*

5.4.10 Expression analysis of human Ishikawa immortalised endometrial epithelial cells treated with bP4HB

No significant differences were found in positively selected genes targeted by biologically regulated stem lineage miRNAs in human Ishikawa immortalised endometrial epithelial cells treated with bP4HB. mRNA expression levels are shown in Figure 5.10.

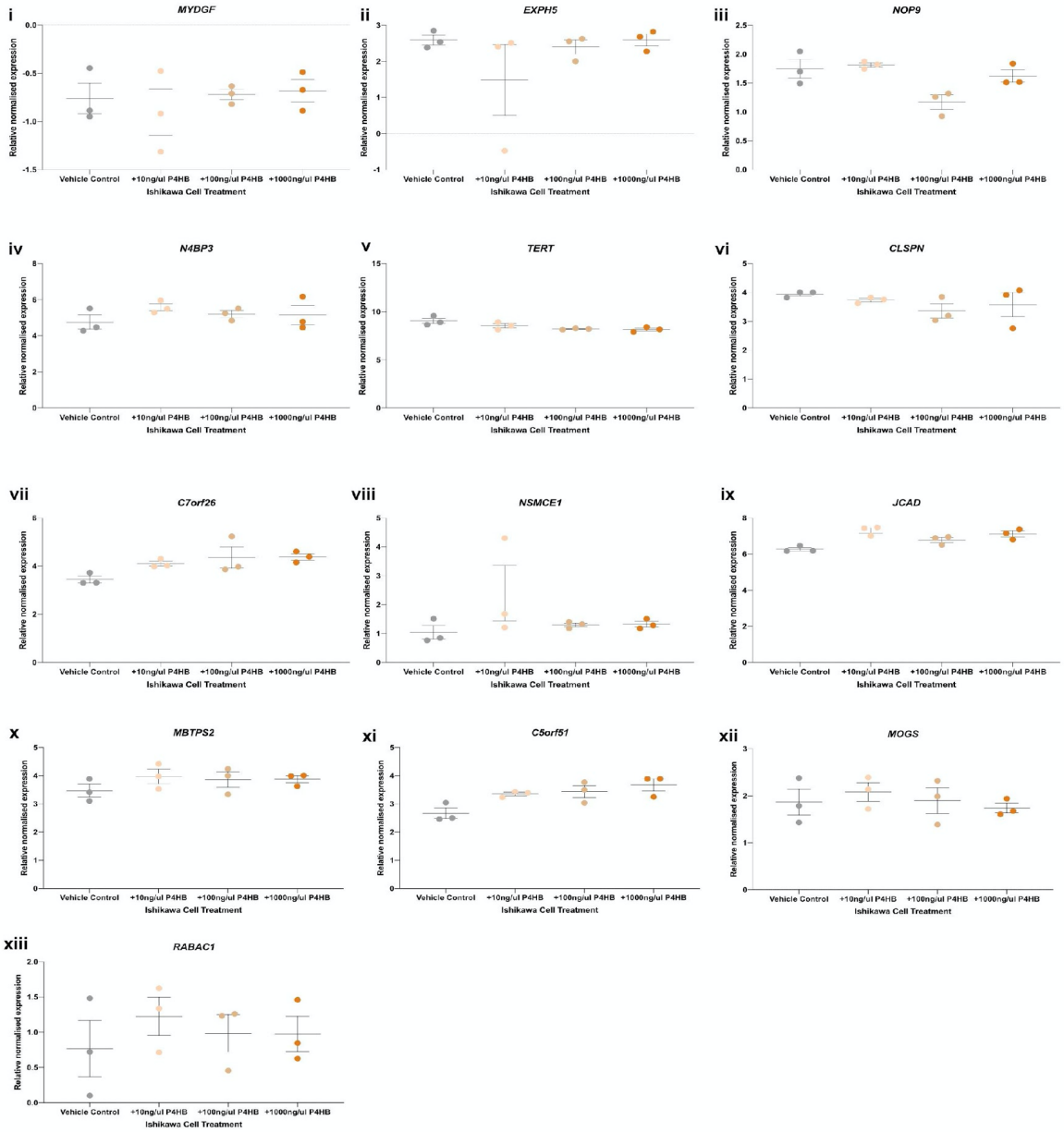
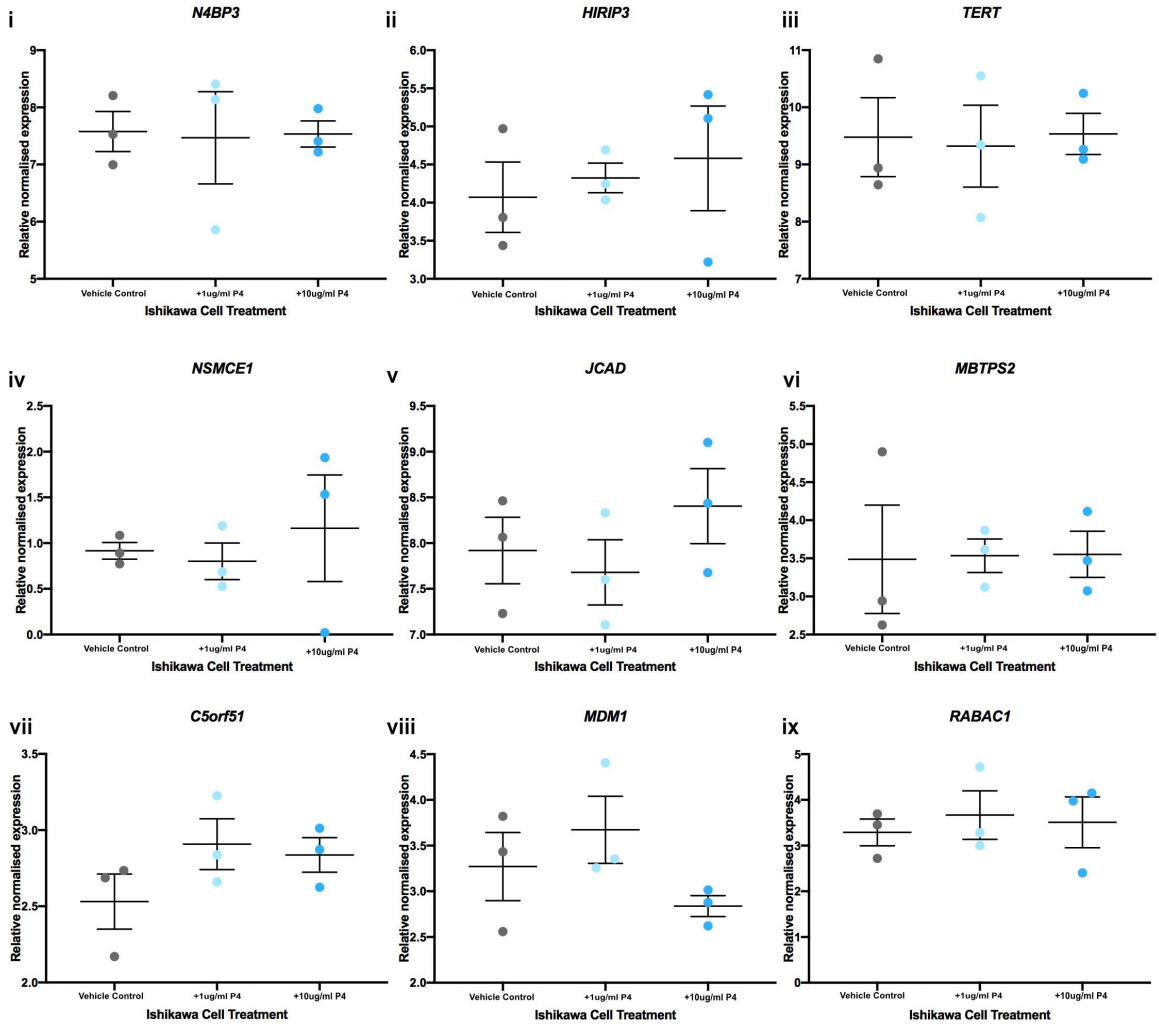


Figure 5.10 Expression of candidate positively selected genes in human Ishikawa immortalised endometrial epithelial cells treated with bP4HB. *Changes in expression of conserved positively selected genes* ***(i)*** *MYDGF,* ***(ii)*** *EXPH5,* ***(iii)*** *NOP9,* ***iv)*** *N4BP3,* ***(v)*** *TERT,* ***(vi)*** *CLSPN,* ***(vii)*** *C7orf26,* ***(viii)*** *NSMCE1,* ***(ix)*** *JCAD,* ***(x)*** *MBTPS2,* ***(xi)*** *C5orf51,* ***(xii)*** *MOGS and* ***(xiii)*** *RABAC1 human Ishikawa immortalised endometrial epithelial cells treated with* ***a)*** *vehicle control,* ***b)*** *10ng/µl bP4HB,* ***c)*** *100ng/µl bP4HB and* ***d)*** *1000ng/µl bP4HB. No significant (p≤0.05) differential expression was determined in response to dose-dependent treatment with bP4HB.*

5.4.11 Expression analysis of human Ishikawa immortalised endometrial epithelial cells treated with P4

One gene, mannosyl-oligosaccharide glucosidase (*MOGS*), was found to be significantly downregulated in human Ishikawa immortalised endometrial epithelial cells treated with 1ng/µl P4. mRNA expression levels are shown in Figure 5.11.



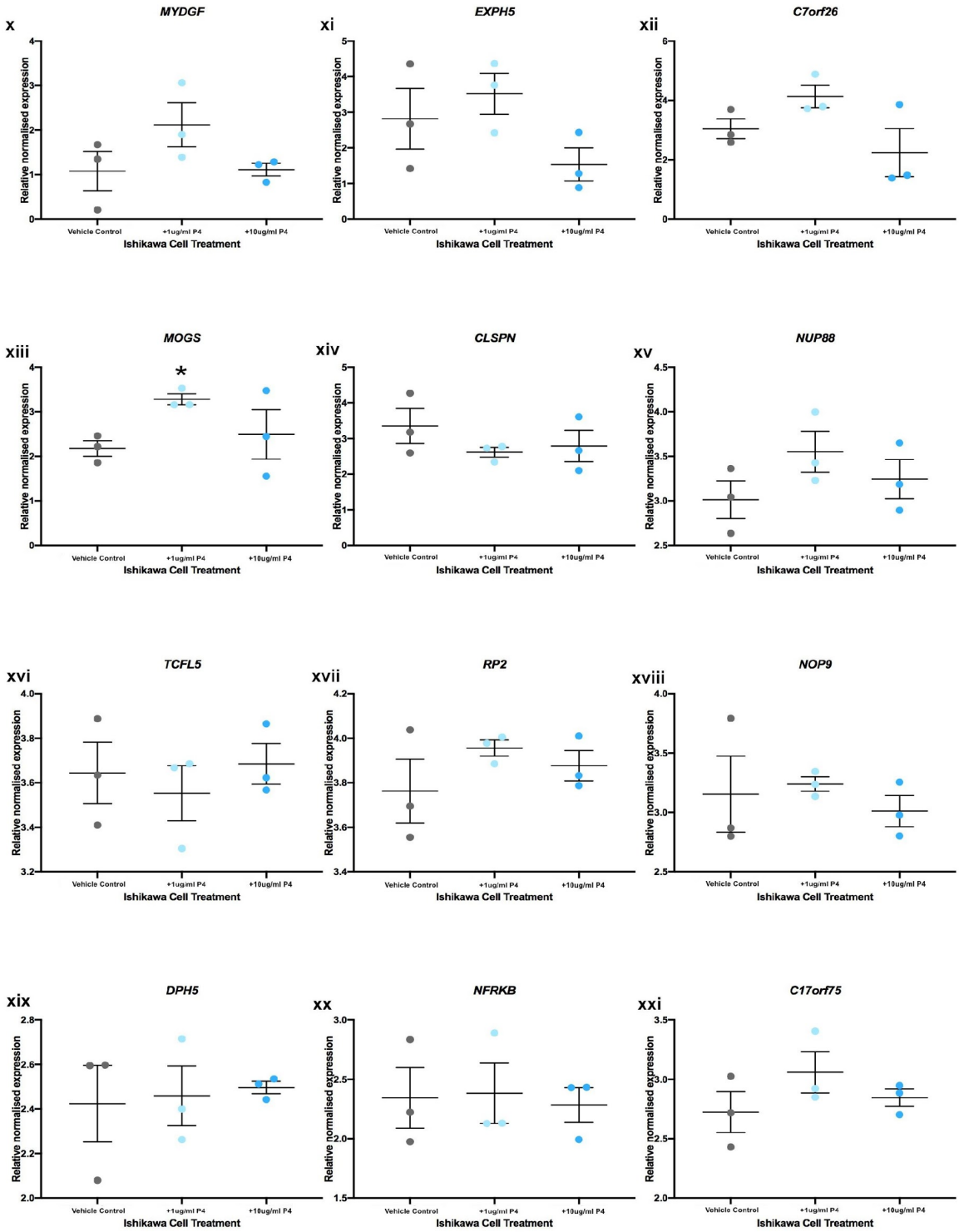


Figure 5.11 Expression of candidate positively selected genes in human Ishikawa immortalised endometrial epithelial cells treated with P4. *Changes in expression of conserved positively selected genes* ***(i)*** *N4BP3,* ***(ii)*** *HIRIP3,* ***(iii)*** *TERT,* ***(iv)*** *NSMCE1,* ***(v)*** *JCAD,* ***(vi)*** *MBTPS2,* ***(vii)*** *C5orf51,* ***(viii)*** *MDM1,* ***(ix)*** *RABAC1,* ***(x)*** *MYDGF,* ***(xi)*** *EXPH5,* ***(xii)*** *C7orf26,* ***(xiii)*** *MOGS,* ***(xiv)*** *CLSPN,* ***(xv)*** *NUP88,* ***(xvi)*** *TCFL5,* ***(xvii)*** *RP2,* ***(xviii)*** *NOP9,* ***(xix)*** *DPH5,* ***(xx)*** *NFRKB and* ***(xxi)*** *C17orf75 in human Ishikawa immortalised endometrial epithelial cells treated with* ***a)*** *vehicle control,* ***b)*** *1ng/µl P4, and* ***c)*** *10ng/µl P4. Significant differences in expression values determined when p≤0.05 are depicted by an asterisk (*).*

5.5 Discussion

In summary, of the 36 candidate human and bovine genes evaluated, two genes, *MOGS* and *NSMCE1* were found to be biologically regulated by progesterone in human immortalised endometrial epithelial cells and bovine primary endometrial stromal cells, respectively.

Previous work (Hume *et al.*, unpublished) determined 3 stem lineage miRNAs; miR-340-5p, miR-542-3p and miR-671-5p; are significantly downregulated in response to treatment of human Ishikawa immortalised endometrial epithelial cells with 10ng/µl P4. When investigating the response of the predicted positively selected targets of these miRNAs, one gene, *MOGS*, was found to be significantly downregulated in response to 1ng/µl P4, returning to levels without significant different from control samples when treated with 10ng/µl P4. miR-671-5p is predicted to have 7mer-m8 binding to the *MOGS* seed sequence from a TargetScan analysis (Agarwal *et al.*, 2015). Taking this predicted regulatory interaction into account, these results indicate a possible feedback loop between miR-671-5p and *MOGS*, that is mediated by P4 in human Ishikawa immortalised endometrial epithelial cells. Under conditions where miR-671-5p is normally expressed (with no significant difference in expression in vehicle control vs 1ng/µl P4), *MOGs* is significantly downregulated. However, under conditions where miR-671-5p is downregulated, *MOGS* is not significantly downregulated (where expression of MOGS in response to 10ng/µl P4 is compared to control, vehicle control and 1ng/µl P4).

At the early stages of pregnancy, P4 is secreted by the corpus luteum and increases until ~12 weeks of gestation, when progesterone production is performed by the placenta. During the follicular phase, serum P4 concentration is 1-2nmol/L (0.3-0.6ng/µl), rising to 35-50 nmol/L (11-15ng/µl) during the mid-luteal phase, 57nmol/L (18ng/µl) at 5 weeks gestation, reaching 80nmol/L (25ng/µl) at 12-13 weeks gestation, when the placenta takes over progesterone production (Kumar and Magon, 2012; Ku *et al.*, 2018). Applying these states to the experimental concentrations of P4, it is possible that *MOGS* is downregulated by miR-671-5p during the follicular phase, when follicles mature to allow ovulation (Monis and Tetrokalashvili, 2019), returning to normal levels during the luteal phase, as the uterus becomes receptive for an implanting conceptus (Tavaniotou *et al.*, 2002). *MOGS*, also known as mannosyl-oligosaccharide

glucosidase, or glucosidase 1, is expressed in the endoplasmic reticulum and is involved in the initial processing of *N*-glycans after their attachment to a glycoprotein (Nairn and Moremen, 2014). A genetic condition, *MOGS-CDG*, a congenital disorder of glycosylation, is the result of a frameshift mutation of the *MOGS* gene, resulting in a truncated protein. This condition affects multiple systems across the body, due to glycosylation being conducted in many different cell types, and can result in widespread food intolerance, seizures, delayed motor skill development and an apparent level of resistance to viral infection (Li *et al.*, 2019). This apparent resistance to viral infection may be explained by the fact that many viral proteins require post-translational modification using glycans. When the *N*-glycosylation process is negatively affected, as in MOGS-CDG, the proteins cannot be adequately modified, affecting viral replication post-entry (Sadat *et al.*, 2014). The same study evaluating MOGS-CDG effects on viral infection also analysed its effects on patient immune systems, noting that MOGS-CDG patients were affected with hypogammaglobulinemia, or insufficient production of gamma globulins, negatively affecting the immune response.

The abnormal production of immune cells in congenital disorders of glycosylation may be explained by the importance of glycosylation in the immune system, with immunoglobulins, complement immune cells, toll-like receptors, host immune receptors and adhesion molecules displaying some form of glycosylation (Lyons, Milner and Rosenzweig, 2015). *N*-linked glycosylation is implicated in the synthesis and action of toll-like receptors 2,3 and 4; the expression and recognition potential of major histocompatibility complex proteins I and II; antibody-antigen recognition by immunoglobulins; leukocyte recruitment and selectin interactions; and the ability of sialic acid-binding immunoglobulin-type lectins (Siglecs), to differentiate between endogenous molecules and antigens (Pascoal *et al.*, 2020).

Considering a major side effect of *MOGS* downregulation or absence is altered immune system reactions and impaired cell glycosylation indicates that *MOGS* downregulation in response to 1ng/µl P4 in human Ishikawa immortalised endometrial epithelial cells may induce an altered immune response, or altered glycosylation affecting cell-cell communication or recognition. While the specific dynamics of immune system alteration during ovulation, implantation and

early pregnancy are not completely understood, significant alterations are required in order to facilitate the implantation of a fetus while avoiding recognition by the maternal immune system as a 'foreign body' (Mor and Cardenas, 2010); modulate the remodelling of the endometrium (Lee, Lee and Lee, 2011); and angiogenesis and trophoblast invasion in the uterus (Faas and De Vos, 2018).

However, in order to fully investigate the role of miR-671-5p:*MOGS* interactions in the post-ovulation uterine landscape, further analyses are essential. Primarily, confirming the interaction using a luciferase or GFP assay between miR-671-5p and *MOGS* by inserting a reporter gene alongside the 3'UTR of *MOGS* and transfecting the recombinant *MOGS* plasmid, along with miR-671-5p into human Ishikawa immortalised endometrial epithelial cells and monitoring fluorescence. The miR-671-5p regulation of *MOGS* is given experimental support if the reporter is suppressed, indicating downregulation of *MOGS* (Kuhn *et al.*, 2008). Once the interaction is confirmed, the regulatory dynamics and effects may in investigated by overexpression of miR-671-5p, which may be conducted alongside knockout/knockdown experiments of miR-671-5p, in order to deduce the specific role of miR-671-5p-mediated downregulation of *MOGS* in response to P4.

In Chapter 3 (Sections 4.4.5 and 4.4.9) one stem lineage miRNA, miR-505-5p, was determined to be significantly upregulated in bovine endometrial epithelial and stromal cells (observed at 0.1µg/mL vs 10µg/mL P4 in epithelial cells and in vehicle control vs 10µg/mL P4 in stromal cells). When investigating the response of the predicted positively selected targets of this miRNA, one gene, *NSMCE1*, was found to be significantly downregulated in bovine endometrial epithelial cells treated with 1µg/mL P4 when compared to vehicle control and 0.1µg/mL P4. miR-505-5p is predicted to have 7-mer-1a binding to the *NSMCE1* 3'UTR seed sequence, with predictions taken from a TargetScan analysis (Agarwal *et al.*, 2015). Considering this predicted regulatory interaction, along with the altered expression patterns of miR-505-5p and its target gene, *NSMCE1*, in bovine endometrial stromal cells subject to dose dependent treatment with progesterone, indicates that miR-505-5p may upregulate *NSMCE1* expression. This pattern of upregulation is suggested by the fact that miR-505-5p is upregulated in bovine endometrial

stromal cells treated with 10µg/mL of progesterone. *NSMCE1* expression in the bovine endometrial stromal cells is decreased at lower concentrations of progesterone (0.1µg/mL vs 1.0µg/mL and vehicle control vs 1.0µg/mL), but shows no significant downregulation at the highest concentration of progesterone (10µg/mL), with expression levels positively correlating with miR-505-5p expression. This positive correlation indicates that miR-505-5p may directly induce *NSMCE1* via miRNA-induced RNA activation (miRNAa), similarly to that observed in the interaction of miR-4281 and FOXP3 during regulatory T cell differentiation (Zhang *et al.*, 2018). Alternatively, miR-505-5p may positively regulate *NSMCE1* expression via another pathway. While miRNAa is a documented method of miRNA upregulation, it has been theorised that miRNAs may induce upregulation of a target gene by engaging in competitive binding with another miRNA that is acting to repress expression of the target gene (Vasudevan, 2012). The observed differential expression of miR-505-5p and *NSMCE1* may support this hypothesis, in that *NSMCE1* does not experience corresponding upregulation at progesterone concentrations that induce miR-505-5p upregulation. Rather, *NSMCE1* expression is no longer significantly downregulated, indicating that increased miR-505-5p expression may antagonise other factors repressing *NSMCE1* expression, thereby removing the effect of these factors.

In terms of function, *NSMCE1* is a homologue of non-structural maintenance of chromosomes element 1 (*NSE1*), with annotated functions in the prevention and repair of DNA damage (Gong *et al.*, 2019). *NSMCE1,* along with *NSMCE2, NSMCE3* (also known as *NDNL2*) and *NSMCE4*, are mammalian-specific members of the SMC5/6 complex, a group of proteins found to be essential in chromosome segregation and replication during the cell cycle (Fernandez-Capetillo, 2016). Members of this complex (*NSMCE1, NSMCE2, NSMCE3 and SMC6*) are among the conserved gene families found to have undergone positive selection on the stem Eutherian lineage (Chapter 2), with this interaction annotated with high confidence in the STRINGv.11 (Szklarczyk *et al.*, 2019) analysis of these conserved gene families (Figure 2.6).

That *NSMCE1* and by extension the SMC5/6 complex is implicated in bovine pregnancy and may be targeted by a progesterone-regulated miRNA, with no corresponding differential expression in a human *in vitro* model of progesterone secretion, may be due to species-specific gene

regulatory networks. The role of progesterone in bovine conceptus development have been previously studied, in order to investigate the effect that increased progesterone concentration may have on the developing bovine conceptus (Forde *et al.*, 2009). This study identified differential expression of genes with functions related to glucose transport and energy synthesis induced by increased progesterone concentration during early pregnancy. However, the data gathered in this study also provided valuable information on the changes induced by endogenous progesterone concentrations on the bovine endometrium prior to maternal recognition of pregnancy and implantation. Among genes found to be differentially expressed in heifers exposed to normal and high concentrations of progesterone on Day 5 of pregnancy, a number of structural maintenance of chromosome (*SMC*) genes (*NSMCE2*, *SMC1*-like protein, *SMC4*-like protein and unidentified genes with homology to *SMC2*) were found to be significantly upregulated when compared to Day 7 of pregnancy, with increased fold change in the group exposed to higher progesterone concentrations. *SMC1*-like protein and *NSMCE2* were also significantly upregulated at Day 7 of pregnancy in heifers exposed to normal progesterone concentrations when compared to Day 13 of pregnancy. However, *SMC1*-like protein was found to be significantly downregulated at Day 7 of pregnancy in heifers exposed to high progesterone concentrations when compared to Day 13 of pregnancy, with no differential expression of *NSMCE2* noted. After Day 13 of pregnancy, no differential expression of *SMC* genes was observed in heifers exposed to normal or high progesterone concentrations.

The role of the SMC5/6 complex, and by extension *NSMCE1*, in early bovine pregnancy has yet to be elucidated. In the above study, *NSMCE1* expression was seen to decrease with increasing concentrations of progesterone and as gestation continued, until progesterone concentrations of 10µg/mL were reached. Circulating concentrations of progesterone in pregnant cattle at Day 24 of pregnancy have been previously measured as >0.001µg/mL, reaching 0.008µg/mL on Day 31 of pregnancy. The progesterone concentrations at which *NSMCE1* expression was downregulated (0.1µg/mL and 1.0µg/mL) would suggest that *NSMCE1* downregulation is induced at or around Day 7 of bovine pregnancy and is continually downregulated, with repression only ceased at concentrations of 10µg/mL. During early pregnancy, increasing progesterone concentrations repress estradiol and estradiol-mediated proliferation in the endometrial cell

population and induces differentiation of epithelial cells to prepare for implantation (Radi, Marusak and Morris, 2009). This purposeful prevention of cell proliferation may provide an explanation for the downregulation of *NSMCE1*. In the absence of normal proliferative activity, the need for chromosomal maintenance by SMC5/6 complex during the cell cycle may be lessened.

The data observed here, while providing evidence for miR-505-5p mediated upregulation of *NSMCE1*, along with possible downregulation of proteins required for normal cell cycle maintenance during early bovine pregnancy, does not provide sufficient information to understand the mechanism by which this is carried out. While confirming the interaction of miR-505-5p and *NSMCE1* is the primary recommendation for future research, using similar methods to those recommended in Section 5.5.2 (luciferase or GFP reporter assays to confirm the interaction, followed by overexpression and knockdown of miR-505-5p to gather further information on the regulatory dynamics of this interaction), additional analyses to investigate the location and method of miR-505-5p binding to *NSMCE1* may provide valuable information on miRNA-mediated upregulation of gene expression. Previous studies into miRNAa have utilised immunoprecipitation assays or luciferase reporter assays, combined with information from target prediction software (Lin *et al.*, 2011; Matsui *et al.*, 2013) to identify the precise location of miRNA binding. Multiple sequence alignments (MSAs) of the full *NSMCE1* sequence may serve to identify promoter elements or RNA motifs in the UTRs of *NSMCE1*, which could provide evidence for an miRNAa-like mechanism of gene expression upregulation, as utilised in the identification of *lin-4* activation (Turner, Jiao and Slack, 2014), which discovered an autoregulatory mechanism in the *lin-4* promotor region, where the mature *lin-4* miRNA binds its own promotor region to promote its own transcription. If, however, no evidence is found for miRNAa in the regulatory interaction of miR-505-5p and *NSMCE1*, a comparison of *NSMCE1* binding sites may be undertaken, to identify binding sites that overlap those for miR-505-5p, indicating possible competitive binding (Calloni and Bonatto, 2019). These results may be used alongside immunoprecipitation assays or reporter assays in bovine endometrial stromal cells treated with media control, vehicle control, 0.1µg/mL progesterone, 1.0µg/mL progesterone and 10µg/mL progesterone, to investigate alterations in miRNA populations near miR-505-5p binding sites on

NSMCE1. By examining which miRNAs bind to regions of potential competitive interactions with miR-505-5p as progesterone concentration increases, it may be possible to observe the displacement of a repressor miRNA with miR-505-5p.

Of the positively selected genes targeted by biologically regulated stem lineage miRNAs, only two genes, *MOGS* and *NSMCE1*, showed significant differential expression in *in vitro* models of early pregnancy. This contradicts the hypothesis that underlay this series of experiments, that altered expression of biologically regulated stem lineage miRNAs would induce corresponding expression changes in their positively selected targets, taking into account the increasing level of binding interactions shown between these miRNAs and genes shown to be positively selected on the stem Eutherian lineage. However, the lack of multiple observations of alterations in gene expression is only a negative result in the specific scenarios tested (*in vitro* models of early pregnancy, treated with molecules implemented in endometrial remodelling, recognition of pregnancy and implantation for 24 hours in static culture).

The successful establishment of pregnancy requires multiple factors, some of which are time- or dose-dependent. The steady increase of progesterone concentration from follicular phase to term pregnancy has been discussed (Kumar and Magon, 2012; Ku *et al.*, 2018). CAPG has also been shown to have altered effects dependant on concentration, with human Ishikawa immortalised endometrial epithelial cells showing significant upregulation of 5 genes in response to 10ng/µl and 100ng/µl of bCAPG, with no significant change in gene expression observed in human Ishikawa cells treated with 1000ng/µl bCAPG, indicating a dose-dependent response (Tinning *et al.*, 2020). IFNT, the pregnancy recognition signal in ruminants, has been shown to have time-dependent effects, inducing different gene expression profiles in bovine endometrial epithelial and stromal cells treated for 2 hours vs 24 hour treatment (Mansouri-Attia *et al.*, 2009; Forde *et al.*, 2011). Considering these observations, the *in vitro* models used in this analysis may not capture the full picture of transcriptional alterations during early pregnancy. While all but one of these genes did not show significant alterations in expression at the concentrations and times analysed, it is possible that they may be altered at lower or higher concentrations, or within shorter or longer windows of treatment. In addition, as discussed in Chapter 4, sample-sample

variation in miRNA expression was observed here. This variation may be obscuring smaller expression differences. Therefore if these experiments were to be repeated or continued, a larger treatment group is recommended.

Another factor that should be taken into account when considering the lack of significant differential expression in *in vitro* models of bovine early pregnancy is genome quality. Of the 37 candidate bovine genes for primer design, 13 had no RefSeq ID. Of the remaining 24, 9 had a baseMean ≤30. This resulted in only 22 candidate genes with possible primers, 10 of which were validated. These issues with sequence information, transcript abundance and primer validation resulted in a small pool of available genes to analyse. Using published RNA-sequencing data analysing the effects of 1000ng/µl bCAPG on human and bovine endometrial cells (Tinning *et al.*, 2020), one gene, *TCFL5* was found to be significantly downregulated in bovine endometrial epithelial cells treated with 1000ng/µl bCAPG. This gene could not be analysed by RT-qPCR, as there was no available RefSeq ID (Table 5.4). No candidate positively selected genes were found to be differentially expressed in bovine endometrial stromal cells, or in human Ishikawa immortalised endometrial epithelial cells. However, these results may be due to the time- and concentration-dependant issues discussed above.

The final factor that may be affecting the results of this analysis may be the method used to quantify the differential expression. RT-qPCR was selected for this analysis due to its flexibility and use in detecting alterations in expression of known genes. However, studies comparing RT-qPCR to RNA-sequencing have highlighted the differences in genes determined to be significantly differentially expressed between the two methods. Many differences in results have been ascribed to low log fold change in gene expression (where log fold change <2) (Everaert *et al.*, 2017). These issues relating to low log fold change in gene expression have been hypothesized to be due to high background noise in RT-qPCR methods of normalisation, in comparison to the low background noise of RNA-sequencing data (Costa *et al.*, 2013). These issues in background noise may be directly related to methods of normalisation, with RNA-sequencing using absolute quantification of gene expression, using the whole genome. Conversely, RT-qPCR methods depend on quantification of gene expression using either normaliser genes, or absolute

quantification based on the genes analysed in the experiment, which may obscure small differences in gene expression. In order to investigate these effects, published RNA-sequencing data of the effects of 10ng/µl progesterone treatment on human Ishikawa immortalised endometrial epithelial cells was used to examine alterations of positively selected genes regulated by miRNAs shown to be differentially expressed in response to progesterone. 6 of these genes (*MYDGF, EXPH5, C7orf26, MOGS, CLSPN* and *NUP88*) were found to be significantly differentially expressed, with none of these alterations in expression detected when the experiment was performed using RT-qPCR. All 6 differentially expressed genes had a log fold change <2, with 5 genes having a log fold change <1. These differences in detection highlight the effect that background noise in RT-qPCR may have on observed results, indicating that the alterations of these genes may occur, but in small amounts. When comparing RT-qPCR results to RNA-sequencing results, observations of altered *MOGS* expression are also altered. When analysed using RT-qPCR, *MOGS* was significantly downregulated in response to 1ng/µl P4, with no significant downregulation observed in response to 10ng/µl P4. RNA-sequencing data shows that *MOGS* is also significantly downregulated in response to 10ng/µl P4, however, the log fold change is low (-0.52). Considering the factors discussed above, if the expression of these positively selected genes were to be analysed in early pregnancy in a subsequent experiment, the use of multiple time points and concentrations, along with RNA-sequencing, may show further alterations in the expression of positively selected genes targeted by biologically regulated miRNAs.

5.6 Conclusion

Using miRNA target predictions, a list of candidate genes that i) experienced selective pressure on the stem Eutherian lineage and were never subsequently lost; and ii) are predicted to be targeted by stem lineage miRNAs that are regulated by molecules implicated in early pregnancy; were analysed in *in vitro* models of early pregnancy. One gene, *MOGS*, was found to be significantly downregulated in response to 1ng/µl P4 in human Ishikawa immortalised endometrial epithelial cells. *MOGS* is a predicted target of miR-671-5p, a stem lineage miRNA which is downregulated in response to 10ng/µl P4 in human Ishikawa immortalised endometrial epithelial cells. Using RNA-sequencing data, *MOGS* was found to be significantly downregulated to a lesser degree in response to 10ng/µl in human Ishikawa immortalised endometrial epithelial cells, indicating a possible P4-mediated regulatory interaction between miR-671-5p and *MOGS*. While the function of this interaction is as yet unknown, the altered immune cell profile of patients affected by *MOGS* deficiency may link this regulation to changes in the immune system during early pregnancy. Another gene, *NSMCE1*, was found to be downregulated at 0.1µg/mL and 1.0µg/mL P4 in bovine endometrial stromal cells, with no significant downregulation of *NSMCE1* at 10µg/mL P4, the concentration of P4 at which miR-505-5p, the stem lineage miRNA regulating *NSMCE1* is upregulated. This alteration in expression may indicate a miRNA-induced RNA activation response, or a competitive interaction by which miR-505-5p removes a repressor miRNA to allow *NSMCE1* expression at increased progesterone concentration. *NSMCE1* is a documented member of the SMC5/6 complex involved in chromosome maintenance and DNA damage repair. Members of the SMC5/6 complex also have previously been found to be upregulated in the endometrium of heifers at Day 5 of pregnancy, with subsequent downregulation as progesterone concentration increases. This differential expression of proteins involved in chromosome maintenance and repair during the cell cycle may be related to progesterone-mediated downregulation of estradiol, repression of cell proliferation and induction of epithelial cell differentiation to facilitate implantation.

No other candidate genes were found to be significantly differentially expressed in *in vitro* models. The lack of significant alterations in gene expression in other candidate genes may

be due to issues with detection of low fold change differences using RT-qPCR; or the *in vitro* models used being unable to accurately depict the dynamic changes experienced during early pregnancy, with more experimental parameters a recommendation if these genes are to be analysed in the context of early pregnancy in the future.

Chapter 6: Discussion

6.1 Discussion

The eutherian mammal placenta is an individualised tissue with a specialised function. It evolved once in the eutherian mammal clade and was never subsequently lost (Roberts, Green and Schulz, 2016). While placental morphology varies in eutherian mammal clades the purpose of the eutherian placenta remains the same: to facilitate physiological interchange between the mother and fetus during gestation (Garratt *et al.*, 2013). Studies of the evolution of novel tissues and traits have proposed that innovation in both protein coding and regulatory elements drive the development of novel traits (Wagner, 2015). Using the most current mammalian phylogeny, along with multiple high-quality vertebrate genomes, we sought to investigate the role of selective pressure and novel miRNA emergence in the evolution of the eutherian placenta.

Previous assessments of the power of a selective pressure analysis have emphasized the importance of an accurate phylogeny and multiple sequence alignment, along with downstream statistical testing (Hughes and Friedman, 2008; Zhai *et al.*, 2012). In Chapter 2, most up-to-date mammalian phylogeny (Morgan *et al.*, 2013; Tarver *et al.*, 2016) was used, along with gene families annotated by Ensembl Compara (Yates *et al.*, 2016) to combat possible false positives caused by inaccurate phylogeny or incorrect gene family annotations. In addition, studies have recommended the use of multiple alignment methods to find the best fitting alignment method and reduce false positives related to inaccurate multiple sequence alignments (Anisimova, M., Cannarozzi, G., & Liberles, 2010; Markova-Raina and Petrov, 2011). The alignment comparison method in the VESPA pipeline (Webb, Walsh and O'Connell, 2016b) was employed to identify the best alignment for each gene family. Using these methods, along with downstream likelihood-ratio tests (Yang, 2007) and filtering for gene families that had undergone a single, fixed substitution on the stem Eutherian lineage (Orr, 2018), 237 gene families were identified that underwent positive selection on the stem Eutherian lineage. Of these, 115 gene families (PSStemFixed) were shared between all sampled Eutherian species. While none of these 115 gene families were explicitly associated with placentation or early pregnancy events, STRINGv.11

(Szklarczyk *et al.*, 2019) and PANTHERv.14 (Muruganujan *et al.*, 2018) were used to investigate predicted functional annotations and interactions. Using these methods, the 115 SGOs that underwent positive selection on the stem eutherian lineage were found to be functionally enriched for telomere activity and were implicated in pathways related to chromosomal maintenance, telomere activity, p53 signalling, cell cycle and the inflammatory immune response, activities which have been associated with the formation of the placenta (Hauguel-de Mouzon and Guerre-Millo, 2006; Woods, Perez-Garcia and Hemberger, 2018; Gal *et al.*, 2019).

In Chapter 3, the most up-to-date eutherian phylogeny (Morgan *et al.*, 2013; Tarver *et al.*, 2016) was used in tandem with a curated database of miRNA annotations (Fromm *et al.*, 2018) to identify the phylogenetic placement of miRNAs that emerged in therian mammals. The choice was made to include miRNAs that emerged in *Theria* (which are live-bearing mammals with a short-lived yolk-sac placenta), along with those that emerged in *Eutheria*, in order to investigate possible co-option of mammalian miRNAs during the evolution of the eutherian placenta. Through the use of homology searches of miRNA stem loop sequences against multiple vertebrate genomic regions (3'UTR, 5'UTR, exons and introns), 112 miRNAs were placed on the therian phylogeny. Of these, 2 miRNAs emerged on the therian stem lineage and 12 emerged on the eutherian stem lineage. Each of these miRNAs were associated with the placenta through expression data and literature sources. In order to investigate possible functions of these 14 miRNAs, TargetScan (Agarwal *et al.*, 2015) was used to predict putative target genes, which were then filtered by the predicted complementarity of the seed sequence. The predicted targets of these miRNAs included 130 gene families that emerged on the eutherian stem lineage (Dunwell, Paps and Holland, 2017). Also among the predicted targets were 15379 genes that evolved endometrial expression on the stem eutherian lineage that are hypothesized to have assisted in the remodelling of the uterine landscape during the evolution of the placenta (Lynch *et al.*, 2015). The Reactome database was used to identify predicted pathways for the targets of the 14 miRNAs that emerged on the therian and eutherian stem lineages (Jassal *et al.*, 2019). The targets of these miRNAs were implicated in pathways associated with INPP5E regulation, Neurophilin interactions and VEGF interactions with its receptor (VEGR), each implicated in angiogenesis. TGF-β signalling and p53 regulation were also among the associated pathways, which are each implicated in cell

proliferation. Furthermore, 84 of the 115 PSStemFixed families are predicted targets of the 14 stem lineage miRNAs. The evolution of novel miRNAs and their binding sites on genes undergoing selective pressure during the evolution of the placenta may indicate the formation of novel regulatory networks. Considering that novel regulatory networks have been theorised to support phenotypic variation (Wagner, 2015; Kin *et al.*, 2016), the interactions between these 14 miRNAs and the 84 predicted targets that underwent positive selection on the stem eutherian lineage, subject to *in vitro* validation, may be targets of future work investigating the impact of these interactions on placental function.

In order to assess the role of the 14 miRNAs that emerged on the stem therian and eutherian lineages in the formation of the placenta, RT-qPCR was used in Chapter 4 to analyse their expression in mammalian reproductive tissues and models of early pregnancy. Expression of the 14 stem lineage miRNAs was confirmed in reproductive tissue from human, bovine, porcine and murine sources. However, expression could not be fully assessed in opossum uterine or mammary gland samples. This could be due to the changes in seed sequence that occurred in the therian miRNAs on the eutherian stem lineage. Alternatively, the 2 miRNAs that emerged on the stem therian lineage may be expressed in marsupial mammary or reproductive tissues, evolving expression in reproductive tissues on the eutherian stem lineage. Using human and bovine *in vitro* models of early pregnancy events, the expression of 12 miRNAs were found to be significantly altered in response to bovine maternal recognition of pregnancy, conceptus-secreted implantation-related proteins and progesterone action. Little overlap was observed in differentially expressed miRNAs between bovine and human *in vitro* models, which may indicate that these stem lineage miRNAs were adapted in a species-specific manner as different early pregnancy mechanisms evolved. However, one miRNA, miR-185-5p, was found to be significantly upregulated in both human and bovine *in vitro* models of a conserved conceptus-secreted protein, P4HB. This similar response in species with markedly different implantation strategies suggests that some part of the implantation-related transcriptional response is regulated in the same manner across different implantation morphologies. With differential expression of 12 stem lineage miRNAs established in bovine and human *in vitro* models of early pregnancy they may be future candidates for investigation into regulatory changes during early pregnancy.

In Chapter 5, PSStemFixed families targeted by miRNAs found to be differentially expressed in bovine and human *in vitro* models of early pregnancy, were assessed to determine if they underwent corresponding differential expression in the same in vitro models. The aim of this analysis was both to investigate functional roles of the positively selected genes and determine the role of predicted functional interactions between stem lineage miRNAs and genes that underwent positive selection on the stem eutherian lineage. Using RT-qPCR, expression of 40 positively selected genes with ≥5 predicted binding sites for stem lineage miRNAs was investigated in human and bovine models of early pregnancy used in Chapter 4. A possible regulatory interaction was observed between *MOGS* and miR-671-5p in human Ishikawa immortalised endometrial epithelial cells treated with progesterone, with *MOGS* found to be significantly downregulated. Downregulation of *MOGS*, a glucosidase involved in the processing of *N*-glycans, is associated with an altered immune response and alterations in cell-cell communication (Li *et al.*, 2019; Pascoal *et al.*, 2020). Significant alterations in both of these pathways are required for successful establishment of pregnancy, which may explain the response of *MOGS* to progesterone treatment. Another possible positive regulatory interaction was observed between *NSMCE1* and miR-505-5p in primary bovine endometrial epithelial cells treated with progesterone, with *NSMCE1* found to be significantly downregulated. *NSMCE1* is a member of the SMC5/6 DNA damage repair complex and is associated with the normal cell cycle (Fernandez-Capetillo, 2016). This complex was found to be upregulated in response to progesterone secretion at Day 5 of bovine pregnancy, with expression decreasing as progesterone concentration increased (Forde *et al.*, 2009). No precise function of *NSMCE1* and the SMC5/6 complex in bovine pregnancy has been established. However, the downregulation of *NSMCE1* in response to increasing concentrations of progesterone may be associated with the anti-proliferative action of increasing progesterone concentration on the endometrial cell population (Radi, Marusak and Morris, 2009). The lack of differential expression observed in other candidate positively selected genes in human and bovine *in vitro* models of early pregnancy may be due to a number of factors. The *in vitro* models of pregnancy used in this analysis are of early pregnancy. The remaining 38 genes may not be involved in early pregnancy events. It is possible that these genes are implicated in later stages of pregnancy, such as placental

maintenance. In addition, the *in vitro* models used were subject to 24 hour treatment with the molecule of interest. As established, miRNA regulation and early pregnancy are dynamic and the *in vitro* models used cannot capture the full transcriptomic landscape of early pregnancy. Finally, studies evaluating the use of RT-qPCR in detecting differential gene expression have suggested that high background noise caused by normaliser genes in RT-qPCR may obscure differential gene expression with a low log-fold change (Costa *et al.*, 2013). This hypothesis may explain some results of this analysis, as published RNA-sequencing data of progesterone treatment on human Ishikawa immortalised endometrial epithelial cells found significant differential expression of 6 of the candidate positively selected genes, each with a log-fold change of <2. Therefore, future analyses of these candidate genes and their associated miRNAs may reveal functions in early pregnancy and placentation not observed here.

This thesis sought to combine computational identification and expression analyses of innovation in protein coding and regulatory elements that may have facilitated the emergence of placentation in eutherian mammals. A conserved set of 115 gene families were found to have undergone positive selection on the stem eutherian lineage. This result provides evidence of adaptations on the protein coding level which may have helped to drive the phenotypic alterations required for the evolution of the placenta. 112 miRNAs were found to have emerged in *Theria* and *Eutheria*. Two of these miRNAs emerged on the therian stem lineage, with a further 12 miRNAs found to have emerged on the eutherian stem lineage. With targets of these stem lineage miRNAs implicated in transcriptional remodelling of the endometrium, angiogenesis and cell proliferation, along with 84 gene families that underwent positive selection on the stem eutherian lineage they are valuable candidates for future analysis of regulatory alterations in placental evolution, early pregnancy, placentation and placental disorders. Using RT-qPCR, 12 stem lineage miRNAs were found to be differentially expressed in *in vitro* models of human and bovine early pregnancy, with a conserved miRNA response found in human and bovine epithelial cells treated with bP4HB, a conserved conceptus-secreted implantation-related protein. 2 of the positively selected genes targeted by stem lineage miRNAs were found to be differentially expressed in human and bovine models of progesterone treatment. These results may indicate a regulatory interaction between stem lineage miRNAs and genes that underwent positive

selection on the stem eutherian lineage during progesterone-mediated remodelling of the endometrium in early pregnancy. From the work presented here, it is clear that innovations in protein coding and regulatory elements occurred during the evolution of eutherian placenta. Interactions between miRNAs that emerged on the therian and eutherian stem lineage and gene families that underwent positive selection on the eutherian stem lineage may indicate a novel regulatory network that emerged at the time of placental evolution. Expression analysis of these elements suggest an involvement in early pregnancy events that establish pregnancy and initiate placentation and provide candidates for future functional work and assessment of these elements in both healthy and disordered pregnancy.

This thesis has provided evidence of a conserved regulatory network that emerged on the root eutherian lineage, members of which have been found in this work to respond to molecules that are necessary for early pregnancy and placental establishment. This work provides further weight to the theory of organ identity networks (Wagner, 2015) and highlights the importance of miRNAs during evolution and development. Furthermore, this thesis has provided evidence of concurrent positive selection, miRNA emergence and regulatory re-wiring at a crucial time in mammalian evolution. That members of this network are involved in early pregnancy and placentation has been shown, both in this work and in literature. However, the elements in this network have been implicated in multiple functions and are not specific to reproduction. Therefore, these molecules may help the understanding of how morphological innovation occurs and how novel organs and systems emerge.

While the functions of members of this conserved eutherian regulatory network are not specific to reproduction, this thesis has provided evidence of their response to early pregnancy signals associated with pregnancy recognition, implantation and endometrial remodelling. This evidence, alongside conserved reactions in human and bovine models indicate a role in establishing pregnancy and the placenta itself. Using this information, members of this network (primarily the miRNAs) may be used to understand altered gene expression during preeclampsia, implantation failure and recurrent miscarriage. This understanding may be exploited in order to prevent or treat these issues, by correcting aberrant gene expression. Presenting a pathway by

which these issues could be treated has consequences clinically and agriculturally, enabling healthy pregnancies, combating causes of infertility and preventing gestational complications.

Chapter 7: Bibliography

7.1 Bibliography

Abdel-Hafiz, H. A. and Horwitz, K. B. (2012) 'Control of progesterone receptor transcriptional synergy by SUMOylation and deSUMOylation', *BMC molecular biology*. BioMed Central, 13, p. 10. doi: 10.1186/1471-2199-13-10.

Agarwal, V. *et al.* (2015) 'Predicting effective microRNA target sites in mammalian mRNAs', *eLife*, 4(AUGUST2015), pp. 1–38. doi: 10.7554/eLife.05005.

Aguilera, F., McDougall, C. and Degnan, B. M. (2017) 'Co-Option and De Novo Gene Evolution Underlie Molluscan Shell Diversity', *Molecular Biology and Evolution*, 34(4), pp. 779–792. doi: 10.1093/molbev/msw294.

Ajayi, A. F. and Akhigbe, R. E. (2020) 'Staging of the estrous cycle and induction of estrus in experimental rodents: an update', *Fertility research and practice*. BioMed Central, 6, p. 5. doi: 10.1186/s40738-020-00074-3.

Almeida, F. C. *et al.* (2014) 'Family Size Evolution in Drosophila Chemosensory Gene Families: A Comparative Analysis with a Critical Appraisal of Methods', *Genome Biology and Evolution*, 6(7), pp. 1669–1682. doi: 10.1093/gbe/evu130.

Altmäe, S. *et al.* (2012) 'Research resource: interactome of human embryo implantation: identification of gene expression pathways, regulation, and integrated regulatory networks', *Molecular endocrinology (Baltimore, Md.)*. 2011/11/10. Endocrine Society, 26(1), pp. 203–217. doi: 10.1210/me.2011-1196.

Altschul, S. F. *et al.* (1990) 'Basic local alignment search tool.', *Journal of molecular biology*, 215(3), pp. 403–10. doi: 10.1016/S0022-2836(05)80360-2.

Alvergne, A. and Högqvist Tabor, V. (2018) 'Is Female Health Cyclical? Evolutionary Perspectives on Menstruation', *Trends in Ecology & Evolution*, 33(6), pp. 399–414. doi: https://doi.org/10.1016/j.tree.2018.03.006.

Anisimova, M., Cannarozzi, G., & Liberles, D. (2010) 'Finding the balance between the mathematical and biological optima in multiple sequence alignment', *Trends in Evolutionary Biology*, 2(1), p. e7. doi: https://doi.org/10.4081/eb.2010.e7.

Anisimova, M., Bielawski, J. P. and Yang, Z. (2001) 'Accuracy and Power of the Likelihood Ratio Test in Detecting Adaptive Molecular Evolution', *Molecular Biology and Evolution*, 18(8), pp. 1585–1592. doi: 10.1093/oxfordjournals.molbev.a003945.

Anisimova, M., Bielawski, J. P. and Yang, Z. (2002) 'Accuracy and Power of Bayes Prediction of Amino Acid Sites Under Positive Selection', *Molecular Biology and Evolution*, 19(6), pp. 950–958. doi: 10.1093/oxfordjournals.molbev.a004152.

Anisimova, M., Nielsen, R. and Yang, Z. (2003) 'Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites', *Genetics*, 164(3), pp. 1229–1236. Available at: https://www.ncbi.nlm.nih.gov/pubmed/12871927.

Anisimova, M. and Yang, Z. (2007) 'Multiple Hypothesis Testing to Detect Lineages under Positive Selection that Affects Only a Few Sites', *Molecular Biology and Evolution*, 24(5), pp. 1219–1228. doi: 10.1093/molbev/msm042.

Areal, H., Abrantes, J. and Esteves, P. J. (2011) 'Signatures of positive selection in Toll-like receptor (TLR) genes in mammals', *BMC Evolutionary Biology*, 11(1), p. 368. doi: 10.1186/1471-2148-11-368.

Arroyo, J. I. *et al.* (2012) 'INSL4 Pseudogenes Help Define the Relaxin Family Repertoire in the Common Ancestor of Placental Mammals', *Journal of Molecular Evolution*, 75(1), pp. 73–78. doi: 10.1007/s00239-012-9517-0.

Ayala, F. J., Powell, J. R. and Tracey, M. L. (1972) 'Enzyme variability in the Drosophila willistoni Group. V. Genic variation in natural populations of Drosophila equinoxialis', *Genetical Research*. 2009/04/14. Cambridge University Press, 20(1), pp. 19–42. doi: DOI: 10.1017/S0016672300013562.

Baerwald, A. R., Adams, G. P. and Pierson, R. A. (2005) 'Form and function of the corpus luteum during the human menstrual cycle', *Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 25(5), pp. 498–507. doi: 10.1002/uog.1891.

Bakker, J. and Baum, M. J. (2000) 'Neuroendocrine regulation of GnRH release in induced ovulators', *Frontiers in Neuroendocrinology*, 21(3), pp. 220–262. doi: 10.1006/frne.2000.0198.

Baldwin, M. W. *et al.* (2014) 'Evolution of sweet taste perception in hummingbirds by transformation of the ancestral umami receptor', *Science*, 345(6199), pp. 929 LP – 933. doi:

10.1126/science.1255097.

Ball, B. A. *et al.* (2019) 'Relationship between anti-Müllerian hormone and fertility in the mare.', *Theriogenology*. United States, 125, pp. 335–341. doi: 10.1016/j.theriogenology.2018.11.005.

Barlow, D. P. *et al.* (1991) 'The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus', *Nature*, 349(6304), pp. 84–87. doi: 10.1038/349084a0.

Bartel, D. P. (2004) 'MicroRNAs: Genomics, Biogenesis, Mechanism, and Function', *Cell*, 116(2), pp. 281–297. doi: https://doi.org/10.1016/S0092-8674(04)00045-5.

Bartel, D. P. (2009) 'MicroRNA Target Recognition and Regulatory Functions', *Cell*, 136(2), pp. 215–233. doi: 10.1016/j.cell.2009.01.002.

Bartolomei, M. S. and Ferguson-Smith, A. C. (2011) 'Mammalian genomic imprinting', *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press, 3(7), p. a002592. doi: 10.1101/cshperspect.a002592.

Bassett, A. R. *et al.* (2014) 'Understanding functional miRNA–target interactions in vivo by site-specific genome engineering', *Nature Communications*, 5(1), p. 4640. doi: 10.1038/ncomms5640.

Bauersachs, S. *et al.* (2006) 'Embryo-induced transcriptome changes in bovine endometrium reveal species-specific and common molecular markers of uterine receptivity', *Reproduction*. Bristol, UK: Society for Reproduction and Fertility, 132(2), pp. 319–331. doi: 10.1530/rep.1.00996.

Bauersachs, S. *et al.* (2012) 'Comparison of the Effects of Early Pregnancy with Human Interferon, Alpha 2 (IFNA2), on Gene Expression in Bovine Endometrium1', *Biology of Reproduction*, 86(2). doi: 10.1095/biolreprod.111.094771.

Bazer, F. W. *et al.* (2011) 'Uterine receptivity to implantation of blastocysts in mammals', *Frontiers in Bioscience*, 3, pp. 745–67. doi: 10.2741/s184.

Bazer, F. W., Song, G. and Thatcher, W. W. (2012) 'Roles of conceptus secretory proteins in establishment and maintenance of pregnancy in ruminants', *Asian-Australasian journal of animal sciences*. Asian-Australasian Association of Animal Production Societies (AAAP) and Korean Society of Animal Science and Technology (KSAST), 25(1), pp. 1–16. doi: 10.5713/ajas.2011.r.08.

Bazer, F. W., Spencer, T. E. and Johnson, G. A. (2009) 'Interferons and Uterine Receptivity',

Semin Reprod Med, 27(01), pp. 90–102. doi: 10.1055/s-0028-1108013.

Bedford, J. M. (2014) 'Singular features of fertilization and their impact on the male reproductive system in eutherian mammals', *REPRODUCTION*. Bristol, UK: Bioscientifica Ltd, 147(2), pp. R43–R52. doi: 10.1530/REP-13-0436.

Behringer, R. R., Eakin, G. S. and Renfree, M. B. (2005) 'Mammalian diversity: gametes, embryos and reproduction', *Reproduction, Fertility and Development*, 18(2), pp. 99–107. Available at: https://doi.org/10.1071/RD05137.

Bellato, M. *et al.* (2019) 'A Bioinformatics Approach to Explore MicroRNAs as Tools to Bridge Pathways Between Plants and Animals. Is DNA Damage Response (DDR) a Potential Target Process? ', *Frontiers in Plant Science* , p. 1535. Available at: https://www.frontiersin.org/article/10.3389/fpls.2019.01535.

Bellomo, D. *et al.* (1996) 'Cell proliferation in mammalian gastrulation: The ventral node and notochord are relatively quiescent', *Developmental Dynamics*. John Wiley & Sons, Ltd, 205(4), pp. 471–485. doi: 10.1002/(SICI)1097-0177(199604)205:4<471::AID-AJA10>3.0.CO;2-4.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing', *Journal of the Royal Statistical Society. Series B (Methodological)*. [Royal Statistical Society, Wiley], 57(1), pp. 289–300. Available at: http://www.jstor.org/stable/2346101.

Benton, R. (2015) 'Multigene Family Evolution: Perspectives from Insect Chemoreceptors', *Trends in Ecology & Evolution*, 30(10), pp. 590–600. doi: https://doi.org/10.1016/j.tree.2015.07.009.

Berezikov, E. (2011) 'Evolution of microRNA diversity and regulation in animals.', *Nature reviews. Genetics*. Nature Publishing Group, 12(12), pp. 846–60. doi: 10.1038/nrg3079.

Bidarimath, M. *et al.* (2017) 'Extracellular vesicle mediated intercellular communication at the porcine maternal-fetal interface: A new paradigm for conceptus-endometrial cross-talk', *Scientific reports*. Nature Publishing Group, 7, p. 40476. doi: 10.1038/srep40476.

Blackburne, B. P. and Whelan, S. (2011) 'Measuring the distance between multiple sequence alignments', *Bioinformatics*, 28(4), pp. 495–502. doi: 10.1093/bioinformatics/btr701.

Booker, T. R., Jackson, B. C. and Keightley, P. D. (2017) 'Detecting positive selection in the genome', *BMC Biology*, 15(1), p. 98. doi: 10.1186/s12915-017-0434-y.

Boos, A. *et al.* (2006) 'Pregnancy effects on distribution of progesterone receptors, oestrogen receptor α, glucocorticoid receptors, Ki-67 antigen and apoptosis in the bovine interplacentomal uterine wall and foetal membranes', *Animal Reproduction Science*, 91(1), pp. 55–76. doi: https://doi.org/10.1016/j.anireprosci.2005.03.012.

Bounds, K. R. *et al.* (2017) 'MicroRNAs: New Players in the Pathobiology of Preeclampsia', *Frontiers in cardiovascular medicine*. Frontiers Media S.A., 4, p. 60. doi: 10.3389/fcvm.2017.00060.

Brandt, J. *et al.* (2005) 'Transposable elements as a source of genetic innovation: expression and evolution of a family of retrotransposon-derived neogenes in mammals', *Gene*, 345(1), pp. 101–111. doi: https://doi.org/10.1016/j.gene.2004.11.022.

Brayman, M. J. *et al.* (2006) 'Progesterone Receptor Isoforms A and B Differentially Regulate MUC1 Expression in Uterine Epithelial Cells', *Molecular Endocrinology*, 20(10), pp. 2278–2291. doi: 10.1210/me.2005-0343.

Britten, R. J. (1986) 'Rates of DNA sequence evolution differ between taxonomic groups', *Science*, 231(4744), pp. 1393 LP – 1398. doi: 10.1126/science.3082006.

Bromham, L. (2011) 'The genome as a life-history character: why rate of molecular evolution varies between mammal species', *Philosophical Transactions of the Royal Society B: Biological Sciences*. Royal Society, 366(1577), pp. 2503–2513. doi: 10.1098/rstb.2011.0014.

Bronson, F. H. (1989) *Mammalian Reproductive Biology*. Chicago: The University of Chicago Press.

Brooks, K., Burns, G. and Spencer, T. E. (2014) 'Conceptus elongation in ruminants: roles of progesterone, prostaglandin, interferon tau and cortisol', *Journal of animal science and biotechnology*. BioMed Central, 5(1), p. 53. doi: 10.1186/2049-1891-5-53.

Calloni, R. and Bonatto, D. (2019) 'Characteristics of the competition among RNAs for the binding of shared miRNAs', *European Journal of Cell Biology*, 98(2), pp. 94–102. doi: https://doi.org/10.1016/j.ejcb.2019.04.001.

Campo-Paysaa, F. *et al.* (2011) 'microRNA complements in deuterostomes: Origin and evolution of microRNAs', *Evolution and Development*, 13(1), pp. 15–27. doi: 10.1111/j.1525-142X.2010.00452.x.

Cardoso-Moreira, M. *et al.* (2019) 'Gene expression across mammalian organ development',

Nature. 2019/06/26, 571(7766), pp. 505–509. doi: 10.1038/s41586-019-1338-5.

Carter, A. M. and Enders, A. C. (2004) 'Comparative aspects of trophoblast development and placentation', *Reproductive Biology and Endocrinology*, 2(1), p. 46. doi: 10.1186/1477-7827-2-46.

Carter, A. M. and Mess, A. M. (2014) 'Mammalian Placentation: Implications for Animal Models', *Pathobiology of Human Disease*, (November), pp. 2423–2442. doi: 10.1016/B978-0-12-386456-7.05013-9.

Chavatte-Palmer, P. and Tarrade, A. (2016) 'Placentation in different mammalian species', *Annales d'Endocrinologie*. Elsevier Masson SAS, 77(2), pp. 67–74. doi: 10.1016/j.ando.2016.04.006.

Chen, D.-B. and Zheng, J. (2014) 'Regulation of placental angiogenesis', *Microcirculation (New York, N.Y. : 1994)*, 21(1), pp. 15–25. doi: 10.1111/micc.12093.

Chen, M.-Y., Liang, D. and Zhang, P. (2017) 'Phylogenomic Resolution of the Phylogeny of Laurasiatherian Mammals: Exploring Phylogenetic Signals within Coding and Noncoding Sequences', *Genome Biology and Evolution*, 9(8), pp. 1998–2012. doi: 10.1093/gbe/evx147.

Chen, Z. *et al.* (2017) 'CRL4B(DCAF11) E3 ligase targets p21 for degradation to control cell cycle progression in human osteosarcoma cells', *Scientific reports*. Nature Publishing Group UK, 7(1), p. 1175. doi: 10.1038/s41598-017-01344-9.

Cheng, A. M. *et al.* (1998) 'Mammalian Grb2 Regulates Multiple Steps in Embryonic Development and Malignant Transformation', *Cell*. Elsevier, 95(6), pp. 793–803. doi: 10.1016/S0092-8674(00)81702-X.

Chi, R. A. *et al.* (2019) 'Human Endometrial Transcriptome and Progesterone Receptor Cistrome Reveal Important Pathways and Epithelial Regulators', *The Journal of Clinical Endocrinology & Metabolism*, 105(4). doi: 10.1210/clinem/dgz117.

Christenson, L. K. and Devoto, L. (2003) 'Cholesterol transport and steroidogenesis by the corpus luteum', *Reproductive biology and endocrinology : RB&E*. BioMed Central, 1, p. 90. doi: 10.1186/1477-7827-1-90.

Chuong, E. B. *et al.* (2013) 'Endogenous retroviruses function as species-specific enhancer elements in the placenta', *Nature Genetics*. Nature Publishing Group, 45(3), pp. 325–329. doi: 10.1038/ng.2553.

Concannon, P. W. (2011) ‘Reproductive cycles of the domestic bitch’, *Animal Reproduction Science*, 124(3), pp. 200–210. doi: https://doi.org/10.1016/j.anireprosci.2010.08.028.

Conn, P. M. and Crowley, W. F. (1991) ‘Gonadotropin-Releasing Hormone and Its Analogues’, *New England Journal of Medicine*. Massachusetts Medical Society, 324(2), pp. 93–103. doi: 10.1056/NEJM199101103240205.

Cordes, K. R. *et al.* (2009) ‘miR-145 and miR-143 regulate smooth muscle cell fate and plasticity’, *Nature*. 2009/07/05, 460(7256), pp. 705–710. doi: 10.1038/nature08195.

Correia de Sousa, M. *et al.* (2019) ‘Deciphering miRNAs’ Action through miRNA Editing’, *International Journal of Molecular Sciences* . doi: 10.3390/ijms20246249.

Costa, C. *et al.* (2013) ‘Comprehensive molecular screening: from the RT-PCR to the RNA-seq’, *Translational lung cancer research*. Pioneer Bioscience Publishing Company, 2(2), pp. 87–91. doi: 10.3978/j.issn.2218-6751.2013.02.05.

Costanzo, V. *et al.* (2018) ‘Exploring the links between cancer and placenta development’, *Open biology*. The Royal Society, 8(6), p. 180081. doi: 10.1098/rsob.180081.

Cretekos, C. J. *et al.* (2008) ‘Regulatory divergence modifies limb length between mammals’, *Genes & development*. Cold Spring Harbor Laboratory Press, 22(2), pp. 141–151. doi: 10.1101/gad.1620408.

Crowe, M. A. and Mullen, M. P. (2013) ‘Regulation and Function of Gonadotropins Throughout the Bovine Oestrous Cycle’, in Vizcarra, J. (ed.) *Gonadotropic*. IntechOpen. doi: DOI: 10.5772/53870.

Cvijović, I., Good, B. H. and Desai, M. M. (2018) ‘The Effect of Strong Purifying Selection on Genetic Diversity’, *Genetics*, 209(4), pp. 1235 LP – 1278. doi: 10.1534/genetics.118.301058.

Datto, M. B. *et al.* (1995) ‘Transforming growth factor beta induces the cyclin-dependent kinase inhibitor p21 through a p53-independent mechanism’, *Proceedings of the National Academy of Sciences of the United States of America*, 92(12), pp. 5545–5549. doi: 10.1073/pnas.92.12.5545.

Davidson, E. H. and Erwin, D. H. (2006) ‘Gene Regulatory Networks and the Evolution of Animal Body Plans’, *Science*, 311(5762), pp. 796 LP – 800. doi: 10.1126/science.1113832.

Degterev, A., Ofengeim, D. and Yuan, J. (2019) ‘Targeting RIPK1 for the treatment of human diseases’, *Proceedings of the National Academy of Sciences*, 116(20), pp. 9714 LP – 9722. doi: 10.1073/pnas.1901179116.

DeSouza, M. M. *et al.* (2000) 'MUC1/episialin: a critical barrier in the female reproductive tract', *Journal of Reproductive Immunology*, 45(2), pp. 127–158. doi: https://doi.org/10.1016/S0165-0378(99)00046-7.

Ding, J. *et al.* (2015) 'MiR-519d-3p Suppresses Invasion and Migration of Trophoblast Cells via Targeting MMP-2', *PLOS ONE*. Public Library of Science, 10(3), p. e0120321. Available at: https://doi.org/10.1371/journal.pone.0120321.

do Imperio, G. E. *et al.* (2018) 'Chorioamnionitis Induces a Specific Signature of Placental ABC Transporters Associated with an Increase of miR-331-5p in the Human Preterm Placenta', *Cellular Physiology and Biochemistry*, 45(2), pp. 591–604. doi: 10.1159/000487100.

Dominguez, F. *et al.* (2005) 'Embryonic implantation and leukocyte transendothelial migration: different processes with similar players?', *The FASEB Journal*. Federation of American Societies for Experimental Biology, 19(9), pp. 1056–1060. doi: 10.1096/fj.05-3781hyp.

Dong, J. *et al.* (2010) 'MicroRNA Networks in Mouse Lung Organogenesis', *PLoS ONE*, 5(5). doi: 10.1371/journal.pone.0010854.

Driggers, P. H. and Segars, J. H. (2002) 'Estrogen action and cytoplasmic signaling pathways. Part II: the role of growth factors and phosphorylation in estrogen signaling', *Trends in endocrinology and metabolism: TEM*, 13(10), pp. 422–427. doi: 10.1016/s1043-2760(02)00634-3.

Dunwell, T. L., Paps, J. and Holland, P. W. H. (2017) 'Novel and divergent genes in the evolution of placental mammals', *Proceedings. Biological sciences*. 2017/10/04. The Royal Society, 284(1864), p. 20171357. doi: 10.1098/rspb.2017.1357.

Duret, L. *et al.* (2006) 'The Xist RNA Gene Evolved in Eutherians by Pseudogenization of a Protein-Coding Gene', *Science*, 312(5780), pp. 1653 LP – 1655. doi: 10.1126/science.1126316.

Dwivedi, Y. *et al.* (2015) 'Chronic corticosterone-mediated dysregulation of microRNA network in prefrontal cortex of rats: relevance to depression pathophysiology', *Translational psychiatry*. Nature Publishing Group, 5(11), pp. e682–e682. doi: 10.1038/tp.2015.175.

Dyson, J. M. *et al.* (2017) 'INPP5E regulates phosphoinositide-dependent cilia transition zone function', *The Journal of cell biology*. 2016/12/20. The Rockefeller University Press, 216(1), pp. 247–263. doi: 10.1083/jcb.201511055.

Eddy, S. R. (2012) 'The C-value paradox, junk DNA and ENCODE', *Current Biology*. Elsevier,

22(21), pp. R898–R899. doi: 10.1016/j.cub.2012.10.002.

Edgar, R. C. (2004) 'MUSCLE: Multiple sequence alignment with high accuracy and high throughput', *Nucleic Acids Research*, 32(5), pp. 1792–1797. doi: 10.1093/nar/gkh340.

Elton, T. S. *et al.* (2013) 'Regulation of the MIR155 host gene in physiological and pathological processes', *Gene*, 532(1), pp. 1–12. doi: https://doi.org/10.1016/j.gene.2012.12.009.

Emera, D., Romero, R. and Wagner, G. (2012) 'The evolution of menstruation: a new model for genetic assimilation: explaining molecular origins of maternal responses to fetal invasiveness', *BioEssays : news and reviews in molecular, cellular and developmental biology*. 2011/11/07, 34(1), pp. 26–35. doi: 10.1002/bies.201100099.

Emera, D. and Wagner, G. P. (2012) 'Transposable element recruitments in the mammalian placenta: Impacts and mechanisms', *Briefings in Functional Genomics*, 11(4), pp. 267–276. doi: 10.1093/bfgp/els013.

Emiliani, S. *et al.* (2005) 'Embryo-maternal interactive factors regulating the implantation process: Implications in assisted reproductive treatment', *Reproductive BioMedicine Online*. Reproductive Healthcare Ltd, Duck End Farm, Dry Drayton, Cambridge CB23 8DB, UK, 10(4), pp. 527–540. doi: 10.1016/S1472-6483(10)60831-0.

Enders, Allen C. and Blankenship, T. N. (1999) 'Comparative placental structure', *Advanced Drug Delivery Reviews*, 38(1), pp. 3–15. doi: 10.1016/S0169-409X(99)00003-4.

Enders, Allen C and Blankenship, T. N. (1999) 'Vascular Invasion During Implantation and Placentation - Embryo Implantation: Molecular, Cellular and Clinical Aspects', in Carson, D. D. (ed.). New York, NY: Springer New York, pp. 23–38. doi: 10.1007/978-1-4612-1548-6_3.

Enright, A. J. *et al.* (2004) 'MicroRNA targets in Drosophila', *Genome Biology*. London: BioMed Central, 5(1), pp. R1–R1. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC395733/.

Esnault, C. *et al.* (2013) 'Differential Evolutionary Fate of an Ancestral Primate Endogenous Retrovirus Envelope Gene, the EnvV Syncytin, Captured for a Function in Placentation', *PLOS Genetics*. Public Library of Science, 9(3), p. e1003400. Available at: https://doi.org/10.1371/journal.pgen.1003400.

Everaert, C. *et al.* (2017) 'Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data', *Scientific reports*. Nature Publishing Group UK, 7(1), p. 1559. doi: 10.1038/s41598-017-01617-3.

Eyre-Walker, A. (2006) 'The genomic rate of adaptive evolution', *Trends in Ecology & Evolution*, 21(10), pp. 569–575. doi: https://doi.org/10.1016/j.tree.2006.06.015.

Faas, M. M. and De Vos, P. (2018) 'Innate immune cells in the placental bed in healthy pregnancy and preeclampsia', *Placenta*, 69, pp. 125–133. doi: https://doi.org/10.1016/j.placenta.2018.04.012.

Fair, T. (2016) 'Embryo maternal immune interactions in cattle', *Animal Reproduction*, 13(3), pp. 346–354. doi: 10.21451/1984-3143-AR877.

Fallen, S. *et al.* (2018) 'Extracellular vesicle RNAs reflect placenta dysfunction and are a biomarker source for preterm labour', *Journal of cellular and molecular medicine*. 2018/03/08. John Wiley and Sons Inc., 22(5), pp. 2760–2773. doi: 10.1111/jcmm.13570.

Fan, X. and Kurgan, L. (2015) 'Comprehensive overview and assessment of computational prediction of microRNA targets in animals', *Briefings in Bioinformatics*, 16(5), pp. 780–794. Available at: http://dx.doi.org/10.1093/bib/bbu044.

Farrokhnia, F. *et al.* (2014) 'MicroRNA regulation of mitogenic signaling networks in the human placenta', *The Journal of biological chemistry*. 2014/07/30. American Society for Biochemistry and Molecular Biology, 289(44), pp. 30404–30416. doi: 10.1074/jbc.M114.587295.

Fendler, W. *et al.* (2016) 'Differential regulation of serum microRNA expression by HNF1β and HNF1α transcription factors', *Diabetologia*. 2016/04/08. Springer Berlin Heidelberg, 59(7), pp. 1463–1473. doi: 10.1007/s00125-016-3945-0.

Fernandez-Capetillo, O. (2016) 'The (elusive) role of the SMC5/6 complex', *Cell cycle (Georgetown, Tex.)*. Taylor & Francis, 15(6), pp. 775–776. doi: 10.1080/15384101.2015.1137713.

Fishel, S. B., Edwards, R. G. and Evans, C. J. (1984) 'Human chorionic gonadotropin secreted by preimplantation embryos cultured in vitro', *Science*, 223(4638), pp. 816 LP – 818. doi: 10.1126/science.6546453.

Floridon, C. *et al.* (2000) 'Localization of E-cadherin in villous, extravillous and vascular trophoblasts during intrauterine, ectopic and molar pregnancy', *Molecular Human Reproduction*, 6(10), pp. 943–950. doi: 10.1093/molehr/6.10.943.

Fluhr, H. *et al.* (2008) 'Human chorionic gonadotropin stimulates matrix metalloproteinases-2 and -9 in cytotrophoblastic cells and decreases tissue inhibitor of metalloproteinases-1, -2, and

-3 in decidualized endometrial stromal cells.', *Fertility and sterility*. United States, 90(4 Suppl), pp. 1390–1395. doi: 10.1016/j.fertnstert.2007.08.023.

Foley, N. M. *et al.* (2018) 'Growing old, yet staying young: The role of telomeres in bats' exceptional longevity', *Science Advances*, 4(2), pp. 1–12. doi: 10.1126/sciadv.aao0926.

Foote, A. D. *et al.* (2015) 'Convergent evolution of the genomes of marine mammals', *Nature Genetics*, 47(3), pp. 272–275. doi: 10.1038/ng.3198.

Forde, N. *et al.* (2009) 'Progesterone-Regulated Changes in Endometrial Gene Expression Contribute to Advanced Conceptus Development in Cattle1', *Biology of Reproduction*, 81(4), pp. 784–794. doi: 10.1095/biolreprod.108.074336.

Forde, N. *et al.* (2011) 'Conceptus-Induced Changes in the Endometrial Transcriptome: How Soon Does the Cow Know She Is Pregnant?1', *Biology of Reproduction*, 85(1), pp. 144–156. doi: 10.1095/biolreprod.110.090019.

Forde, N. *et al.* (2012) 'Evidence for an early endometrial response to pregnancy in cattle: both dependent upon and independent of interferon tau', *Physiological Genomics*. American Physiological Society, 44(16), pp. 799–810. doi: 10.1152/physiolgenomics.00067.2012.

Forde, N. *et al.* (2015) '"Conceptualizing" the Endometrium: Identification of Conceptus-Derived Proteins During Early Pregnancy in Cattle', *Biology of reproduction*. 2015/05/06. Society for the Study of Reproduction, Inc., 92(6), p. 156. doi: 10.1095/biolreprod.115.129296.

Freyer, C. and Renfree, M. B. (2009) 'The mammalian yolk sac placenta', *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 312B(6), pp. 545–554. doi: 10.1002/jez.b.21239.

Fromm, B. *et al.* (2015) 'A Uniform System for the Annotation of Vertebrate microRNA Genes and the Evolution of the Human microRNAome', *Annual Review of Genetics*. Annual Reviews, 49(1), pp. 213–242. doi: 10.1146/annurev-genet-120213-092023.

Fromm, B. *et al.* (2018) 'MirGeneDB2.0: the curated microRNA Gene Database', *bioRxiv*.

Fu, G. *et al.* (2013) 'MicroRNAs in human placental development and pregnancy complications', *International Journal of Molecular Sciences*, 14(3), pp. 5519–5544. doi: 10.3390/ijms14035519.

Furukawa, S., Kuroda, Y. and Sugiyama, A. (2014) 'A comparison of the histological structure of the placenta in experimental animals.', *Journal of toxicologic pathology*, 27(1), pp. 11–8. doi: 10.1293/tox.2013-0060.

Gal, H. *et al.* (2019) 'Molecular pathways of senescence regulate placental structure and function', *The EMBO journal*. 2019/08/19. John Wiley and Sons Inc., 38(18), pp. e100849–e100849. doi: 10.15252/embj.2018100849.

Garratt, M. *et al.* (2013) 'Diversification of the eutherian placenta is associated with changes in the pace of life', *Proc. Natl. Acad. Sci. USA*, 110(19), pp. 7760–7765. doi: 10.1073/pnas.1305018110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1305018110.

Gebert, L. F. R. and MacRae, I. J. (2019) 'Regulation of microRNA function in animals', *Nature Reviews Molecular Cell Biology*, 20(1), pp. 21–37. doi: 10.1038/s41580-018-0045-7.

Gharib, W. H. and Robinson-Rechavi, M. (2013) 'The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in GC', *Molecular biology and evolution*. 2013/04/04. Oxford University Press, 30(7), pp. 1675–1686. doi: 10.1093/molbev/mst062.

Gibbs, G. M., Roelants, K. and O'Bryan, M. K. (2008) 'The CAP Superfamily: Cysteine-Rich Secretory Proteins, Antigen 5, and Pathogenesis-Related 1 Proteins—Roles in Reproduction, Cancer, and Immune Defense', *Endocrine Reviews*, 29(7), pp. 865–897. doi: 10.1210/er.2008-0032.

Gilbert, S. (2000) 'Fusion of the Genetic Material', in *Developmental Biology*. 6th edn. Sunderland (MA): Sinauer Associates. Available at: https://www.ncbi.nlm.nih.gov/books/NBK10055/.

Gillet, V. *et al.* (2019) 'miRNA Profiles in Extracellular Vesicles From Serum Early in Pregnancies Complicated by Gestational Diabetes Mellitus', *The Journal of Clinical Endocrinology & Metabolism*, 104(11), pp. 5157–5169. doi: 10.1210/jc.2018-02693.

Giribet, G. (2005) 'TNT: Tree Analysis Using New Technology', *Systematic Biology*, 54(1), pp. 176–178.

GIUDICE, L. C. *et al.* (1990) 'Insulin-Like Growth Factor Binding Proteins in Maternal Serum Throughout Gestation and in the Puerperium: Effects of a Pregnancy-Associated Serum Protease Activity*', *The Journal of Clinical Endocrinology & Metabolism*, 71(4), pp. 806–816. doi: 10.1210/jcem-71-4-806.

Godkin, J. D. *et al.* (1982) 'Purification and properties of a major, low molecular weight protein

released by the trophoblast of sheep blastocysts at Day 13–21', *Reproduction*. Bristol, UK: Bioscientifica Ltd, 65(1), pp. 141–150. doi: 10.1530/jrf.0.0650141.

Goldman, N. and Yang, Z. (1994) 'A codon-based model of nucleotide substitution for protein-coding DNA sequences.', *Molecular Biology and Evolution*, 11(5), pp. 725–736. doi: 10.1093/oxfordjournals.molbev.a040153.

Gong, M. *et al.* (2019) 'A transcriptomic analysis of Nsmce1 overexpression in mouse hippocampal neuronal cell by RNA sequencing', *Functional & Integrative Genomics*. doi: 10.1007/s10142-019-00728-6.

Graham, J. D. and Clarke, C. L. (1997) 'Physiological Action of Progesterone in Target Tissues*', *Endocrine Reviews*, 18(4), pp. 502–519. doi: 10.1210/edrv.18.4.0308.

Greening, D. W. *et al.* (2016) 'Modulating the endometrial epithelial proteome and secretome in preparation for pregnancy: The role of ovarian steroid and pregnancy hormones', *Journal of Proteomics*, 144, pp. 99–112. doi: https://doi.org/10.1016/j.jprot.2016.05.026.

Griffiths, S. K. and Campbell, J. P. (2015) 'Placental structure, function and drug transfer', *Continuing Education in Anaesthesia, Critical Care and Pain*, 15(2), pp. 84–89. doi: 10.1093/bjaceaccp/mku013.

GrÜnweiler, A. and Hartmann, R. K. (2007) 'Locked Nucleic Acid Oligonucleotides', *BioDrugs*, 21(4), pp. 235–243. doi: 10.2165/00063030-200721040-00004.

Gu, Y. *et al.* (2013) 'Differential miRNA expression profiles between the first and third trimester human placentas', *American Journal of Physiology - Endocrinology and Metabolism*. Bethesda, MD: American Physiological Society, 304(8), pp. E836–E843. doi: 10.1152/ajpendo.00660.2012.

Guillomot, M. (1995) 'Cellular interactions during implantation in domestic ruminants.', *Journal of reproduction and fertility. Supplement*, 49, pp. 39–51. doi: 10.1530/biosciprocs.3.004.

Gulyaeva, L. F. and Kushlinskiy, N. E. (2016) 'Regulatory mechanisms of microRNA expression', *Journal of Translational Medicine*, 14(1), p. 143. doi: 10.1186/s12967-016-0893-x.

Guodong, F. *et al.* (2013) 'MicroRNA-376c Impairs Transforming Growth Factor-β and Nodal Signaling to Promote Trophoblast Cell Proliferation and Invasion', *Hypertension*. American Heart Association, 61(4), pp. 864–872. doi: 10.1161/HYPERTENSIONAHA.111.203489.

Haig, D. (1996) 'Gestational drive and the green-bearded placenta', *Proceedings of the National Academy of Sciences of the United States of America*, 93(13), pp. 6547–6551. doi:

10.1073/pnas.93.13.6547.

Haig, D. (2012) 'Retroviruses and the placenta', *Current Biology*. Elsevier Ltd, 22(15), pp. R609–R613. doi: 10.1016/j.cub.2012.06.002.

Halasz, M. and Szekeres-Bartho, J. (2013) 'The role of progesterone in implantation and trophoblast invasion', *Journal of Reproductive Immunology*, 97(1), pp. 43–50. doi: https://doi.org/10.1016/j.jri.2012.10.011.

Halliday, T. J. D. and Goswami, A. (2016) 'Eutherian morphological disparity across the end-Cretaceous mass extinction', *Biological Journal of the Linnean Society*. John Wiley & Sons, Ltd, 118(1), pp. 152–168. doi: 10.1111/bij.12731.

Hallström, B. M. *et al.* (2007) 'Phylogenomic data analyses provide evidence that Xenarthra and Afrotheria are sister groups', *Molecular Biology and Evolution*, 24(9), pp. 2059–2068. doi: 10.1093/molbev/msm136.

Hallström, B. M. *et al.* (2011) 'A genomic approach to examine the complex evolution of laurasiatherian mammals', *PloS one*. 2011/12/02. Public Library of Science, 6(12), pp. e28199–e28199. doi: 10.1371/journal.pone.0028199.

Hallström, B. M. and Janke, A. (2008) 'Resolution among major placental mammal interordinal relationships with genome data imply that speciation influenced their earliest radiations', *BMC Evolutionary Biology*, 8(1), p. 162. doi: 10.1186/1471-2148-8-162.

Han, J. *et al.* (2004) 'The Drosha-DGCR8 complex in primary microRNA processing', *Genes & Development* , 18(24), pp. 3016–3027. doi: 10.1101/gad.1262504.

Han, S. W., Lei, Z. M. and Rao, C. V. (1999) 'Treatment of human endometrial stromal cells with chorionic gonadotropin promotes their morphological and functional differentiation into decidua', *Molecular and Cellular Endocrinology*, 147(1), pp. 7–16. doi: https://doi.org/10.1016/S0303-7207(98)00240-8.

Harapan, H. and Andalas, M. (2015) 'The role of microRNAs in the proliferation, differentiation, invasion, and apoptosis of trophoblasts during the occurrence of preeclampsia—A systematic review', *Tzu Chi Medical Journal*, 27(2), pp. 54–64. doi: https://doi.org/10.1016/j.tcmj.2015.05.001.

Hauguel-de Mouzon, S. and Guerre-Millo, M. (2006) 'The Placenta Cytokine Network and Inflammatory Signals', *Placenta*, 27(8), pp. 794–798. doi:

https://doi.org/10.1016/j.placenta.2005.08.009.

Hawkins, J. A. *et al.* (2019) 'A metaanalysis of bat phylogenetics and positive selection based on genomes and transcriptomes from 18 species', *Proceedings of the National Academy of Sciences*, 116(23), pp. 11351 LP – 11360. doi: 10.1073/pnas.1814995116.

Heimberg, A. M. *et al.* (2008) 'MicroRNAs and the advent of vertebrate morphological complexity', *Proceedings of the National Academy of Sciences of the United States of America*, 105(8), pp. 2946–50. doi: 10.1073/pnas.0712259105.

Hellstrom, M. *et al.* (1999) 'Role of PDGF-B and PDGFR-beta in recruitment of vascular smooth muscle cells and pericytes during embryonic blood vessel formation in the mouse', *Development*, 126(14), pp. 3047 LP – 3055. Available at: http://dev.biologists.org/content/126/14/3047.abstract.

Hertel, J. *et al.* (2006) 'The expansion of the metazoan microRNA repertoire', *BMC Genomics*, 7(1), p. 25. doi: 10.1186/1471-2164-7-25.

Hilgers, L. *et al.* (2018) 'Novel Genes, Ancient Genes, and Gene Co-Option Contributed to the Genetic Basis of the Radula, a Molluscan Innovation', *Molecular Biology and Evolution*, 35(7), pp. 1638–1652. doi: 10.1093/molbev/msy052.

Hinman, V. F. *et al.* (2003) 'Developmental gene regulatory network architecture across 500 million years of echinoderm evolution', *Proceedings of the National Academy of Sciences*, 100(23), pp. 13356 LP – 13361. doi: 10.1073/pnas.2235868100.

Hornstein, E. *et al.* (2005) 'The microRNA miR-196 acts upstream of Hoxb8 and Shh in limb development', *Nature*. Nature Publishing Group, 438, p. 671. Available at: http://dx.doi.org/10.1038/nature04138.

Hosseini, M. K. *et al.* (2018) 'MicroRNA expression profiling in placenta and maternal plasma in early pregnancy loss', *Molecular medicine reports*. 2018/01/31. D.A. Spandidos, 17(4), pp. 4941–4952. doi: 10.3892/mmr.2018.8530.

Hu, J. *et al.* (2010) 'Cellular cholesterol delivery, intracellular processing and utilization for biosynthesis of steroid hormones', *Nutrition & metabolism*. BioMed Central, 7, p. 47. doi: 10.1186/1743-7075-7-47.

Hu, X.-Q. and Zhang, L. (2017) 'Angiogenesis during pregnancy: all routes lead to MAPKs', *The Journal of physiology*. 2017/06/08. John Wiley and Sons Inc., 595(14), pp. 4571–4572. doi:

10.1113/JP274489.

Hue, I., Degrelle, S. A. and Turenne, N. (2012) 'Conceptus elongation in cattle: Genes, models and questions', *Animal Reproduction Science*, 134(1), pp. 19–28. doi: https://doi.org/10.1016/j.anireprosci.2012.08.007.

Hughes, A. L. and Friedman, R. (2008) 'Codon-based tests of positive selection, branch lengths, and the evolution of mammalian immune system genes', *Immunogenetics*, 60(9), p. 495. doi: 10.1007/s00251-008-0304-4.

Hunkapiller, N. M. *et al.* (2011) 'A role for Notch signaling in trophoblast endovascular invasion and in the pathogenesis of pre-eclampsia', *Development (Cambridge, England)*. Company of Biologists, 138(14), pp. 2987–2998. doi: 10.1242/dev.066589.

Ioannidis, J. and Donadeu, F. X. (2017) 'Changes in circulating microRNA levels can be identified as early as day 8 of pregnancy in cattle', *PloS one*. Public Library of Science, 12(4), pp. e0174892–e0174892. doi: 10.1371/journal.pone.0174892.

Ireland, J. J., Murphee, R. L. and Coulson, P. B. (1980) 'Accuracy of Predicting Stages of Bovine Estrous Cycle by Gross Appearance of the Corpus Luteum', *Journal of Dairy Science*, 63(1), pp. 155–160. doi: 10.3168/jds.S0022-0302(80)82901-8.

Ito, M. *et al.* (2015a) 'A trans-homologue interaction between reciprocally imprinted miR-127 and Rtl1 regulates placenta development', *Development*, 142(14), pp. 2425 LP – 2430. Available at: http://dev.biologists.org/content/142/14/2425.abstract.

Ito, M. *et al.* (2015b) 'A trans-homologue interaction between reciprocally imprinted miR-127 and Rtl1 regulates placenta development', *Development (Cambridge, England)*. The Company of Biologists, 142(14), pp. 2425–2430. doi: 10.1242/dev.121996.

Jacobs, E. E. and Sanadi, D. R. (1960) 'The reversible removal of cytochrome c from mitochondria.', *The Journal of biological chemistry*, 235(2), pp. 531–534.

Jassal, B. *et al.* (2019) 'The reactome pathway knowledgebase', *Nucleic Acids Research*, 48(D1), pp. D498–D503. doi: 10.1093/nar/gkz1031.

Jirkovska, M., Korabecna, M. and Lassakova, S. (2019) 'Telomeres and Telomerase Activity in the Human Placenta', in Morrish, D. T. A. (ed.) *Telomerase and non-Telomerase Mechanisms of Telomere Maintenance*. IntechOpen. doi: DOI: 10.5772/intechopen.86327.

Johnson, W. E. (2019) 'Origins and evolutionary consequences of ancient endogenous retroviruses', *Nature Reviews Microbiology*, 17(6), pp. 355–370. doi: 10.1038/s41579-019-0189-2.

Kadokawa, Y. *et al.* (1989) 'Expression Pattern of E- and P-Cadherin in Mouse Embryos and Uteri during the Periimplantation Period', *Development, Growth & Differentiation*. John Wiley & Sons, Ltd, 31(1), pp. 23–30. doi: 10.1111/j.1440-169X.1989.00023.x.

Kaneko-Ishino, T. and Ishino, F. (2012) 'The role of genes domesticated from LTR retrotransposons and retroviruses in mammals', *Frontiers in Microbiology*. Frontiers Research Foundation, 3, p. 262. doi: 10.3389/fmicb.2012.00262.

Karmakar, S. and Das, C. (2002) 'Regulation of Trophoblast Invasion by IL-1β and TGF- β1', *American Journal of Reproductive Immunology*. John Wiley & Sons, Ltd, 48(4), pp. 210–219. doi: 10.1034/j.1600-0897.2002.01151.x.

Katoh, K. *et al.* (2002) 'MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.', *Nucleic acids research*, 30(14), pp. 3059–3066. doi: 10.1093/nar/gkf436.

Kawamata, T. and Tomari, Y. (2010) 'Making RISC', *Trends in Biochemical Sciences*. Elsevier, 35(7), pp. 368–376. doi: 10.1016/j.tibs.2010.03.009.

Kaya, H. S. *et al.* (2015) 'Roles of progesterone receptor A and B isoforms during human endometrial decidualization', *Molecular Endocrinology*, 29(6), pp. 882–895. doi: 10.1210/me.2014-1363.

Keniry, A. *et al.* (2012) 'The H19 lincRNA is a developmental reservoir of miR-675 which suppresses growth and Igf1r', *Nature cell biology*, 14(7), pp. 659–665. doi: 10.1038/ncb2521.

Kertesz, M. *et al.* (2007) 'The role of site accessibility in microRNA target recognition', *Nature Genetics*, 39(10), pp. 1278–1284. doi: 10.1038/ng2135.

Khalaj, K. *et al.* (2019) 'Extracellular vesicles from endometriosis patients are characterized by a unique miRNA-lncRNA signature', *JCI insight*. American Society for Clinical Investigation, 4(18), p. e128846. doi: 10.1172/jci.insight.128846.

Kim, S.-K. *et al.* (2006) 'miTarget: microRNA target gene prediction using a support vector machine', *BMC Bioinformatics*. London: BioMed Central, 7, p. 411. doi: 10.1186/1471-2105-7-411.

Kim, S.-M. and Kim, J.-S. (2017) ‘A Review of Mechanisms of Implantation’, *Development & reproduction*. 2017/12/31. The Korean Society of Developmental Biology, 21(4), pp. 351–359. doi: 10.12717/DR.2017.21.4.351.

Kimura, M. (1968) ‘Evolutionary Rate at the Molecular Level’, *Nature*, 217(5129), pp. 624–626. doi: 10.1038/217624a0.

Kimura, M. (1980) ‘A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences’, *Journal of Molecular Evolution*, 16(2), pp. 111–120. doi: 10.1007/BF01731581.

Kimura, M. (1991) ‘The neutral theory of molecular evolution: A review of recent evidence’, *The Japanese Journal of Genetics*, pp. 367–386. doi: 10.1266/jjg.66.367.

Kimura, M. and Crow, J. F. (1964) ‘The Number of Alleles That Can Be Maintained in a Finite Population’, *Genetics*, 49(4), pp. 725–738. Available at: https://pubmed.ncbi.nlm.nih.gov/14156929.

Kimura, M. and Ohta, T. (1974) ‘On some principles governing molecular evolution’, *Proceedings of the National Academy of Sciences of the United States of America*, 71(7), pp. 2848–2852. doi: 10.1073/pnas.71.7.2848.

Kin, K. *et al.* (2016) ‘The transcriptomic evolution of mammalian pregnancy: gene expression innovations in endometrial stromal fibroblasts’, *Genome Biology and Evolution*, p. evw168. doi: 10.1093/gbe/evw168.

King, J. L. and Jukes, T. H. (1969) ‘Non-Darwinian Evolution’, *Science*, 164(3881), pp. 788 LP – 798. doi: 10.1126/science.164.3881.788.

Klohonatz, K. M. *et al.* (2019) ‘Non-Coding RNA Sequencing of Equine Endometrium During Maternal Recognition of Pregnancy’, *Genes*. MDPI, 10(10), p. 821. doi: 10.3390/genes10100821.

Knutson, T. P. *et al.* (2012) ‘Phosphorylated and sumoylation-deficient progesterone receptors drive proliferative gene signatures during breast cancer progression’, *Breast cancer research : BCR*. BioMed Central, 14(3), pp. R95–R95. doi: 10.1186/bcr3211.

Koppers, A. J., Reddy, T. and O’Bryan, M. K. (2011) ‘The role of cysteine-rich secretory proteins in male fertility’, *Asian journal of andrology*. 2010/10/25. Nature Publishing Group, 13(1), pp. 111–117. doi: 10.1038/aja.2010.77.

Kosakovsky Pond, S. L. *et al.* (2011) ‘A random effects branch-site model for detecting episodic

diversifying selection', *Molecular biology and evolution*. 2011/06/13. Oxford University Press, 28(11), pp. 3033–3043. doi: 10.1093/molbev/msr125.

Kosakovsky Pond, S. L. *et al.* (2019) 'HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies', *Molecular Biology and Evolution*, 37(1), pp. 295–299. doi: 10.1093/molbev/msz197.

Kosakovsky Pond, S. L. and Frost, S. D. W. (2005) 'Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection', *Molecular Biology and Evolution*, 22(5), pp. 1208–1222. doi: 10.1093/molbev/msi105.

Koscianska, E. *et al.* (2015) 'Cooperation meets competition in microRNA-mediated DMPK transcript regulation', *Nucleic Acids Research*, 43(19), pp. 9500–9518. doi: 10.1093/nar/gkv849.

Kosiol, C. *et al.* (2008) 'Patterns of Positive Selection in Six Mammalian Genomes', *PLOS Genetics*. Public Library of Science, 4(8), p. e1000144. Available at: https://doi.org/10.1371/journal.pgen.1000144.

Ku, C. W. *et al.* (2018) 'Serum progesterone distribution in normal pregnancies compared to pregnancies complicated by threatened miscarriage from 5 to 13 weeks gestation: a prospective cohort study', *BMC pregnancy and childbirth*. BioMed Central, 18(1), p. 360. doi: 10.1186/s12884-018-2002-z.

Kubiczkova, L. *et al.* (2012) 'TGF-β – an excellent servant but a bad master', *Journal of Translational Medicine*, 10(1), p. 183. doi: 10.1186/1479-5876-10-183.

Kuhn, D. E. *et al.* (2008) 'Experimental validation of miRNA targets', *Methods (San Diego, Calif.)*, 44(1), pp. 47–54. doi: 10.1016/j.ymeth.2007.09.005.

Kumar, P. *et al.* (2013) 'The c-Myc-regulated microRNA-17~92 (miR-17~92) and miR-106a~363 clusters target hCYP19A1 and hGCM1 to inhibit human trophoblast differentiation', *Molecular and cellular biology*. 2013/02/25. American Society for Microbiology, 33(9), pp. 1782–1796. doi: 10.1128/MCB.01228-12.

Kumar, P. and Magon, N. (2012) 'Hormones in pregnancy', *Nigerian medical journal : journal of the Nigeria Medical Association*. Medknow Publications & Media Pvt Ltd, 53(4), pp. 179–183. doi: 10.4103/0300-1652.107549.

Kuokkanen, S. *et al.* (2010) 'Genomic Profiling of MicroRNAs and Messenger RNAs Reveals Hormonal Regulation in MicroRNA Expression in Human Endometrium1', *Biology of*

Reproduction, 82(4), pp. 791–801. doi: 10.1095/biolreprod.109.081059.

Kurita, T. *et al.* (1998) 'Stromal Progesterone Receptors Mediate the Inhibitory Effects of Progesterone on Estrogen-Induced Uterine Epithelial Cell Deoxyribonucleic Acid Synthesis**This work was supported by NIH Grants AG-13784 and HD-07857.', *Endocrinology*, 139(11), pp. 4708–4713. doi: 10.1210/endo.139.11.6317.

Kutzler, M. A. (2015) 'Alternative methods for feline fertility control: Use of melatonin to suppress reproduction.', *Journal of feline medicine and surgery*. England, 17(9), pp. 753–757. doi: 10.1177/1098612X15594988.

Laddha, S. V *et al.* (2013) 'Genome-wide analysis reveals downregulation of miR-379/miR-656 cluster in human cancers.', *Biology direct*. Biology Direct, 8(1), p. 10. doi: 10.1186/1745-6150-8-10.

Lan, T., Wang, X.-R. and Zeng, Q.-Y. (2013) 'Structural and functional evolution of positively selected sites in pine glutathione S-transferase enzyme family', *The Journal of biological chemistry*. 2013/07/11. American Society for Biochemistry and Molecular Biology, 288(34), pp. 24441–24451. doi: 10.1074/jbc.M113.456863.

de Lange, T. (2002) 'Protection of mammalian telomeres', *Oncogene*, 21(4), pp. 532–540. doi: 10.1038/sj.onc.1205080.

Langley, C. H. and Fitch, W. M. (1974) 'An examination of the constancy of the rate of molecular evolution', *Journal of Molecular Evolution*, 3(3), pp. 161–177. doi: 10.1007/BF01797451.

Large, M. J. *et al.* (2014) 'The epidermal growth factor receptor critically regulates endometrial function during early pregnancy', *PLoS genetics*. Public Library of Science, 10(6), pp. e1004451–e1004451. doi: 10.1371/journal.pgen.1004451.

Lee, J. Y., Lee, M. and Lee, S. K. (2011) 'Role of endometrial immune cells in implantation', *Clinical and experimental reproductive medicine*. 2011/09/30. The Korean Society for Reproductive Medicine, 38(3), pp. 119–125. doi: 10.5653/cerm.2011.38.3.119.

Lee, K. *et al.* (2006) 'Molecular mechanisms involved in progesterone receptor regulation of uterine function', *The Journal of Steroid Biochemistry and Molecular Biology*, 102(1), pp. 41–50. doi: https://doi.org/10.1016/j.jsbmb.2006.09.006.

Lee, R. C., Feinbaum, R. L. and Ambros, V. (1993) 'The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14', *Cell*, 75(5), pp. 843–854. doi:

https://doi.org/10.1016/0092-8674(93)90529-Y.

van der Lee, R. *et al.* (2017) 'Genome-scale detection of positive selection in nine primates predicts human-virus evolutionary conflicts', *Nucleic acids research*. Oxford University Press, 45(18), pp. 10634–10648. doi: 10.1093/nar/gkx704.

Lee, Y. *et al.* (2002) 'MicroRNA maturation: stepwise processing and subcellular localization', *The EMBO Journal*, 21(17), pp. 4663 LP – 4670. Available at: http://emboj.embopress.org/content/21/17/4663.abstract.

Levene, H. (1960) 'Robust tests for equality of variances.', in *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Palo Alto, CA: Stanford University Press (Stanford studies in mathematics and statistics), pp. 278–292. Available at: https://books.google.co.uk/books?id=ZUSsAAAAIAAJ.

Lewis, B. P., Burge, C. B. and Bartel, D. P. (2005a) 'Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets', *Cell*. Elsevier, 120(1), pp. 15–20. doi: 10.1016/j.cell.2004.12.035.

Lewis, B. P., Burge, C. B. and Bartel, D. P. (2005b) 'Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets', *Cell*. Elsevier, 120(1), pp. 15–20. doi: 10.1016/j.cell.2004.12.035.

Lewontin, R. C. (1974) *The Genetic Basis of Evolutionary Change*. New York, NY: Columbia University Press.

Li, L. *et al.* (2020) 'Transcriptome sequencing of endometrium revealed alterations in mRNAs and lncRNAs after ovarian stimulation.', *Journal of assisted reproduction and genetics*, 37(1), pp. 21–32. doi: 10.1007/s10815-019-01616-5.

Li, M. *et al.* (2019) 'Compound heterozygous variants in MOGS inducing congenital disorders of glycosylation (CDG) IIb', *Journal of Human Genetics*, 64(3), pp. 265–268. doi: 10.1038/s10038-018-0552-6.

Li, Q. (2014) 'Transforming growth factor β signaling in uterine development and function', *Journal of animal science and biotechnology*. BioMed Central, 5(1), p. 52. doi: 10.1186/2049-1891-5-52.

Li, W. H. *et al.* (1990) 'Molecular phylogeny of Rodentia, Lagomorpha, Primates, Artiodactyla, and Carnivora and molecular clocks', *Proceedings of the National Academy of Sciences of the*

United States of America, 87(17), pp. 6703–6707. doi: 10.1073/pnas.87.17.6703.

Li, W. H. (1993) 'Unbiased estimation of the rates of synonymous and nonsynonymous substitution', *Journal of Molecular Evolution*, 36(1), pp. 96–99. doi: 10.1007/BF02407308.

Li, W. H., Wu, C. I. and Luo, C. C. (1985) 'A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes.', *Molecular Biology and Evolution*, 2(2), pp. 150–174. doi: 10.1093/oxfordjournals.molbev.a040343.

Licht, P., Russu, V. and Wildt, L. (2001) 'On the Role of Human Chorionic Gonadotropin (hCG) in the Embryo-Endometrial Microenvironment: Implications for Differentiation and Implantation', *Semin Reprod Med*, 19(01), pp. 37–48. doi: 10.1055/s-2001-13909.

Lillegraven, J. A. *et al.* (1987) 'The origin of eutherian mammals', *Biological Journal of the Linnean Society*, 32(3), pp. 281–336. doi: 10.1111/j.1095-8312.1987.tb00434.x.

Lim, H. *et al.* (1997) 'Multiple Female Reproductive Failures in Cyclooxygenase 2 Deficient Mice', *Cell*. Elsevier, 91(2), pp. 197–208. doi: 10.1016/S0092-8674(00)80402-X.

Lin, C.-C. *et al.* (2011) 'A KLF4-miRNA-206 autoregulatory feedback loop can promote or inhibit protein translation depending upon cell context', *Molecular and cellular biology*. 2011/04/25. American Society for Microbiology, 31(12), pp. 2513–2527. doi: 10.1128/MCB.01189-10.

Liu, N. *et al.* (2019) 'Telomere dysfunction impairs epidermal stem cell specification and differentiation by disrupting BMP/pSmad/P63 signaling', *PLOS Genetics*. Public Library of Science, 15(9), p. e1008368. Available at: https://doi.org/10.1371/journal.pgen.1008368.

Livak, K. J. and Schmittgen, T. D. (2001) 'Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method', *Methods*, 25(4), pp. 402–408. doi: https://doi.org/10.1006/meth.2001.1262.

Lombardi, J. (1998) 'Reproductive Regulation', in *Comparative Vertebrate Reproduction*. Springer, Boston, MA, pp. 155–187. doi: https://doi.org/10.1007/978-1-4615-4937-6.

Lonergan, P. (2011) 'Influence of progesterone on oocyte quality and embryo development in cows', *Theriogenology*. Elsevier Inc., 76(9), pp. 1594–1601. doi: 10.1016/j.theriogenology.2011.06.012.

Lonergan, P. and Forde, N. (2014) 'Maternal-embryo interaction leading up to the initiation of implantation of pregnancy in cattle', *animal*. 2014/03/28. Cambridge University Press, 8(s1), pp.

64–69. doi: DOI: 10.1017/S1751731114000470.

Loughran, N. B. *et al.* (2012) 'Functional consequence of positive selection revealed through rational mutagenesis of human myeloperoxidase', *Molecular Biology and Evolution*, 29(8), pp. 2039–2046. doi: 10.1093/molbev/mss073.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), p. 550. doi: 10.1186/s13059-014-0550-8.

Luo, L. *et al.* (2012) 'MicroRNA-378a-5p promotes trophoblast cell survival, migration and invasion by targeting Nodal', *Journal of Cell Science*, 125(13), pp. 3124 LP – 3132. Available at: http://jcs.biologists.org/content/125/13/3124.abstract.

Lv, Y. *et al.* (2016) 'miRNA and target gene expression in menstrual endometria and early pregnancy decidua', *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 197, pp. 27–30. doi: https://doi.org/10.1016/j.ejogrb.2015.11.003.

Lynch, M. (2007) 'The frailty of adaptive hypotheses for the origins of organismal complexity', *Proceedings of the National Academy of Sciences of the United States of America*. 2007/05/09. National Academy of Sciences, 104 Suppl(Suppl 1), pp. 8597–8604. doi: 10.1073/pnas.0702207104.

Lynch, V. J. *et al.* (2004) 'Adaptive evolution of HoxA-11 and HoxA-13 at the origin of the uterus in mammals', *Proceedings of the Royal Society B: Biological Sciences*, 271(1554), pp. 2201–2207. doi: 10.1098/rspb.2004.2848.

Lynch, V. J. *et al.* (2011) 'Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals.', *Nature genetics*. Nature Publishing Group, 43(11), pp. 1154–9. doi: 10.1038/ng.917.

Lynch, V. J. *et al.* (2015) 'Ancient transposable elements transformed the uterine regulatory landscape and transcriptome during the evolution of mammalian pregnancy', *Cell Reports*, 10(4), pp. 551–562. doi: 10.1016/j.celrep.2014.12.052.

Lyons, J. J., Milner, J. D. and Rosenzweig, S. D. (2015) 'Glycans Instructing Immunity: The Emerging Role of Altered Glycosylation in Clinical Immunology ', *Frontiers in Pediatrics* , p. 54. Available at: https://www.frontiersin.org/article/10.3389/fped.2015.00054.

Mahmoudi, S. *et al.* (2009) 'Wrap53, a Natural p53 Antisense Transcript Required for p53

Induction upon DNA Damage', *Molecular Cell*. Elsevier, 33(4), pp. 462–471. doi: 10.1016/j.molcel.2009.01.028.

Maillot, G. *et al.* (2009) 'Widespread Estrogen-Dependent Repression of microRNAs Involved in Breast Tumor Cell Growth', *Cancer Research*, 69(21), pp. 8332 LP – 8340. doi: 10.1158/0008-5472.CAN-09-2206.

Mamane, Y. *et al.* (1999) 'Interferon regulatory factors: the next generation', *Gene*, 237(1), pp. 1–14. doi: https://doi.org/10.1016/S0378-1119(99)00262-0.

Mansouri-Attia, N. *et al.* (2009) 'Gene expression profiles of bovine caruncular and intercaruncular endometrium at implantation', *Physiological Genomics*. American Physiological Society, 39(1), pp. 14–27. doi: 10.1152/physiolgenomics.90404.2008.

Markova-Raina, P. and Petrov, D. (2011) 'High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 Drosophila genomes', *Genome research*. 2011/03/10. Cold Spring Harbor Laboratory Press, 21(6), pp. 863–874. doi: 10.1101/gr.115949.110.

Martin, R. D. (2007) 'The evolution of human reproduction: A primatological perspective', *American Journal of Physical Anthropology*. John Wiley & Sons, Ltd, 134(S45), pp. 59–84. doi: 10.1002/ajpa.20734.

Martini, F. *et al.* (2017) *Fundamentals of Anatomy & Physiology*. 11th edn. Pearson.

Massingham, T. and Goldman, N. (2005) 'Detecting Amino Acid Sites Under Positive Selection and Purifying Selection', *Genetics*, 169(3), pp. 1753 LP – 1762. doi: 10.1534/genetics.104.032144.

Maston, G. A. and Ruvolo, M. (2002) 'Chorionic Gonadotropin Has a Recent Origin Within Primates and an Evolutionary History of Selection', *Molecular Biology and Evolution*, 19(3), pp. 320–335. Available at: http://dx.doi.org/10.1093/oxfordjournals.molbev.a004085.

Matsui, M. *et al.* (2013) 'Promoter RNA links transcriptional regulation of inflammatory pathway genes', *Nucleic acids research*. 2013/09/02. Oxford University Press, 41(22), pp. 10086–10109. doi: 10.1093/nar/gkt777.

Mayor, P. *et al.* (2019) 'Menstrual cycle in four New World primates: Poeppig's woolly monkey (Lagothrix poeppigii), red uakari (Cacajao calvus), large-headed capuchin (Sapajus macrocephalus) and nocturnal monkey (Aotus nancymaae).', *Theriogenology*. United States,

123, pp. 11–21. doi: 10.1016/j.theriogenology.2018.09.019.

McDonald, J. H. and Kreitman, M. (1991) 'Adaptive protein evolution at the Adh locus in Drosophila', *Nature*, 351(6328), pp. 652–654. doi: 10.1038/351652a0.

McGowen, M. R. *et al.* (2014) 'The evolution of embryo implantation', *The International journal of developmental biology*, 58(2–4), pp. 155–161. doi: 10.1387/ijdb.140020dw.

McNab, B. K. (2008) 'An analysis of the factors that influence the level and scaling of mammalian BMR', *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 151(1), pp. 5–28. doi: https://doi.org/10.1016/j.cbpa.2008.05.008.

Meredith, R. W. *et al.* (2011) 'Impacts of the Cretaceous Terrestrial Revolution and KPg Extinction on Mammal Diversification', *Science*, 334(6055), pp. 521–524. doi: 10.1126/science.1211028.

Mess, A. (2014) 'Placental Evolution within the Supraordinal Clades of Eutheria with the Perspective of Alternative Animal Models for Human Placentation', *Advances in Biology*, 2014, pp. 1–21. doi: 10.1155/2014/639274.

Messer, P. W. and Petrov, D. A. (2013) 'Frequent adaptation and the McDonald–Kreitman test', *Proceedings of the National Academy of Sciences*, 110(21), pp. 8615 LP – 8620. doi: 10.1073/pnas.1220835110.

Messinis, I. E., Messini, C. I. and Dafopoulos, K. (2014) 'Novel aspects of the endocrinology of the menstrual cycle', *Reproductive BioMedicine Online*. Reproductive Healthcare Ltd., 28(6), pp. 714–722. doi: 10.1016/j.rbmo.2014.02.003.

Mestdagh, P. *et al.* (2014) 'Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study', *Nature Methods*, 11(8), pp. 809–815. doi: 10.1038/nmeth.3014.

Michael Roberts, R., Xie, S. and Mathialagan, N. (1996) 'Maternal Recognition of Pregnancy', *Biology of Reproduction*, 54(2), pp. 294–302. doi: 10.1095/biolreprod54.2.294.

Miranda, K. C. *et al.* (2006) 'A Pattern-Based Method for the Identification of MicroRNA Binding Sites and Their Corresponding Heteroduplexes', *Cell*, 126(6), pp. 1203–1217. doi: https://doi.org/10.1016/j.cell.2006.07.031.

Misao, R. *et al.* (1998) 'Expression of progesterone receptor isoforms in corpora lutea of human subjects: correlation with serum oestrogen and progesterone concentrations.', *Molecular*

Human Reproduction, 4(11), pp. 1045–1052. doi: 10.1093/molehr/4.11.1045.

Mitchell-Olds, T., Willis, J. H. and Goldstein, D. B. (2007) 'Which evolutionary processes influence natural genetic variation for phenotypic traits?', *Nature reviews. Genetics*, 8(11), pp. 845–56. doi: 10.1038/nrg2207.

Mitxelena, J. *et al.* (2016) 'E2F7 regulates transcription and maturation of multiple microRNAs to restrain cell proliferation', *Nucleic acids research*. 2016/03/09. Oxford University Press, 44(12), pp. 5557–5570. doi: 10.1093/nar/gkw146.

Moldovan, L. *et al.* (2013) 'Analyzing the circulating microRNAs in exosomes/extracellular vesicles from serum or plasma by qRT-PCR', *Methods in molecular biology (Clifton, N.J.)*, 1024, pp. 129–145. doi: 10.1007/978-1-62703-453-1_10.

Monis, C. N. and Tetrokalashvili, M. (2019) 'Menstrual Cycle Proliferative And Follicular Phase'. StatPearls Publishing, Treasure Island (FL). Available at: http://europepmc.org/abstract/MED/31194386.

Montgomery, S. H. and Mundy, N. I. (2014) 'Microcephaly genes evolved adaptively throughout the evolution of eutherian mammals', *BMC Evolutionary Biology*, 14(1), p. 120. doi: 10.1186/1471-2148-14-120.

Mor, G. and Cardenas, I. (2010) 'The immune system in pregnancy: a unique complexity', *American journal of reproductive immunology (New York, N.Y. : 1989)*. 2010/03/29, 63(6), pp. 425–433. doi: 10.1111/j.1600-0897.2010.00836.x.

Morales-Prieto, D. M. *et al.* (2012) 'MicroRNA expression profiles of trophoblastic cells', *Placenta*. Elsevier, 33(9), pp. 725–734. doi: 10.1016/j.placenta.2012.05.009.

Morales, P. *et al.* (2004) 'Extracellular localization of proteasomes in human sperm', *Molecular Reproduction and Development*. John Wiley & Sons, Ltd, 68(1), pp. 115–124. doi: 10.1002/mrd.20052.

Moran, R., Morgan, C. and O'Connell, M. (2015) 'A Guide to Phylogenetic Reconstruction Using Heterogeneous Models—A Case Study from the Root of the Placental Mammal Tree', *Computation*, 3(2), pp. 177–196. doi: 10.3390/computation3020177.

Moran, Y. *et al.* (2017) 'The evolutionary origin of plant and animal microRNAs', *Nature ecology & evolution*, 1, p. 27. doi: 10.1038/s41559-016-0027.

Morgan, C. C. *et al.* (2013) 'Heterogeneous Models Place the Root of the Placental Mammal

Phylogeny', *Molecular Biology and Evolution*, 30(9), pp. 2145–2156. doi: doi:10.1093/molbev/mst117.

Morgan, F. J. and Canfield, R. E. (1971) 'Nature of the Subunits of Human Chorionic Gonadotropin', *Endocrinology*, 88(4), pp. 1045–1053. doi: 10.1210/endo-88-4-1045.

Mossman, H. W. (1987) 'Vertebrate Fetal Membranes', *Rutgers, New Brunswick*, (November), p. 383.

van Mourik, M. S. M., Macklon, N. S. and Heijnen, C. J. (2009) 'Embryonic implantation: cytokines, adhesion molecules, and immune cells in establishing an implantation environment', *Journal of Leukocyte Biology*. John Wiley & Sons, Ltd, 85(1), pp. 4–19. doi: 10.1189/jlb.0708395.

Munaut, C. *et al.* (2016) 'Dysregulated circulating miRNAs in preeclampsia', *Biomedical reports*. 2016/10/14. D.A. Spandidos, 5(6), pp. 686–692. doi: 10.3892/br.2016.779.

Murphy, W. J. *et al.* (2001) 'Molecular phylogenetics and the origins of placental mammals.', *Nature*, 409(February), pp. 614–618. doi: 10.1038/35054550.

Muruganujan, A. *et al.* (2018) 'PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools', *Nucleic Acids Research*, 47(D1), pp. D419–D426. doi: 10.1093/nar/gky1038.

Nairn, A. V and Moremen, K. W. (2014) 'Mannosyl-Oligosaccharide Glucosidase (Glucosidase I, MOGS) BT - Handbook of Glycosyltransferases and Related Genes', in Taniguchi, N. et al. (eds). Tokyo: Springer Japan, pp. 1273–1282. doi: 10.1007/978-4-431-54240-7_10.

NCBI (2017) *NCBI BLAST FAQ*. Available at: https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=FAQ (Accessed: 14 November 2018).

Nei, M., Suzuki, Y. and Nozawa, M. (2010) 'The Neutral Theory of Molecular Evolution in the Genomic Era', *Annual Review of Genomics and Human Genetics*. Annual Reviews, 11(1), pp. 265–289. doi: 10.1146/annurev-genom-082908-150129.

Nei, M. and Tajima, F. (1981) 'Genetic Drift and Estimation of Effective Population Size', *Genetics*, 98(3), pp. 625 LP – 640. Available at: http://www.genetics.org/content/98/3/625.abstract.

Nielsen, R. and Yang, Z. (1998) 'Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene', *Genetics*, 148(3), pp. 929–936. Available at:

https://pubmed.ncbi.nlm.nih.gov/9539414.

Nilsson, D. and Pelger, S. (1994) 'A pessimistic estimate of the time required for an eye to evolve', *Proceedings of the Royal Society of London. Series B: Biological Sciences*. Royal Society, 256(1345), pp. 53–58. doi: 10.1098/rspb.1994.0048.

Nishihara, H. *et al.* (2016) 'Coordinately Co-opted Multiple Transposable Elements Constitute an Enhancer for wnt5a Expression in the Mammalian Secondary Palate', *PLOS Genetics*, 12(10), p. e1006380. doi: 10.1371/journal.pgen.1006380.

Nishihara, H., Okada, N. and Hasegawa, M. (2007) 'Rooting the eutherian tree: the power and pitfalls of phylogenomics.', *Genome biology*, 8(9), p. R199. doi: 10.1186/gb-2007-8-9-r199.

Nishikura, K. (2010) 'Functions and regulation of RNA editing by ADAR deaminases', *Annual review of biochemistry*, 79, pp. 321–349. doi: 10.1146/annurev-biochem-060208-105251.

Niswender, G. D. *et al.* (1994) 'Luteal Function: The Estrous Cycle and Early Pregnancy', *Biology of Reproduction*, 50(2), pp. 239–247. doi: 10.1095/biolreprod50.2.239.

Novakovic, B. *et al.* (2016) 'DNA methylation mediated up-regulation of TERRA non-coding RNA is coincident with elongated telomeres in the human placenta', *Molecular Human Reproduction*, 22(11), pp. 791–799. doi: 10.1093/molehr/gaw053.

O'Brien, J. *et al.* (2018) 'Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation ', *Frontiers in Endocrinology* , p. 402. Available at: https://www.frontiersin.org/article/10.3389/fendo.2018.00402.

O'Connell, M. J. and McInerney, J. O. (2005) 'Gamma Chain Receptor Interleukins: Evidence for Positive Selection Driving the Evolution of Cell-to-Cell Communicators in the Mammalian Immune System', *Journal of Molecular Evolution*, 61(5), pp. 608–619. doi: 10.1007/s00239-004-0313-3.

O'Leary, N. A. *et al.* (2016) 'Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation', *Nucleic acids research*. 2015/11/08. Oxford University Press, 44(D1), pp. D733–D745. doi: 10.1093/nar/gkv1189.

Ohta, T. and Gillespie, J. H. (1996) 'The Development of Neutral and Nearly Neutral Theories', *Theoretical Population Biology*, 49(0007), pp. 128–142. Available at: http://ac.els-cdn.com/S0040580996900076/1-s2.0-S0040580996900076-main.pdf?_tid=db129526-08e7-11e7-a54c-00000aab0f6b&acdnat=1489517936_86d51a61f7f0b8d31ee3d70efd4f837c.

Okada, H., Tsuzuki, T. and Murata, H. (2018) 'Decidualization of the human endometrium', *Reproductive medicine and biology*. John Wiley and Sons Inc., 17(3), pp. 220–227. doi: 10.1002/rmb2.12088.

Okumu, L. A. *et al.* (2010) 'The effect of elevated progesterone and pregnancy status on mRNA expression and localisation of progesterone and oestrogen receptors in the bovine uterus', *REPRODUCTION*. Bristol, UK: BioScientifica, 140(1), pp. 143–153. doi: 10.1530/REP-10-0113.

Olena, A. F. and Patton, J. G. (2010) 'Genomic Organization of microRNAs', *Journal of cellular physiology*, 222(3), pp. 540–545. doi: 10.1002/jcp.21993.

Oliveira, A. C. *et al.* (2017) 'Combining Results from Distinct MicroRNA Target Prediction Tools Enhances the Performance of Analyses ', *Frontiers in Genetics* , p. 59. Available at: https://www.frontiersin.org/article/10.3389/fgene.2017.00059.

Oliveira, A. C. *et al.* (2019) 'Understanding the Modus Operandi of MicroRNA Regulatory Clusters', *Cells*. MDPI, 8(9), p. 1103. doi: 10.3390/cells8091103.

Orr, D. (2018) 'Post Reader'.

Paiva, P. *et al.* (2011) 'Human chorionic gonadotrophin regulates FGF2 and other cytokines produced by human endometrial epithelial cells, providing a mechanism for enhancing endometrial receptivity', *Human Reproduction*, 26(5), pp. 1153–1162. doi: 10.1093/humrep/der027.

Pascoal, C. *et al.* (2020) 'CDG and immune response: From bedside to bench and back', *Journal of Inherited Metabolic Disease*. John Wiley & Sons, Ltd, 43(1), pp. 90–124. doi: 10.1002/jimd.12126.

Pasquinelli, A. E. *et al.* (2000) 'Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA', *Nature*, 408(6808), pp. 86–89. doi: 10.1038/35040556.

Pasten, C., Morales, P. and Kong, M. (2005) 'Role of the sperm proteasome during fertilization and gamete interaction in the mouse', *Molecular Reproduction and Development*. John Wiley & Sons, Ltd, 71(2), pp. 209–219. doi: 10.1002/mrd.20280.

Pearson, W. R. (2013) 'An introduction to sequence similarity ("homology") searching', *Current protocols in bioinformatics*, Chapter 3, p. Unit3.1-Unit3.1. doi: 10.1002/0471250953.bi0301s42.

Peñalva-Arana, D. C., Lynch, M. and Robertson, H. M. (2009) 'The chemoreceptor genes of the waterflea Daphnia pulex: many Grs but no Ors', *BMC Evolutionary Biology*, 9(1), p. 79. doi:

10.1186/1471-2148-9-79.

Phillippe, M., Sawyer, M. R. and Edelson, P. K. (2019) 'The telomere gestational clock: increasing short telomeres at term in the mouse', *American Journal of Obstetrics & Gynecology*. Elsevier, 220(5), pp. 496.e1-496.e8. doi: 10.1016/j.ajog.2019.01.218.

Piehler, J. *et al.* (2012) 'Structural and dynamic determinants of type I interferon receptor assembly and their functional interpretation', *Immunological reviews*, 250(1), pp. 317–334. doi: 10.1111/imr.12001.

Pineles, B. L. *et al.* (2007) 'Distinct subsets of microRNAs are expressed differentially in the human placentas of patients with preeclampsia', *American Journal of Obstetrics and Gynecology*, 196(3), pp. 261.e1-261.e6. doi: https://doi.org/10.1016/j.ajog.2007.01.008.

Platanias, L. C. (2005) 'Mechanisms of type-I- and type-II-interferon-mediated signalling', *Nature Reviews Immunology*, 5(5), pp. 375–386. doi: 10.1038/nri1604.

Pobezinskaya, Y. L. and Liu, Z. (2012) 'The role of TRADD in death receptor signaling', *Cell cycle (Georgetown, Tex.)*. 2012/03/01. Landes Bioscience, 11(5), pp. 871–876. doi: 10.4161/cc.11.5.19300.

Pohler, K. G. *et al.* (2017) 'Circulating microRNA as candidates for early embryonic viability in cattle', *Molecular reproduction and development*, 84(8), pp. 731–743. doi: 10.1002/mrd.22856.

Pond, S. L. K., Frost, S. D. W. and Muse, S. V (2004) 'HyPhy: hypothesis testing using phylogenies', *Bioinformatics*, 21(5), pp. 676–679. doi: 10.1093/bioinformatics/bti079.

Ponsuksili, S. *et al.* (2014) 'Differential Expression of miRNAs and Their Target mRNAs in Endometria Prior to Maternal Recognition of Pregnancy Associates with Endometrial Receptivity for In Vivo- and In Vitro-Produced Bovine Embryos1', *Biology of Reproduction*, 91(6). doi: 10.1095/biolreprod.114.121392.

Pratt, A. J. and MacRae, I. J. (2009) 'The RNA-induced Silencing Complex: A Versatile Gene-silencing Machine', *The Journal of Biological Chemistry*. 9650 Rockville Pike, Bethesda, MD 20814, U.S.A.: American Society for Biochemistry and Molecular Biology, 284(27), pp. 17897–17901. doi: 10.1074/jbc.R900012200.

Prochnik, S. E., Rokhsar, D. S. and Aboobaker, A. A. (2007) 'Evidence for a microRNA expansion in the bilaterian ancestor', *Development Genes and Evolution*, 217(1), pp. 73–77. doi: 10.1007/s00427-006-0116-1.

Quah, S., Hui, J. H. L. and Holland, P. W. H. (2015) 'A Burst of miRNA Innovation in the Early Evolution of Butterflies and Moths', *Molecular biology and evolution*. 2015/01/08. Oxford University Press, 32(5), pp. 1161–1174. doi: 10.1093/molbev/msv004.

Radi, Z. A., Marusak, R. A. and Morris, D. L. (2009) 'Species Comparison of the Role of p38 MAP Kinase in the Female Reproductive System', *Journal of toxicologic pathology*. 2009/07/07. The Japanese Society of Toxicologic Pathology, 22(2), pp. 109–124. doi: 10.1293/tox.22.109.

Raheem, K. A. (2017) 'An insight into maternal recognition of pregnancy in mammalian species', *Journal of the Saudi Society of Agricultural Sciences*, 16(1), pp. 1–6. doi: https://doi.org/10.1016/j.jssas.2015.01.002.

Rahman, M. L. *et al.* (2018) 'Regulation of birthweight by placenta-derived miRNAs: evidence from an arsenic-exposed birth cohort in Bangladesh', *Epigenetics*. Taylor & Francis, 13(6), pp. 573–590. doi: 10.1080/15592294.2018.1481704.

Rasier, G. *et al.* (2006) 'Female sexual maturation and reproduction after prepubertal exposure to estrogens and endocrine disrupting chemicals: A review of rodent and human data', *Molecular and Cellular Endocrinology*, 254–255, pp. 187–201. doi: https://doi.org/10.1016/j.mce.2006.04.002.

Rasweiler, J. J. and Badwaik, N. K. (2000) '5 - Anatomy and Physiology of the Female Reproductive Tract', in Crichton, E. G. and Krutzsch, P. H. B. T.-R. B. of B. (eds) *Reproductive Biology of Bats*. London: Academic Press, pp. 157–219. doi: https://doi.org/10.1016/B978-012195670-7/50006-0.

Revel, A. *et al.* (2011) 'MicroRNAs are associated with human embryo implantation defects', *Human Reproduction*, 26(10), pp. 2830–2840. doi: 10.1093/humrep/der255.

Roberts, R. M., Green, J. A. and Schulz, L. C. (2016) 'The evolution of the placenta', *Reproduction (Cambridge, England)*. 2016/08/02, 152(5), pp. R179–R189. doi: 10.1530/REP-16-0325.

Robertson, H. M., Warr, C. G. and Carlson, J. R. (2003) 'Molecular evolution of the insect chemoreceptor gene superfamily in Drosophila melanogaster', *Proceedings of the National Academy of Sciences*, 100(suppl 2), pp. 14537 LP – 14542. doi: 10.1073/pnas.2335847100.

Romiguier, J. *et al.* (2010) 'Contrasting GC-content dynamics across 33 mammalian genomes: relationship with life-history traits and chromosome sizes', *Genome research*. 2010/06/07. Cold

Spring Harbor Laboratory Press, 20(8), pp. 1001–1009. doi: 10.1101/gr.104372.109.

Rote, N. S., Chakrabarti, S. and Stetzer, B. P. (2004) 'The role of human endogenous retroviruses in trophoblast differentiation and placental development.', *Placenta*, 25(8–9), pp. 673–83. doi: 10.1016/j.placenta.2004.02.008.

Sadat, M. A. *et al.* (2014) 'Glycosylation, Hypogammaglobulinemia, and Resistance to Viral Infections', *New England Journal of Medicine*. Massachusetts Medical Society, 370(17), pp. 1615–1625. doi: 10.1056/NEJMoa1302846.

Sætrom, P. *et al.* (2007) 'Distance constraints between microRNA target sites dictate efficacy and cooperativity', *Nucleic Acids Research*, 35(7), pp. 2333–2342. doi: 10.1093/nar/gkm133.

Sanetra, M. *et al.* (2005) 'Conservation and co-option in developmental programmes: the importance of homology relationships', *Frontiers in Zoology*, 2(1), p. 15. doi: 10.1186/1742-9994-2-15.

Santamaria, X. and Taylor, H. (2014) 'MicroRNA and gynecological reproductive diseases', *Fertility and Sterility*. Elsevier, 101(6), pp. 1545–1551. doi: 10.1016/j.fertnstert.2014.04.044.

Sartori, R. *et al.* (2004) 'Comparison of Ovarian Function and Circulating Steroids in Estrous Cycles of Holstein Heifers and Lactating Cows', *Journal of Dairy Science*. Elsevier, 87(4), pp. 905–920. doi: 10.3168/jds.S0022-0302(04)73235-X.

Sato, J., Nasu, M. and Tsuchitani, M. (2016) 'Comparative histopathology of the estrous or menstrual cycle in laboratory animals', *Journal of toxicologic pathology*. 2016/05/16. Japanese Society of Toxicologic Pathology, 29(3), pp. 155–162. doi: 10.1293/tox.2016-0021.

Schlafke, S. and Enders, A. C. (1975) 'Cellular Basis of Interaction Between Trophoblast and Uterus at Implantation', *Biology of Reproduction*, 12(1), pp. 41–65. doi: 10.1095/biolreprod12.1.41.

Schneider, A. *et al.* (2009) 'Estimates of Positive Darwinian Selection Are Inflated by Errors in Sequencing, Annotation, and Alignment', *Genome Biology and Evolution*, 1, pp. 114–118. doi: 10.1093/gbe/evp012.

Schneider, W. *et al.* (1991) 'Murine progesterone receptor exists predominantly as the 83-kilodalton "A" form', *The Journal of Steroid Biochemistry and Molecular Biology*, 38(3), pp. 285–291. doi: https://doi.org/10.1016/0960-0760(91)90099-Q.

Schroeder, M. *et al.* (2018) 'Placental miR-340 mediates vulnerability to activity based anorexia

in mice', *Nature Communications*, 9(1), p. 1596. doi: 10.1038/s41467-018-03836-2.

Seino, S., Bell, G. I. and Li, W. H. (1992) 'Sequences of primate insulin genes support the hypothesis of a slower rate of molecular evolution in humans and apes than in monkeys.', *Molecular Biology and Evolution*, 9(2), pp. 193–203. doi: 10.1093/oxfordjournals.molbev.a040713.

Seluanov, A. *et al.* (2007) 'Telomerase activity coevolves with body mass not lifespan', *Aging cell*. 2006/12/14, 6(1), pp. 45–52. doi: 10.1111/j.1474-9726.2006.00262.x.

Sempere, L. F. *et al.* (2006) 'The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint.', *Journal of experimental zoology. Part B, Molecular and developmental evolution*. United States, 306(6), pp. 575–588. doi: 10.1002/jez.b.21118.

Seshagiri, P. B. *et al.* (2009) 'Cellular and molecular regulation of mammalian blastocyst hatching', *Journal of Reproductive Immunology*, 83(1), pp. 79–84. doi: https://doi.org/10.1016/j.jri.2009.06.264.

Seya, T. *et al.* (2012) 'TLR3/TICAM-1 signaling in tumor cell RIP3-dependent necroptosis', *OncoImmunology*. Taylor & Francis, 1(6), pp. 917–923. doi: 10.4161/onci.21244.

Sheikh, A. A., Hooda, O. K. and Dang, A. K. (2018) 'JAK3 and PI3K mediate bovine Interferon-tau stimulated gene expression in the blood neutrophils', *Journal of Cellular Physiology*. John Wiley & Sons, Ltd, 233(6), pp. 4885–4894. doi: 10.1002/jcp.26296.

Shi, Q. J. *et al.* (1993) 'Novel role of human chorionic gonadotropin in differentiation of human cytotrophoblasts', *Endocrinology*, 132(3), pp. 1387–1395. doi: 10.1210/en.132.3.1387.

Shivdasani, R. (2006) 'MicroRNAs: regulators of gene expression and cell differentiation', *Blood*, 108(12), pp. 3646–3653. doi: 10.1182/blood-2006-01-030015.

Shubin, N., Tabin, C. and Carroll, S. (2009) 'Deep homology and the origins of evolutionary novelty', *Nature*, 457(7231), pp. 818–823. doi: 10.1038/nature07891.

Sibley, C. P. *et al.* (2004) 'Placental-specific insulin-like growth factor 2 (Igf2) regulates the diffusional exchange characteristics of the mouse placenta', *Proceedings of the National Academy of Sciences of the United States of America*. 2004/05/18. National Academy of Sciences, 101(21), pp. 8204–8208. doi: 10.1073/pnas.0402508101.

Silverthorn, D. U. (2010) *Human Physiology: An Integrated Approach*. 5th Editio. Edited by D. Espinoza. San Francisco: Pearson, Benjamin Cummings.

Sironi, M. *et al.* (2015) 'Evolutionary insights into host–pathogen interactions from mammalian sequence data', *Nature Reviews Genetics*, 16(4), pp. 224–236. doi: 10.1038/nrg3905.

Skarzynski, D. *et al.* (2013) 'Growth and regression in bovine corpora lutea: Regulation by local survival and death pathways', *Reproduction in Domestic Animals*, 48(SUPPL.1), pp. 25–37. doi: 10.1111/rda.12203.

Smith, M. D. *et al.* (2015) 'Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection', *Molecular biology and evolution*. 2015/02/19. Oxford University Press, 32(5), pp. 1342–1353. doi: 10.1093/molbev/msv022.

Smith, N. G. C. and Eyre-Walker, A. (2002) 'Adaptive protein evolution in Drosophila', *Nature*, 415(6875), pp. 1022–1024. doi: 10.1038/4151022a.

Sohr, S. and Engeland, K. (2011) 'The tumor suppressor p53 induces expression of the pregnancy-supporting human chorionic gonadotropin (hCG) CGB7 gene', *Cell cycle (Georgetown, Tex.)*. 2011/11/01. Landes Bioscience, 10(21), pp. 3758–3767. doi: 10.4161/cc.10.21.17946.

Song, G. and Wang, L. (2008) 'MiR-433 and miR-127 Arise from Independent Overlapping Primary Transcripts Encoded by the miR-433-127 Locus', *PLOS ONE*. Public Library of Science, 3(10), p. e3574. Available at: https://doi.org/10.1371/journal.pone.0003574.

Soumillon, M. *et al.* (2013) 'Birth and expression evolution of mammalian microRNA genes', *Genome Research*, pp. 34–45. doi: 10.1101/gr.140269.112.Freely.

Spencer, T. E. and Bazer, F. W. (1995) 'Temporal and Spatial Alterations in Uterine Estrogen Receptor and Progesterone Receptor Gene Expression During the Estrous Cycle and Early Pregnancy in the Ewe1', *Biology of Reproduction*, 53(6), pp. 1527–1543. doi: 10.1095/biolreprod53.6.1527.

Stadler, H. S., Higgins, K. M. and Capecchi, M. R. (2001) 'Loss of Eph-receptor expression correlates with loss of cell adhesion and chondrogenic capacity in Hoxa13 mutant limbs', *Development (Cambridge, England)*. Howard Hughes Medical Institute, University of Utah School of Medicine, Salt Lake City, UT 84112-5331, USA., 128(21), pp. 4177–4188. Available at: http://europepmc.org/abstract/MED/11684655.

Starega-Roslan, J. *et al.* (2011) 'Structural basis of microRNA length variety', *Nucleic Acids Research*. Oxford University Press, 39(1), pp. 257–268. doi: 10.1093/nar/gkq727.

Starokadomskyy, P. *et al.* (2013) 'CCDC22 deficiency in humans blunts activation of proinflammatory NF-κB signaling', *The Journal of Clinical Investigation*. The American Society for Clinical Investigation, 123(5), pp. 2244–2256. doi: 10.1172/JCI66466.

Stocco, C., Telleria, C. and Gibori, G. (2007) 'The Molecular Control of Corpus Luteum Formation, Function, and Regression', *Endocrine Reviews*, 28(1), pp. 117–149. doi: 10.1210/er.2006-0022.

Su, R.-W. and Fazleabas, A. T. (2015) 'Implantation and Establishment of Pregnancy in Human and Nonhuman Primates', *Advances in anatomy, embryology, and cell biology*, 216, pp. 189–213. doi: 10.1007/978-3-319-15856-3_10.

Summers, K. and Crespi, B. (2005) 'Cadherins in maternal-foetal interactions: red queen with a green beard?', *Proceedings. Biological sciences*. The Royal Society, 272(1563), pp. 643–649. doi: 10.1098/rspb.2004.2890.

Szklarczyk, D. *et al.* (2019) 'STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets', *Nucleic acids research*. Oxford University Press, 47(D1), pp. D607–D613. doi: 10.1093/nar/gky1131.

Tajima, F. (1989) 'Statistical method for testing the neutral mutation hypothesis by DNA polymorphism', *Genetics*, 123(3), pp. 585–595. Available at: https://pubmed.ncbi.nlm.nih.gov/2513255.

Tang, Q. *et al.* (2013) 'miR-141 Contributes to Fetal Growth Restriction by Regulating PLAG1 Expression', *PLOS ONE*. Public Library of Science, 8(3), p. e58737. Available at: https://doi.org/10.1371/journal.pone.0058737.

Tanzer, A. and Stadler, P. F. (2006) 'Evolution of MicroRNAs', in Ying, S.-Y. (ed.) *MicroRNA Protocols*. Totowa, NJ: Humana Press, pp. 335–350. doi: 10.1385/1-59745-123-1:335.

Tarver, J. E. *et al.* (2016) 'The Interrelationships of Placental Mammals and the Limits of Phylogenetic Inference.', *Genome biology and evolution*, 8(2), pp. 330–344. doi: 10.1093/gbe/evv261.

Tarver, J. E. *et al.* (2018) 'Well-Annotated microRNAomes Do Not Evidence Pervasive miRNA Loss', *Genome Biology and Evolution*, 10(6), pp. 1457–1470. doi: 10.1093/gbe/evy096.

Tarver, J. E., Donoghue, P. C. J. and Peterson, K. J. (2012) 'Do miRNAs have a deep evolutionary history?', *BioEssays*. John Wiley & Sons, Ltd, 34(10), pp. 857–866. doi: 10.1002/bies.201200055.

Tavaniotou, A. *et al.* (2002) 'Impact of ovarian stimulation on corpus luteum function and embryonic implantation', *Journal of Reproductive Immunology*, 55(1), pp. 123–130. doi: https://doi.org/10.1016/S0165-0378(01)00134-6.

Taylor, H. S., Vanden Heuvel, G. B. and Igarashi, P. (1997) 'A Conserved Hox Axis in the Mouse and Human Female Reproductive System: Late Establishment and Persistent Adult Expression of the Hoxa Cluster Genes', *Biology of Reproduction*, 57(6), pp. 1338–1345. doi: 10.1095/biolreprod57.6.1338.

Teeling, E. C. and Hedges, S. B. (2013) 'Making the impossible possible: Rooting the tree of placental mammals', *Molecular Biology and Evolution*, 30(9), pp. 1999–2000. doi: 10.1093/molbev/mst118.

Tepekoy, F., Akkoyunlu, G. and Demir, R. (2015) 'The role of Wnt signaling members in the uterus and embryo during pre-implantation and implantation', *Journal of assisted reproduction and genetics*. 2014/12/24. Springer US, 32(3), pp. 337–346. doi: 10.1007/s10815-014-0409-7.

Terenina, E. *et al.* (2016) 'Differentially expressed genes and gene networks involved in pig ovarian follicular atresia', *Physiological Genomics*. American Physiological Society, 49(2), pp. 67–80. doi: 10.1152/physiolgenomics.00069.2016.

Teruyama, K. *et al.* (2001) 'Neurophilin-1 is a downstream target of transcription factor Ets-1 in human umbilical vein endothelial cells', *FEBS Letters*, 504(1), pp. 1–4. doi: https://doi.org/10.1016/S0014-5793(01)02724-7.

Than, N. G. *et al.* (2009) 'A primate subfamily of galectins expressed at the maternal–fetal interface that promote immune cell death', *Proceedings of the National Academy of Sciences*, 106(24), pp. 9731 LP – 9736. doi: 10.1073/pnas.0903568106.

Thompson, J. D. *et al.* (2001) 'Towards a reliable objective function for multiple sequence alignments', *Journal of Molecular Biology*, 314(4), pp. 937–951. doi: 10.1006/jmbi.2001.5187.

Thomson, R. C. *et al.* (2014) 'A critical appraisal of the use of microRNA data in phylogenetics', *Proceedings of the National Academy of Sciences of the United States of America*. 2014/07/28. National Academy of Sciences, 111(35), pp. E3659–E3668. doi: 10.1073/pnas.1407207111.

Tinning, H. *et al.* (2020) 'The role of CAPG in molecular communication between the embryo and the uterine endometrium: Is its function conserved in species with different implantation strategies?', *bioRxiv*.

Tochigi, H. *et al.* (2017) 'Loss of miR-542-3p enhances IGFBP-1 expression in decidualizing human endometrial stromal cells', *Scientific Reports*. The Author(s), 7, p. 40001. Available at: https://doi.org/10.1038/srep40001.

True, J. R. and Carroll, S. B. (2002) 'Gene Co-Option in Physiological and Morphological Evolution', *Annual Review of Cell and Developmental Biology*. Annual Reviews, 18(1), pp. 53–80. doi: 10.1146/annurev.cellbio.18.020402.140619.

Turner, M. J., Jiao, A. L. and Slack, F. J. (2014) 'Autoregulation of lin-4 microRNA transcription by RNA activation (RNAa) in C. elegans', *Cell cycle (Georgetown, Tex.)*. 2014/01/07. Landes Bioscience, 13(5), pp. 772–781. doi: 10.4161/cc.27679.

Upham, N. S., Esselstyn, J. A. and Jetz, W. (2019) 'Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation', *PLOS Biology*. Public Library of Science, 17(12), p. e3000494. Available at: https://doi.org/10.1371/journal.pbio.3000494.

Válóczi, A. *et al.* (2004) 'Sensitive and specific detection of microRNAs by northern blot analysis using LNA-modified oligonucleotide probes', *Nucleic acids research*. Oxford University Press, 32(22), pp. e175–e175. doi: 10.1093/nar/gnh171.

Vanhaesebroeck, B., Stephens, L. and Hawkins, P. (2012) 'PI3K signalling: the path to discovery and understanding', *Nature Reviews Molecular Cell Biology*, 13(3), pp. 195–203. doi: 10.1038/nrm3290.

Vasudevan, S. (2012) 'Posttranscriptional Upregulation by MicroRNAs', *WIREs RNA*. John Wiley & Sons, Ltd, 3(3), pp. 311–330. doi: 10.1002/wrna.121.

Vesely, C. *et al.* (2012) 'Adenosine deaminases that act on RNA induce reproducible changes in abundance and sequence of embryonic miRNAs', *Genome research*. 2012/02/06. Cold Spring Harbor Laboratory Press, 22(8), pp. 1468–1476. doi: 10.1101/gr.133025.111.

Vicens, A. and Treviño, C. L. (2018) 'Positive Selection in the Evolution of Mammalian CRISPs', *Journal of Molecular Evolution*, 86(9), pp. 635–645. doi: 10.1007/s00239-018-9872-6.

Wagner, G. P. (2015) 'Evolutionary innovations and novelties: Let us get down to business!', *Zoologischer Anzeiger*, 256(October), pp. 75–81. doi: 10.1016/j.jcz.2015.04.006.

Wagner, G. P. and Lynch, V. J. (2010) 'Evolutionary Novelties', *Current Biology*, 20(2), pp. R48–R52. doi: http://dx.doi.org/10.1016/j.cub.2009.11.010.

Wahid, F. *et al.* (2010) 'MicroRNAs: Synthesis, mechanism, function, and recent clinical trials', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1803(11), pp. 1231–1243. doi: https://doi.org/10.1016/j.bbamcr.2010.06.013.

Walsh, T. A. (2013) *The evolution of the mammal placenta - a computational approach to the identification and analysis of placenta-specific genes and microRNAs*.

Wang, B. *et al.* (2008) 'High Expression of l-Selectin Ligand in Secretory Endometrium is Associated with Better Endometrial Receptivity and Facilitates Embryo Implantation in Human Being', *American Journal of Reproductive Immunology*. John Wiley & Sons, Ltd, 60(2), pp. 127–134. doi: 10.1111/j.1600-0897.2008.00604.x.

Wang, H. *et al.* (1998) 'Progesterone receptor subtype B is differentially regulated in human endometrial stroma', *Molecular Human Reproduction*, 4(4), pp. 407–412. doi: 10.1093/molehr/4.4.407.

Wang, Y. *et al.* (2015) 'MicroRNAome in decidua: a new approach to assess the maintenance of pregnancy', *Fertility and Sterility*. Elsevier, 103(4), pp. 980-989.e6. doi: 10.1016/j.fertnstert.2015.01.003.

Webb, A. E. *et al.* (2015) 'Adaptive evolution as a predictor of species-specific innate immune response', *Molecular Biology and Evolution*, 32(7), pp. 1717–1729. doi: 10.1093/molbev/msv051.

Webb, A. E., Walsh, T. A. and O'Connell, M. J. (2016a) 'VESPA: Very large-scale Evolutionary and Selective Pressure Analyses', *PeerJ Preprints*, 4, p. e1895v1. doi: 10.7287/PEERJ.PREPRINTS.1895V1.

Webb, A. E., Walsh, T. A. and O'Connell, M. J. (2016b) 'VESPA: Very large-scale Evolutionary and Selective Pressure Analyses', *PeerJ Preprints*, 4, p. e1895v1. doi: 10.7287/PEERJ.PREPRINTS.1895V1.

Welch, J. J. (2006) 'Estimating the Genomewide Rate of Adaptive Protein Evolution in Drosophila', *Genetics*, 173(2), pp. 821 LP – 837. doi: 10.1534/genetics.106.056911.

Welch, J. J., Bininda-Emonds, O. R. P. and Bromham, L. (2008) 'Correlates of substitution rate variation in mammalian protein-coding sequences', *BMC evolutionary biology*. BioMed Central, 8, p. 53. doi: 10.1186/1471-2148-8-53.

Wetzel-Strong, S. E. *et al.* (2016) 'Cohort of estrogen-induced microRNAs regulate

adrenomedullin expression', *American journal of physiology. Regulatory, integrative and comparative physiology*. 2015/11/18. American Physiological Society, 310(2), pp. R209–R216. doi: 10.1152/ajpregu.00305.2014.

Wilkinson, K. A. and Henley, J. M. (2010) 'Mechanisms, regulation and consequences of protein SUMOylation', *The Biochemical journal*, 428(2), pp. 133–145. doi: 10.1042/BJ20100158.

Wiltbank, M. C. *et al.* (2012) 'Comparison of endocrine and cellular mechanisms regulating the corpus luteum of primates and ruminants', *Animal reproduction*, 9(3), pp. 242–259. Available at: https://pubmed.ncbi.nlm.nih.gov/23750179.

Winter, J. *et al.* (2009) 'Many roads to maturity: microRNA biogenesis pathways and their regulation', *Nature Cell Biology*, 11(3), pp. 228–234. doi: 10.1038/ncb0309-228.

Witkos, T. M., Koscianska, E. and Krzyzosiak, W. J. (2011) 'Practical Aspects of microRNA Target Prediction', *Current Molecular Medicine*. Bentham Science Publishers Ltd, 11(2), pp. 93–109. doi: 10.2174/156652411794859250.

Wong, W. S. W. and Nielsen, R. (2004) 'Detecting selection in noncoding regions of nucleotide sequences', *Genetics*, 167(2), pp. 949–958. doi: 10.1534/genetics.102.010959.

Woods, L., Perez-Garcia, V. and Hemberger, M. (2018) 'Regulation of Placental Development and Its Impact on Fetal Growth—New Insights From Mouse Models ', *Frontiers in Endocrinology* , p. 570. Available at: https://www.frontiersin.org/article/10.3389/fendo.2018.00570.

Wu, C. I. and Li, W. H. (1985) 'Evidence for higher rates of nucleotide substitution in rodents than in man', *Proceedings of the National Academy of Sciences of the United States of America*, 82(6), pp. 1741–1745. doi: 10.1073/pnas.82.6.1741.

Wu, S. *et al.* (2010) 'Multiple microRNAs modulate p21Cip1/Waf1 expression by directly targeting its 3' untranslated region', *Oncogene*, 29(15), pp. 2302–2308. doi: 10.1038/onc.2010.34.

Xiao, C. *et al.* (2003) 'Ecsit is required for Bmp signaling and mesoderm formation during mouse embryogenesis', *Genes & Development* , 17(23), pp. 2933–2949. doi: 10.1101/gad.1145603.

Xiao, C. *et al.* (2007) 'MiR-150 Controls B Cell Differentiation by Targeting the Transcription Factor c-Myb', *Cell*. Elsevier, 131(1), pp. 146–159. doi: 10.1016/j.cell.2007.07.021.

Xin, H., Liu, D. and Songyang, Z. (2008) 'The telosome/shelterin complex and its functions',

Genome Biology, 9(9), p. 232. doi: 10.1186/gb-2008-9-9-232.

Yang, C. H. *et al.* (2015) 'The Type I IFN-Induced miRNA, miR-21', *Pharmaceuticals (Basel, Switzerland)*. MDPI, 8(4), pp. 836–847. doi: 10.3390/ph8040836.

Yang, H. *et al.* (2019) 'Differential expression of microRNA-411 and 376c is associated with hypertension in pregnancy ', *Brazilian Journal of Medical and Biological Research* . scielo .

Yang, Q. *et al.* (2018) 'Association of the peripheral blood levels of circulating microRNAs with both recurrent miscarriage and the outcomes of embryo transfer in an in vitro fertilization process', *Journal of translational medicine*. BioMed Central, 16(1), p. 186. doi: 10.1186/s12967-018-1556-x.

Yang, W. *et al.* (2006) 'Modulation of microRNA processing and expression through RNA editing by ADAR deaminases', *Nature structural & molecular biology*. 2005/12/20, 13(1), pp. 13–21. doi: 10.1038/nsmb1041.

Yang, Y. *et al.* (2017) 'Protein SUMOylation modification and its associations with disease', *Open biology*. The Royal Society, 7(10), p. 170167. doi: 10.1098/rsob.170167.

Yang, Z. (1997) 'PAML: a program package for phylogenetic analysis by maximum likelihood', *Computer Applications in BioSciences*, 13, pp. 555–556.

Yang, Z. *et al.* (2000) 'Codon-substitution models for heterogeneous selection pressure at amino acid sites', *Genetics*, 155(1), pp. 431–449. Available at: https://pubmed.ncbi.nlm.nih.gov/10790415.

Yang, Z. (2007) 'PAML 4: Phylogenetic analysis by maximum likelihood', *Molecular Biology and Evolution*, 24(8), pp. 1586–1591. doi: 10.1093/molbev/msm088.

Yang, Z. and Nielsen, R. (2002) 'Codon-Substitution Models for Detecting Molecular Adaptation at Individual Sites Along Specific Lineages', *Molecular Biology and Evolution*, 19(6), pp. 908–917. doi: 10.1093/oxfordjournals.molbev.a004148.

Yang, Z. and dos Reis, M. (2010) 'Statistical Properties of the Branch-Site Test of Positive Selection', *Molecular Biology and Evolution*, 28(3), pp. 1217–1228. doi: 10.1093/molbev/msq303.

Yang, Z., Wong, W. S. W. and Nielsen, R. (2005) 'Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection', *Molecular Biology and Evolution*, 22(4), pp. 1107–1118. doi: 10.1093/molbev/msi097.

Yates, A. *et al.* (2016) 'Ensembl 2016', *Nucleic Acids Research*, 44(D1), pp. D710–D716. doi: 10.1093/nar/gkv1157.

Ye, J. *et al.* (2012) 'Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction', *BMC bioinformatics*. BioMed Central, 13, p. 134. doi: 10.1186/1471-2105-13-134.

Ye, S. *et al.* (2014) 'Bioinformatics Method to Predict Two Regulation Mechanism: TF–miRNA–mRNA and lncRNA–miRNA–mRNA in Pancreatic Cancer', *Cell Biochemistry and Biophysics*, 70(3), pp. 1849–1858. doi: 10.1007/s12013-014-0142-y.

Yi, Y. J. *et al.* (2007) 'Mechanism of sperm-zona pellucida penetration during mammalian fertilization: 26S proteasome as a candidate egg coat lysin.', *Society of Reproduction and Fertility supplement*, 63, pp. 385–408. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-34547850913&partnerID=40&md5=2f19f1533c220232982bcb44c995d69b.

Young, S. L. (2013) 'Oestrogen and progesterone action on endometrium: a translational approach to understanding endometrial receptivity', *Reproductive biomedicine online*. 2013/06/25, 27(5), pp. 497–505. doi: 10.1016/j.rbmo.2013.06.010.

Yu, J. *et al.* (2008) 'MicroRNA-184 antagonizes microRNA-205 to maintain SHIP2 levels in epithelia', *Proceedings of the National Academy of Sciences of the United States of America*. 2008/11/25. National Academy of Sciences, 105(49), pp. 19300–19305. doi: 10.1073/pnas.0803992105.

Yuan, Z. *et al.* (2010) 'Origin and evolution of a placental-specific microRNA family in the human genome', *BMC Evolutionary Biology*, 10(1), p. 346. doi: 10.1186/1471-2148-10-346.

Zeng, Y., Yi, R. and Cullen, B. R. (2003) 'MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms', *Proceedings of the National Academy of Sciences*, 100(17), pp. 9779 LP – 9784. Available at: http://www.pnas.org/content/100/17/9779.abstract.

Zerbino, D. R. *et al.* (2017) 'Ensembl 2018', *Nucleic Acids Research*, 46(D1), pp. D754–D761. doi: 10.1093/nar/gkx1098.

Zhai, W. *et al.* (2012) 'Looking for Darwin in Genomic Sequences—Validity and Success of Statistical Methods', *Molecular Biology and Evolution*, 29(10), pp. 2889–2893. doi: 10.1093/molbev/mss104.

Zhang, B., Wang, Q. and Pan, X. (2006) 'MicroRNAs and their regulatory roles in animals and

plants', *Journal of Cellular Physiology*. Wiley-Blackwell, 210(2), pp. 279–289. doi: 10.1002/jcp.20869.

Zhang, J. (2004) 'Frequent False Detection of Positive Selection by the Likelihood Method with Branch-Site Models', *Molecular Biology and Evolution*, 21(7), pp. 1332–1339. doi: 10.1093/molbev/msh117.

Zhang, J., Nielsen, R. and Yang, Z. (2005) 'Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level', *Molecular Biology and Evolution*, 22(12), pp. 2472–2479. doi: 10.1093/molbev/msi237.

Zhang, Y. *et al.* (2012) 'Protective role of estrogen-induced miRNA-29 expression in carbon tetrachloride-induced mouse liver injury', *The Journal of biological chemistry*. 2012/03/05. American Society for Biochemistry and Molecular Biology, 287(18), pp. 14851–14862. doi: 10.1074/jbc.M111.314922.

Zhang, Yiwen *et al.* (2018) 'A Cellular MicroRNA Facilitates Regulatory T Lymphocyte Development by Targeting the FOXP3 Promoter TATA-Box Motif', *The Journal of Immunology*, 200(3), pp. 1053 LP – 1063. doi: 10.4049/jimmunol.1700196.

Zhou, X.-G. *et al.* (2018) 'Identifying miRNA and gene modules of colon cancer associated with pathological stage by weighted gene co-expression network analysis', *OncoTargets and therapy*. Dove Medical Press, 11, pp. 2815–2830. doi: 10.2147/OTT.S163891.

Zhou, X. *et al.* (2012) 'Phylogenetic and transcriptomic analysis of chemosensory receptors in a pair of divergent ant species reveals sex-specific signatures of odor coding', *PLoS genetics*. 2012/08/30. Public Library of Science, 8(8), pp. e1002930–e1002930. doi: 10.1371/journal.pgen.1002930.

Zhou, Y. *et al.* (1997) 'Human cytotrophoblasts adopt a vascular phenotype as they differentiate. A strategy for successful endovascular invasion?', *The Journal of clinical investigation*. American Society for Clinical Investigation, 99(9), pp. 2139–2151. doi: 10.1172/JCI119387.

de Ziegler, D. *et al.* (2007) 'Roles of FSH and LH during the follicular phase: Insight into natural cycle IVF', *Reproductive BioMedicine Online*, 15(5), pp. 507–513. doi: 10.1016/S1472-6483(10)60381-1.

Supplementary Table 1

Gene Families	lnL	Parameter Estimates	Positively Selected Sites
PTHR10088	-8851.9075	p2=0.03498 w2=999.00000	Alignment (12 BEB sites): 66 76 77 82 85 88 89 90 91 94 337 560
PTHR10237	-9642.8479	p2=0.01770 w2=8.91623	Alignment (7 BEB sites): 38 63 191 401 417 452 499
PTHR10289	-6803.7805	p2=0.02454 w2=48.06706	Alignment (3 BEB sites): 43 181 415
PTHR10458	-4162.7548	p2=0.08187 w2=999.00000	Alignment (268 BEB sites): 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268

PTHR10573	-7359.3932	p2=0.13834 w2=16.85287	Alignment (37 BEB sites): 24 27 28 34 38 43 57 59 62 67 70 82 88 118 125 135 146 147 149 154 167 168 216 218 228 230 240 246 247 258 262 265 267 268 269 281 295
PTHR10815	-4394.4175	p2=0.07978 w2=127.07694	Alignment (13 BEB sites): 77 92 104 126 132 141 183 201 224 236 241 245 258
PTHR10942	-8427.9776	p2=0.06323 w2=162.92639	Alignment (10 BEB sites): 181 182 183 216 236 261 292 351 399 659
PTHR11365	-23733.612	p2=0.01231 w2=20.20186	Alignment (17 BEB sites): 117 186 415 446 564 704 744 748 852 987 1027 1183 1352 1396 1402 1411 1419
PTHR11510	-8349.4766	p2=0.10541 w2=999.00000	Alignment (31 BEB sites): 3 63 65 71 84 133 148 151 157 160 176 218 411 417 419 448 456 469 472 477 485 486 495 504 509 516 518 589 590 595 600
PTHR11522	-4486.2667	p2=0.06831 w2=999.00000	Alignment (29 BEB sites): 7 26 34 42 44 45 46 53 58 60 63 75 78 88 96 99 113 134 136 140 172 196 275 276 287 288 291 299 303
PTHR11943	-7545.1524	p2=0.02014 w2=742.65630	Alignment (5 BEB sites): 2 64 231 240 383
PTHR12493	-6232.6158	p2=0.04513 w2=19.32148	Alignment (6 BEB sites): 76 110 148 152 235 387
PTHR13041	-2450.956	p2=0.07902 w2=15.25148	Alignment (8 BEB sites): 15 30 32 49 93 98 100 101

PTHR13099	-1955.8493	p2=0.08445 w2=33.63521	Alignment (6 BEB sites): 31 33 39 68 85 102
PTHR13191	-6415.096	p2=0.02660 w2=998.99555	Alignment (9 BEB sites): 27 41 68 154 189 211 243 249 268
PTHR13255	-10900.267	p2=0.16884 w2=3.50327	Alignment (62 BEB sites): 45 79 80 83 89 92 96 101 102 103 107 110 113 118 129 133 135 142 153 154 156 161 167 170 171 176 180 205 206 212 216 232 239 245 258 259 262 265 271 298 326 338 344 356 381 389 396 397 400 410 412 416 426 433 501 503 504 522 533 540 544 549
PTHR13318	-6217.158	p2=0.14330 w2=360.57521	Alignment (38 BEB sites): 85 87 113 119 120 124 132 137 149 151 153 165 166 180 186 187 198 200 204 225 228 229 239 251 259 265 266 267 269 282 288 296 305 307 308 330 335 338
PTHR13558	-3159.1839	p2=0.04470 w2=666.30259	Alignment (4 BEB sites): 39 56 70 93
PTHR13595	-6449.9755	p2=0.03247 w2=998.95618	Alignment (6 BEB sites): 196 252 339 367 433 455
PTHR14162	-10079.386	p2=0.11137 w2=25.88982	Alignment (29 BEB sites): 33 50 57 65 67 81 112 151 153 157 173 174 187 193 199 217 218 219 222 234 257 343 351 357 399 415 429 441 475
PTHR14312	-5953.5428	p2=0.03960 w2=16.89209	Alignment (9 BEB sites): 88 108 161 230 238 249 267 268 399
PTHR14343	-12437.092	p2=0.01650 w2=998.96877	Alignment (4 BEB sites): 287 324 482 530

PTHR14520	-2446.9486	p2=0.06885 w2=999.00000	Alignment (110 BEB sites): 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110
PTHR14657	-8287.2304	p2=0.05387 w2=999.00000	Alignment (10 BEB sites): 41 57 112 181 253 336 368 392 401 405
PTHR14696	-14877.784	p2=0.19866 w2=2.90242	Alignment (161 BEB sites): 75 90 91 92 100 104 117 118 131 132 133 149 153 158 177 178 184 187 189 190 202 203 210 215 217 230 246 261 266 267 269 272 276 278 285 288 296 300 301 307 313 320 327 340 343 346 351 365 373 379 383 385 390 397 407 409 419 422 424 427 431 440 441 444 452 459 462 465 477 481 494 512 513 518 522 523 525 531 535 539 543 560 568 570 579 581 582 583 585 589 590 592 609 611 614 622 631 645 667 681 682 684 686 693 709 711 719 722 726 733 734 738 752 754 763 765 767 769 779 780 781 785 790 791 795 796 806 812 816 820 824 827 829 830 831 832 833 837 838 840 841 848 851 860 861 864 866 869 870 876 878 881 889 909 913 915 917 919 920 921 924
PTHR14870	-11170.714	p2=0.09178 w2=999.00000	Alignment (589 BEB sites): 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162

			163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589

PTHR15009	-14457.473	p2=0.06841 w2=12.82842	Alignment (39 BEB sites): 28 35 139 144 161 163 193 199 305 317 337 370 384 469 470 479 494 499 514 516 524 548 555 571 573 578 583 636 639 641 662 664 674 683 701 703 742 743 754
PTHR15054	-15789.658	p2=0.06082 w2=999.00000	Alignment (11 NEB sites): 162 296 558 574 799 829 941 967 1061 1065 1067
PTHR15139	-7368.9688	p2=0.03252 w2=103.52149	Alignment (9 BEB sites): 75 97 116 137 143 205 210 213 258
PTHR15153	-7090.6829	p2=0.11468 w2=285.81123	Alignment (29 BEB sites): 1 68 74 112 163 175 186 201 206 216 217 227 235 237 239 241 242 252 253 264 267 277 290 300 312 321 323 335 340
PTHR15463	-27031.909	p2=0.00346 w2=844.73097	Alignment (15 BEB sites): 115 381 447 546 574 723 808 836 878 996 1005 1088 1147 1148 1373
PTHR15486	-6665.3487	p2=0.13127 w2=4.07226	Alignment (50 BEB sites): 53 74 94 97 100 109 110 121 128 161 167 178 180 186 203 216 222 232 234 238 242 247 267 280 288 290 309 325 327 351 359 373 378 380 384 395 407 415 416 418 419 432 434 436 443 444 447 448 469 472
PTHR15510	-8701.4655	p2=0.04669 w2=999.00000	Alignment (6 BEB sites): 395 424 429 459 482 490
PTHR15511	-7502.9234	p2=0.08682 w2=648.24382	Alignment (23 BEB sites): 25 60 130 132 136 138 146 147 148 157 176 190 211 218 268 297 299 308 310 312 314 317 318
PTHR15555	-7627.3244	p2=0.05686 w2=119.54536	Alignment (10 BEB sites): 80 104 217 301 335 378 405 483 522 530

PTHR15926	-4294.0754	p2=0.03156 w2=998.85924	Alignment (2 BEB sites): 167 229
PTHR16078	-19774.829	p2=0.03958 w2=999.00000	Alignment (7 BEB sites): 102 139 289 302 459 577 830
PTHR16134	-8151.4651	p2=0.08440 w2=56.86778	Alignment (26 BEB sites): 30 33 68 70 92 117 127 135 138 143 195 197 199 213 238 242 243 245 246 250 258 281 297 317 331 366
PTHR16275	-31613.198	p2=0.01668 w2=998.96975	Alignment (10 BEB sites): 292 306 477 576 672 686 752 905 938 1211
PTHR17125	-3767.7065	p2=0.25940 w2=999.00000	Alignment (267 BEB sites): 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267

PTHR19372	-9379.6976	p2=0.06170 w2=150.38507	Alignment (23 BEB sites): 17 34 35 44 62 67 72 76 101 114 145 147 171 327 331 367 382 403 411 424 484 512 537
PTHR20879	-4569.5601	p2=0.10319 w2=191.66558	Alignment (17 BEB sites): 3 5 6 7 8 9 35 38 39 48 132 135 142 158 160 163 185
PTHR21035	-5198.6647	p2=0.05373 w2=466.98450	Alignment (3 BEB sites): 252 255 286
PTHR21248	-3736.9176	p2=0.05269 w2=20.87059	Alignment (7 BEB sites): 31 88 102 103 106 169 230
PTHR21465	-63138.678	p2=0.05892 w2=261.16630	Alignment (65 BEB sites): 738 739 740 896 900 1634 2520 2594 2599 2620 2624 2642 2746 2774 2891 2944 2945 2951 2954 2957 2981 3016 3064 3076 3077 3081 3084 3091 3111 3136 3141 3148 3183 3190 3225 3236 3264 3387 3445 3523 3532 3542 3556 3569 3582 3605 3614 3620 3762 3800 3804 3839 3856 3878 3885 3888 3905 3992 4055 4058 4091 4096 4098 4099 4119
PTHR21651	-4192.5876	p2=0.02871 w2=18.15864	Alignment (6 BEB sites): 19 99 114 137 200 255
PTHR21824	-6244.3924	p2=0.08047 w2=51.02658	Alignment (11 BEB sites): 68 112 148 174 178 211 223 269 298 329 340
PTHR21831	-21474.559	p2=0.04866 w2=999.00000	Alignment (28 BEB sites): 253 285 324 363 452 480 515 586 596 649 665 671 695 748 797 819 825 840 875 885 895 912 923 977 1009 1063 1080 1131

PTHR21844	-5134.7991	p2=0.12905 w2=999.00000	Alignment (28 BEB sites): 116 142 143 254 255 271 272 273 275 283 301 302 326 328 337 338 344 347 348 355 361 362 365 366 367 374 401 421
PTHR22708	-14900.253	p2=0.04552 w2=937.78597	Alignment (13 BEB sites): 24 105 119 185 189 191 296 319 396 414 537 556 604
PTHR22938	-20269.928	p2=0.04675 w2=30.34414	Alignment (29 BEB sites): 22 24 54 258 291 301 444 494 539 554 564 602 611 637 647 664 717 754 756 767 785 791 807 862 877 899 1033 1045 1049
PTHR23099	-7529.4868	p2=0.16140 w2=18.38904	Alignment (85 BEB sites): 47 48 180 192 214 223 228 230 233 241 269 277 279 283 291 292 296 301 313 345 347 368 369 372 374 389 394 396 398 400 401 412 413 422 424 426 430 431 434 488 489 492 493 504 506 511 514 516 519 543 547 548 549 561 563 568 576 584 590 595 598 604 612 615 617 639 648 651 652 692 697 706 709 716 719 727 730 731 732 739 743 745 755 763 765
PTHR23214	-17780.316	p2=0.04612 w2=117.17668	Alignment (18 BEB sites): 68 69 91 133 144 177 221 252 535 561 562 589 794 1147 1148 1164 1176 1199
PTHR23400	-12030.841	p2=0.03364 w2=328.56314	Alignment (13 BEB sites): 98 127 139 143 185 215 252 254 263 339 359 489 492
PTHR24388	-16089.348	p2=0.01328 w2=998.99166	Alignment (6 BEB sites): 111 169 439 565 720 832
PTHR31462	-12487.761	p2=0.02175 w2=271.58343	Alignment (8 BEB sites): 208 253 269 393 461 464 509 510

PTHR31667	-8727.2733	p2=0.14376 w2=999.00000	Alignment (36 NEB sites): 28 30 31 38 48 49 58 64 86 89 106 115 118 135 146 173 197 198 199 203 205 208 230 278 299 307 332 333 359 368 379 381 384 394 416 428
PTHR31840	-7756.3966	p2=0.11323 w2=213.90829	Alignment (32 BEB sites): 42 49 50 122 227 230 245 249 250 253 265 269 282 293 297 312 343 345 358 359 364 373 377 383 397 403 420 426 437 441 445 506
PTHR31932	-5675.9599	p2=0.05916 w2=766.33612	Alignment (10 BEB sites): 16 43 45 86 91 106 112 118 191 196
PTHR31974	-2945.0103	p2=0.27264 w2=999.00000	Alignment (51 BEB sites): 4 5 9 10 11 13 29 35 46 47 55 60 73 74 76 89 92 101 109 114 127 132 133 135 138 142 146 149 151 155 157 160 161 162 164 168 169 174 176 181 189 191 195 198 199 202 209 212 213 221 222
PTHR31991	-12300.579	p2=0.02251 w2=24.85564	Alignment (3 BEB sites): 50 290 569
TF101159	-15792.582	p2=0.05187 w2=748.90116	Alignment (30 BEB sites): 14 49 108 266 280 328 334 351 402 417 431 449 488 514 630 635 727 728 730 736 739 743 745 753 760 762 776 784 851 863
TF105825	-6457.37	p2=0.03343 w2=6.99519	Alignment (7 BEB sites): 95 144 166 169 176 309 332
TF328638	-4396.4554	p2=0.07457 w2=504.74615	Alignment (17 BEB sites): 222 230 234 242 255 268 269 283 287 293 299 300 315 316 325 327 334
TF328734	-8646.9205	p2=0.08797 w2=683.84849	Alignment (37 BEB sites): 121 128 129 154 156 172 229 233 234 238 319 347 353 358 391 395 397 398 399 401 403 405 406 410 416 418 423 425 428 429 459 461 473 480 482 483 508

TF328826	-51095.623	p2=0.08519 w2=4.19610	Alignment (133 BEB sites): 751 762 797 798 825 828 909 954 959 968 979 987 1015 1046 1069 1099 1103 1134 1211 1242 1287 1300 1301 1357 1364 1397 1398 1459 1494 1516 1536 1538 1552 1554 1555 1560 1561 1610 1615 1664 1671 1677 1696 1743 1769 1775 1790 1791 1799 1820 1835 1837 1859 1897 2004 2010 2028 2044 2068 2069 2101 2107 2112 2122 2128 2147 2162 2169 2171 2206 2232 2287 2289 2302 2303 2308 2419 2444 2456 2459 2474 2482 2489 2490 2508 2523 2530 2531 2560 2577 2601 2605 2607 2632 2647 2649 2710 2711 2724 2764 2775 2784 2786 2793 2799 2813 2881 2883 2884 2895 2902 2922 2930 2961 2976 2977 2986 2993 3008 3013 3016 3019 3063 3076 3079 3152 3172 3182 3191 3205 3273 3283 3291
TF329474	-4218.4537	p2=0.11052 w2=999.00000	Alignment (24 BEB sites): 58 61 73 74 78 104 109 119 129 131 136 139 140 142 162 168 169 182 191 193 212 213 259 268
TF330945	-9198.7889	p2=0.06744 w2=74.39295	Alignment (26 BEB sites): 61 81 94 99 124 155 190 212 229 234 319 323 340 350 378 402 407 417 424 437 440 446 457 480 499 520
TF332158	-18625.382	p2=0.06567 w2=320.63057	Alignment (47 BEB sites): 2 70 75 88 97 98 118 143 158 163 217 231 244 246 292 320 342 350 353 357 358 390 410 449 456 472 514 520 552 639 643 682 713 718 733 770 780 811 814 827 859 866 883 910 945 956 982
TF332961	-3697.1128	p2=0.18650 w2=82.32177	Alignment (42 BEB sites): 5 10 14 35 39 41 43 44 67 70 82 83 84 85 86 96 101 103 104 109 113 114 116 118 123 130 131 134 135 139 141 143 151 161 163 167 175 178 191 193 194 203
TF336144	-3408.9005	p2=0.25367 w2=999.00000	Alignment (192 BEB sites): 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91

			92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192
TF337247	-4652.8138	p2=0.04651 w2=972.34918	Alignment (6 BEB sites): 33 51 75 88 110 200
TF337679	-5652.9254	p2=0.03614 w2=570.75457	Alignment (11 BEB sites): 349 359 373 377 389 392 395 438 448 452 457
TF337703	-5761.1039	p2=0.06395 w2=999.00000	Alignment (5 BEB sites): 67 110 154 197 287
TF338349	-9332.7064	p2=0.04644 w2=998.82348	Alignment (10 BEB sites): 114 115 129 199 259 264 279 296 381 404
TF338954	-7620.1001	p2=0.07340 w2=315.72351	Alignment (19 BEB sites): 14 16 31 38 53 86 95 103 136 141 175 263 277 278 398 420 423 426 444
TF339799	-2869.6233	p2=0.07317 w2=999.00000	Alignment (6 BEB sites): 68 72 104 131 142 190
TF343523	-7018.4277	p2=0.21119 w2=375.81677	Alignment (86 BEB sites): 54 58 70 72 75 76 99 100 101 191 193 196 199 200 201 202 214 215 216 219 221 222 225 227 229 233 238 239 254 256 263 267 269 275 278 280 288 290 292 295 296 297

			298 300 301 302 303 304 307 308 310 311 314 315 316 317 318 320 321 324 325 330 331 332 342 344 354 385 397 399 401 402 406 447 470 478 490 514 534 589 597 600 606 616 620 624
TF600204	-15745.63	p2=0.01727 w2=52.59658	Alignment (7 BEB sites): 199 229 276 539 591 738 851
TF600216	-13035.657	p2=0.04231 w2=999.00000	Alignment (20 NEB sites): 71 86 327 336 338 342 437 461 489 497 509 541 551 552 570 581 583 584 616 638
TF600281	-28043.444	p2=0.01609 w2=107.58107	Alignment (5 BEB sites): 101 330 559 1055 1410
TF600298	-30207.558	p2=0.02860 w2=678.38808	Alignment (10 BEB sites): 149 189 309 468 478 667 730 761 1269 1410
TF600326	-5202.9632	p2=0.11991 w2=49.70853	Alignment (35 BEB sites): 14 18 21 22 25 26 43 52 60 64 75 81 89 104 108 112 123 124 125 151 155 181 184 191 195 212 217 222 223 243 245 250 267 270 285
TF600354	-36270.077	p2=0.02052 w2=60.18097	Alignment (19 BEB sites): 128 244 316 371 465 521 530 687 692 698 1009 1102 1127 1181 1185 1256 1266 1267 1483
TF600483	-9735.4858	p2=0.02685 w2=172.66583	Alignment (10 NEB sites): 15 65 100 199 369 410 462 505 515 532
TF600535	-6177.9323	p2=0.02399 w2=34.91606	Alignment (7 BEB sites): 27 54 96 137 164 222 314

TF600733	-11526.6	p2=0.02437 w2=16.01015	Alignment (7 BEB sites): 25 37 72 99 160 177 428
TF600816	-5276.0986	p2=0.04712 w2=998.80746	Alignment (7 BEB sites): 5 105 173 271 272 274 277
TF600873	-6023.498	p2=0.01998 w2=859.77290	Alignment (7 BEB sites): 7 32 223 366 370 428 434
TF601258	-3847.041	p2=0.04489 w2=25.04142	Alignment (11 BEB sites): 36 77 99 100 108 124 136 137 152 162 199
TF601520	-5678.5072	p2=0.03369 w2=333.28641	Alignment (8 BEB sites): 1 27 31 66 105 187 190 320
TF601591	-14083.419	p2=0.03632 w2=26.62131	Alignment (20 BEB sites): 191 494 563 632 633 647 648 650 697 698 699 700 705 837 843 862 867 875 907 910
TF601946	-6989.1726	p2=0.05162 w2=50.48616	Alignment (3 BEB sites): 96 122 147
TF602085	-3165.1676	p2=0.03259 w2=675.46904	Alignment (2 BEB sites): 1 96
TF602138	-15879.967	p2=0.03785 w2=4.25260	Alignment (6 BEB sites): 251 382 513 614 843 889

TF602256	-3523.7291	p2=0.10542 w2=104.18222	Alignment (20 BEB sites): 24 26 27 31 34 39 53 56 59 82 94 99 120 121 134 148 155 163 198 202
TF602914	-4100.7506	p2=0.08648 w2=999.00000	Alignment (44 BEB sites): 18 19 47 48 51 55 56 57 58 63 64 65 67 68 76 77 84 85 108 110 116 117 118 120 122 126 128 130 131 134 140 145 148 159 162 170 177 189 194 204 207 210 211 293
TF603138	-13547.771	p2=0.03431 w2=870.24078	Alignment (5 BEB sites): 182 310 490 528 535
TF603251	-5434.0457	p2=0.06113 w2=447.35176	Alignment (7 BEB sites): 4 34 81 111 155 156 274
TF603726	-14179.601	p2=0.19720 w2=22.45680	Alignment (97 BEB sites): 11 17 19 34 49 50 55 58 67 79 96 114 116 117 129 131 134 137 141 143 144 149 150 159 160 161 163 167 169 172 184 190 193 219 228 231 232 258 265 268 270 277 282 299 302 303 311 314 315 316 321 322 325 334 336 340 345 346 357 358 361 363 366 367 370 383 385 390 392 393 394 415 417 418 424 428 437 439 442 449 452 459 462 465 468 469 472 477 487 507 512 513 519 524 528 531 532
TF603987	-6986.2105	p2=0.09868 w2=11.49587	Alignment (18 BEB sites): 7 37 41 52 53 57 66 68 106 114 133 151 252 264 268 276 302 312
TF604204	-6636.4951	p2=0.05526 w2=362.82278	Alignment (21 BEB sites): 48 50 52 54 55 59 61 64 68 82 86 91 92 212 216 274 333 355 363 374 380
TF604370	-17568.818	p2=0.02142 w2=998.99542	Alignment (5 BEB sites): 49 114 192 712 718

TF605924	-2985.0502	p2=0.05792 w2=393.41665	Alignment (7 BEB sites): 115 116 117 118 120 128 129
TF607796	-10128.12	p2=0.16509 w2=999.00000	Alignment (116 BEB sites): 6 11 12 15 17 19 20 25 27 32 42 48 52 56 58 59 62 66 69 70 72 73 76 78 80 81 83 84 86 87 88 89 90 95 96 97 98 99 114 157 160 161 168 172 178 183 185 186 191 192 202 203 204 214 222 227 239 260 261 265 272 283 285 286 287 300 308 319 320 328 347 357 367 379 380 386 388 389 396 397 409 412 418 419 431 443 444 463 466 474 497 498 501 507 510 524 527 530 538 544 545 546 562 587 588 597 614 623 656 684 685 687 688 689 692 710
TF607946	-33347.21	p2=0.01550 w2=12.42581	Alignment (10 BEB sites): 152 303 334 377 605 612 753 1240 1241 1329
TF611168	-2911.0186	p2=0.04915 w2=999.00000	Alignment (3 BEB sites): 95 108 149
TF611725	-3105.9978	p2=0.04826 w2=999.00000	Alignment (9 BEB sites): 50 56 97 99 115 128 147 167 175
TF619452	-12188.189	p2=0.08679 w2=274.98026	Alignment (53 BEB sites): 18 20 36 82 138 140 143 145 146 180 183 195 219 221 255 302 324 326 331 333 335 338 341 342 343 348 350 383 398 414 425 432 463 466 469 517 521 523 544 559 570 576 593 601 602 603 622 632 666 677 680 705 708
TF623325	-1226.0519	p2=0.19512 w2=999.00000	Alignment (52 BEB sites): 3 4 5 6 9 11 12 13 17 21 22 24 26 27 30 32 37 39 40 42 43 44 45 48 50 52 53 58 60 63 64 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86

TF623981	-11869.847	p2=0.02756 w2=10.44499	Alignment (13 BEB sites): 421 437 630 661 673 701 726 818 854 919 920 973 978
TF626028	-9320.6905	p2=0.03501 w2=93.32813	Alignment (5 BEB sites): 233 280 306 346 402
TF626096	-14611.152	p2=0.07342 w2=29.70776	Alignment (61 BEB sites): 3 18 34 43 45 50 52 53 55 69 72 93 105 107 120 135 137 138 178 258 272 283 310 320 339 345 361 362 380 412 428 431 442 485 499 522 528 568 576 584 592 640 644 728 729 731 735 760 764 770 774 790 865 866 929 939 944 947 974 976 978
TF626791	-2935.1786	p2=0.05653 w2=7.69610	Alignment (3 BEB sites): 34 71 109
TF628883	-5629.3087	p2=0.02705 w2=19.89383	Alignment (5 BEB sites): 24 214 249 265 331
TF628955	-3797.081	p2=0.08745 w2=59.37246	Alignment (17 BEB sites): 42 43 53 61 70 102 107 111 113 144 168 177 205 216 229 231 263
TF644338	-18856.847	p2=0.00944 w2=256.49896	Alignment (4 BEB sites): 209 230 333 747
TF645735	-4275.6388	p2=0.04853 w2=12.35049	Alignment (7 BEB sites): 61 135 138 155 162 205 211
TF654454	-7390.3923	p2=0.04323 w2=362.91842	Alignment (9 BEB sites): 57 58 120 131 156 174 261 446 449

TF655513	-2637.7559	p2=0.05013 w2=17.18527	Alignment (5 BEB sites): 8 79 115 139 152
TF662368	-11110.272	p2=0.03802 w2=999.00000	Alignment (9 BEB sites): 12 582 662 764 771 919 930 931 932
TF662962	-7692.2088	p2=0.02849 w2=12.54014	Alignment (7 BEB sites): 25 219 242 306 357 419 433
PTHR10035	-3646.717	p2=0.02432 w2=27.89232	Alignment (3 BEB sites): 31 80 116
PTHR10412	-17061.204	p2=0.02612 w2=998.93927	Alignment (8 BEB sites): 20 91 128 138 242 340 364 428
PTHR10429	-8212.6583	p2=0.02857 w2=17.12545	Alignment (6 BEB sites): 124 128 246 365 382 433
PTHR10560	-7290.5721	p2=0.08058 w2=36.11785	Alignment (32 BEB sites): 57 91 97 119 136 149 204 218 221 224 229 237 247 248 251 253 254 255 256 276 280 307 325 326 327 338 374 376 377 382 386 387
PTHR10643	-12903.763	p2=0.01254 w2=21.98806	Alignment (7 BEB sites): 14 52 59 60 80 281 344
PTHR10805	-6011.0969	p2=0.01639 w2=328.35867	Alignment (5 BEB sites): 7 12 76 109 231

PTHR10882	-5377.0946	p2=0.02414 w2=12.55409	Alignment (6 BEB sites): 60 99 141 157 218 268
PTHR11001	-3288.1321	p2=0.04386 w2=999.00000	Alignment (3 BEB sites): 12 50 105
PTHR11215	-6838.8218	p2=0.08575 w2=357.75079	Alignment (24 BEB sites): 2 62 64 66 71 92 147 184 188 233 234 238 263 271 283 300 314 326 331 341 345 365 367 397
PTHR12045	-9623.7859	p2=0.02693 w2=998.95305	Alignment (7 BEB sites): 15 115 223 334 394 410 421
PTHR12066	-36003.667	p2=0.02302 w2=954.86770	Alignment (9 BEB sites): 121 122 140 166 197 508 878 1020 1110
PTHR12069	-12933.485	p2=0.00343 w2=191.51329	Alignment (2 BEB sites): 305 309
PTHR12083	-11078.33	p2=0.07817 w2=427.19687	Alignment (43 BEB sites): 5 8 9 23 49 52 60 63 65 78 89 91 163 164 171 222 226 229 311 367 369 376 391 392 394 411 422 428 462 474 482 498 500 503 509 535 557 563 578 579 648 655 659
PTHR12455	-13508.219	p2=0.03133 w2=61.42112	Alignment (14 BEB sites): 40 66 67 71 239 310 497 517 531 541 575 590 619 638
PTHR12770	-9293.4842	p2=0.02258 w2=999.00000	Alignment (4 BEB sites): 99 141 250 364

PTHR12863	-9110.0056	p2=0.02900 w2=998.97724	Alignment (7 BEB sites): 139 150 289 303 356 383 494
PTHR12872	-15754.853	p2=0.01612 w2=998.99713	Alignment (6 BEB sites): 398 450 483 486 725 730
PTHR12963	-12056.202	p2=0.02046 w2=8.69939	Alignment (9 BEB sites): 86 151 429 473 490 502 510 632 676
PTHR12981	-5517.8411	p2=0.06001 w2=999.00000	Alignment (16 BEB sites): 66 77 123 147 161 170 183 205 206 220 253 258 282 305 313 342
PTHR13052	-24815.851	p2=0.01017 w2=8.03002	Alignment (27 BEB sites): 43 186 193 200 220 221 362 366 414 473 498 528 749 769 827 843 850 862 989 1106 1121 1166 1203 1223 1229 1238 1248
PTHR13102	-13325.497	p2=0.02403 w2=999.00000	Alignment (10 BEB sites): 15 26 58 63 70 115 120 472 552 642
PTHR13113	-11154.189	p2=0.03013 w2=998.93799	Alignment (7 BEB sites): 175 196 237 253 360 378 447
PTHR13147	-7332.725	p2=0.07557 w2=13.51006	Alignment (22 BEB sites): 19 30 48 50 54 60 71 74 77 82 95 110 124 134 161 253 355 359 425 443 455 456
PTHR13211	-10149.383	p2=0.04821 w2=37.80743	Alignment (16 BEB sites): 72 88 114 225 253 260 261 331 423 480 493 506 510 513 522 528

PTHR13213	-32880.836	p2=0.06932 w2=386.89538	Alignment (36 BEB sites): 37 40 114 119 166 168 234 276 302 320 322 439 462 587 607 613 638 640 670 683 689 700 810 888 943 1035 1061 1069 1075 1122 1189 1211 1358 1416 1467 1470
PTHR13257	-12882.361	p2=0.02030 w2=21.84731	Alignment (6 BEB sites): 223 273 542 567 571 720
PTHR13325	-8222.4755	p2=0.37701 w2=2.54627	Alignment (104 BEB sites): 4 9 11 15 17 20 22 23 28 29 30 32 39 41 60 63 67 69 71 78 84 91 92 106 107 108 109 112 150 152 153 154 162 167 174 176 182 189 205 208 237 244 245 247 248 249 252 256 259 268 274 280 300 301 303 307 310 311 313 314 323 325 334 336 343 344 348 350 355 360 363 365 366 368 369 377 380 387 390 394 399 405 406 411 422 430 435 436 453 457 470 483 489 493 495 497 499 500 501 502 505 508 512 515
PTHR13603	-5129.157	p2=0.07653 w2=86.88080	Alignment (10 BEB sites): 11 26 65 74 106 110 114 123 201 205
PTHR13779	-13379.122	p2=0.04567 w2=7.07857	Alignment (21 BEB sites): 27 56 59 91 139 168 314 316 326 377 408 413 430 431 442 445 460 466 482 496 672
PTHR14115	-9760.342	p2=0.02535 w2=998.98803	Alignment (5 BEB sites): 366 372 374 450 454
PTHR14344	-28274.32	p2=0.11121 w2=28.17500	Alignment (98 BEB sites): 37 38 48 50 54 58 96 126 127 152 160 161 163 165 168 183 208 217 231 236 260 271 278 288 316 318 319 324 329 340 344 347 362 449 461 462 479 485 489 496 498 511 512 514 517 530 532 534 544 562 591 596 620 661 664 665 691 703 704 710 768 776 785 788 790

			826 833 835 846 851 852 877 896 905 908 919 950 962 963 999 1017 1024 1042 1081 1096 1119 1191 1192 1194 1249 1268 1270 1272 1274 1284 1291 1294 1323
PTHR14396	-27157.901	p2=0.00741 w2=37.10304	Alignment (12 BEB sites): 9 80 174 199 348 554 567 603 729 862 1044 1058
PTHR14416	-7883.3931	p2=0.00816 w2=173.41126	Alignment (3 BEB sites): 405 417 418
PTHR14487	-15296.106	p2=0.10315 w2=30.62227	Alignment (40 BEB sites): 138 143 148 178 185 200 202 206 208 209 214 269 278 296 298 314 347 363 368 379 390 394 395 412 414 450 470 474 524 530 532 533 545 548 561 571 575 586 685 698
PTHR14540	-8618.4496	p2=0.01146 w2=270.42635	Alignment (4 BEB sites): 109 161 489 491
PTHR14580	-5966.1714	p2=0.07959 w2=999.00000	Alignment (50 BEB sites): 5 20 24 27 57 60 61 63 66 68 69 73 76 80 85 87 96 100 101 102 110 114 115 119 120 126 127 128 129 132 133 144 155 158 159 160 162 186 189 211 238 266 269 272 287 289 290 292 294 323
PTHR14913	-7879.6917	p2=0.10221 w2=373.09107	Alignment (31 BEB sites): 136 156 158 159 164 167 171 177 184 201 205 214 226 231 269 270 280 325 341 342 350 354 358 361 372 384 419 426 438 442 446
PTHR15171	-5352.1137	p2=0.08028 w2=314.13386	Alignment (6 BEB sites): 17 46 136 227 228 229
PTHR15182	-12057.216	p2=0.03144 w2=998.92906	Alignment (10 BEB sites): 46 152 204 210 249 254 315 323 336 405

PTHR15289	-19904.41	p2=0.08322 w2=696.21190	Alignment (52 BEB sites): 12 13 55 79 128 224 256 273 275 276 277 282 283 284 285 286 287 288 292 299 319 322 356 367 384 390 412 440 466 467 516 522 550 601 631 637 638 641 663 665 666 687 697 729 806 849 865 897 955 962 1013 1032
PTHR15294	-32346.842	p2=0.01482 w2=54.78740	Alignment (4 BEB sites): 317 318 456 1112
PTHR15319	-26305.43	p2=0.05492 w2=197.30083	Alignment (46 BEB sites): 8 30 32 45 66 168 174 178 179 180 185 194 199 200 208 210 217 219 221 224 238 247 254 289 306 309 338 406 437 513 582 596 608 612 713 745 759 818 842 847 867 869 898 1084 1099 1178
PTHR15380	-7339.3371	p2=0.00867 w2=998.99853	Alignment (3 BEB sites): 74 239 304
PTHR15402	-8418.3083	p2=0.03517 w2=24.77636	Alignment (7 BEB sites): 52 133 138 188 200 224 266
PTHR15410	-13606.632	p2=0.10479 w2=999.00000	Alignment (34 BEB sites): 34 50 91 117 120 152 158 163 236 264 273 274 284 296 321 324 326 334 397 399 413 424 454 464 465 486 494 497 590 605 621 650 681 682
PTHR15440	-6889.1676	p2=0.01526 w2=12.98832	Alignment (5 BEB sites): 138 145 168 176 251
PTHR15499	-8144.0695	p2=0.00472 w2=17.65651	Alignment (2 BEB sites): 115 489

PTHR15561	-3642.3865	p2=0.03636 w2=999.00000	Alignment (3 BEB sites): 8 47 212
PTHR15666	-4874.062	p2=0.07307 w2=43.55499	Alignment (12 BEB sites): 45 46 85 102 138 141 147 191 210 247 254 257
PTHR15668	-12456.496	p2=0.03321 w2=19.14588	Alignment (16 BEB sites): 99 113 147 168 225 282 313 356 359 383 420 453 460 473 515 635
PTHR15705	-16306.941	p2=0.01311 w2=998.98566	Alignment (6 BEB sites): 126 242 280 345 409 642
PTHR15837	-3723.4699	p2=0.06985 w2=917.21787	Alignment (12 BEB sites): 107 165 203 204 206 216 231 246 257 263 271 279
PTHR15914	-2165.3174	p2=0.04249 w2=998.95977	Alignment (7 BEB sites): 19 60 62 63 69 82 85
PTHR15993	-12058.254	p2=0.04203 w2=999.00000	Alignment (9 BEB sites): 82 272 326 342 343 402 558 599 746
PTHR16049	-10186.932	p2=0.06507 w2=999.00000	Alignment (105 BEB sites): 15 20 22 23 113 116 117 118 120 126 137 141 142 143 144 149 159 160 161 164 165 166 167 168 171 173 177 178 182 184 186 189 190 191 199 205 206 207 208 209 210 211 213 214 215 218 219 220 221 222 225 226 227 229 230 231 232 234 235 238 240 242 244 245 246 247 249 250 261 262 263 265 266 268 290 291 293 296 298 300 305 320 323 327 332 333 368 369 374 376 377 379 381 382 383 384 386 387 388 390 391 394 432 520 522

PTHR16234	-4887.1864	p2=0.03104 w2=853.78840	Alignment (4 BEB sites): 21 75 177 218
PTHR16236	-3233.1013	p2=0.20170 w2=999.00000	Alignment (116 BEB sites): 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 32 35 36 40 44 47 48 49 50 57 58 59 60 61 63 65 67 68 73 74 75 76 79 81 84 85 87 89 92 105 107 115 119 122 130 133 135 136 138 149 150 156 157 161 164 167 173 174 175 177 178 181 182 189 195 201 202 204 205 206 207 208 209 211 212 213 214 215 216 217 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235
PTHR16510	-14562.156	p2=0.13137 w2=999.00000	Alignment (38 NEB sites): 17 19 24 80 94 179 191 194 204 206 232 237 243 250 277 287 292 305 314 329 361 370 438 447 472 490 496 539 542 560 577 579 590 613 621 624 627 820
PTHR17530	-6805.235	p2=0.07434 w2=17.13963	Alignment (11 BEB sites): 11 43 67 71 144 227 239 274 289 306 327
PTHR17920	-14603.731	p2=0.03103 w2=117.84086	Alignment (17 BEB sites): 90 91 102 328 333 341 355 368 391 430 488 551 556 567 609 679 686
PTHR19265	-12005.886	p2=0.02091 w2=7.55449	Alignment (4 BEB sites): 116 256 415 418
PTHR19317	-3122.5539	p2=0.05517 w2=999.00000	Alignment (8 NEB sites): 61 64 70 73 86 174 179 180
PTHR19847	-7625.5589	p2=0.04553 w2=999.00000	Alignment (579 BEB sites): 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59

			60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530

			531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579
PTHR20882	-13055.155	p2=0.01928 w2=998.96572	Alignment (5 BEB sites): 29 218 444 484 622
PTHR20973	-6177.9826	p2=0.14525 w2=2.71075	Alignment (27 BEB sites): 1 3 29 33 35 39 74 82 84 85 87 89 97 102 122 130 132 172 177 199 211 224 237 242 288 291 307
PTHR21180	-10218.535	p2=0.02343 w2=999.00000	Alignment (11 BEB sites): 3 16 120 136 188 242 384 394 505 559 567
PTHR21183	-5878.6179	p2=0.01315 w2=998.99403	Alignment (1 BEB sites): 229
PTHR21193	-3858.8551	p2=0.08746 w2=386.53646	Alignment (8 BEB sites): 75 76 94 115 117 135 138 179
PTHR21244	-3414.2741	p2=0.04120 w2=998.95985	Alignment (6 BEB sites): 30 33 60 83 127 143
PTHR21362	-11698.437	p2=0.18711 w2=999.00000	Alignment (210 BEB sites): 3 4 6 8 14 15 16 25 26 30 31 32 33 36 41 48 49 59 69 72 84 85 87 89 114 115 122 127 128 129 131 134 135 137 138 139 141 143 147 148 150 152 153 154 155 160 165 166 167 168 169 172 176 178 180 181 183 191 193 194 197 200 201 203 205 206 207 208 209 210 211 212 213 214 215 216 217 219 221 222 223 224 225 226 228 229 230 231 232 233 234 235 236 238

			239 240 242 243 245 246 247 248 251 252 253 256 257 258 261 262 263 264 266 268 269 271 272 273 274 275 277 280 281 282 283 284 285 287 288 290 291 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 313 314 315 316 318 319 323 324 325 326 327 333 337 346 348 349 350 351 354 357 364 366 373 374 386 388 391 401 405 407 409 410 412 413 428 429 430 440 442 443 455 481 487 489 496 499 508 512 513 514 520 522 526 531 535 538 540 541 542 544 550 551 552 558
PTHR21469	-45067.959	p2=0.02001 w2=998.97647	Alignment (31 BEB sites): 104 149 162 291 336 394 458 473 594 742 785 798 1036 1061 1073 1136 1188 1390 1469 1471 1486 1488 1496 1499 1511 1608 1644 1665 1931 1968 1994
PTHR21494	-14948.739	p2=0.01424 w2=999.00000	Alignment (6 BEB sites): 39 71 133 476 598 632
PTHR21695	-9574.0173	p2=0.08572 w2=42.99025	Alignment (19 BEB sites): 24 55 73 76 77 83 86 123 140 145 162 202 328 329 349 403 451 465 488
PTHR22684	-16504.042	p2=0.07928 w2=2.89212	Alignment (48 BEB sites): 42 60 67 80 82 95 152 179 181 182 186 187 188 223 225 226 227 243 245 256 257 264 265 272 288 297 306 327 353 359 377 388 396 399 409 413 444 511 528 565 568 586 595 597 609 611 614 695
PTHR22726	-12061.056	p2=0.01424 w2=108.18546	Alignment (3 BEB sites): 40 120 328
PTHR22915	-8869.0826	p2=0.01897 w2=12.35657	Alignment (7 BEB sites): 80 81 141 165 297 303 366

PTHR23011	-17432.741	p2=0.03965 w2=389.20277	Alignment (14 BEB sites): 146 151 160 164 169 322 343 351 610 696 705 900 901 922
PTHR23051	-7090.505	p2=0.01758 w2=90.33640	Alignment (4 BEB sites): 51 165 402 409
PTHR23052	-28548.954	p2=0.02651 w2=27.00338	Alignment (17 BEB sites): 8 42 100 255 330 477 478 547 561 659 684 708 750 818 916 986 1064
PTHR23216	-18826.044	p2=0.06229 w2=263.53013	Alignment (37 BEB sites): 35 41 47 64 65 67 72 96 112 156 158 166 176 183 188 205 360 365 385 411 471 525 662 665 671 672 675 707 708 903 916 1163 1164 1165 1210 1256 1285
PTHR23289	-6860.5971	p2=0.00920 w2=998.99964	Alignment (5 BEB sites): 2 14 34 254 312
PTHR23403	-12444.64	p2=0.03599 w2=22.97382	Alignment (10 BEB sites): 62 80 134 349 372 392 473 525 565 597
PTHR24164	-17662.785	p2=0.08343 w2=999.00000	Alignment (71 BEB sites): 8 64 81 83 86 88 98 101 110 111 136 139 142 154 157 158 167 183 190 196 198 199 200 209 210 212 214 222 223 224 225 226 228 235 245 246 247 253 342 355 359 368 388 392 408 423 424 426 445 462 469 477 490 493 496 527 528 530 535 584 608 651 654 657 658 659 745 746 793 910 917
PTHR24394	-27955.022	p2=0.00838 w2=58.32394	Alignment (5 BEB sites): 372 816 866 1170 1243

PTHR31230	-3651.4555	p2=0.05845 w2=24.85461	Alignment (11 BEB sites): 39 41 45 61 63 89 95 110 158 162 174
PTHR31514	-24193.07	p2=0.01147 w2=998.97748	Alignment (5 BEB sites): 24 27 62 196 774
PTHR31609	-6827.4304	p2=0.15477 w2=224.68706	Alignment (48 BEB sites): 12 23 36 48 73 75 76 79 86 99 103 108 114 144 150 153 158 161 167 168 175 177 182 186 189 193 194 203 207 212 214 215 219 221 225 226 259 261 286 289 299 301 307 309 310 312 314 319
PTHR31822	-5815.4635	p2=0.03446 w2=64.77052	Alignment (7 BEB sites): 3 38 127 197 247 255 265
PTHR31835	-5988.8934	p2=0.03472 w2=999.00000	Alignment (9 BEB sites): 88 154 157 179 182 185 270 311 335
PTHR32022	-17924.297	p2=0.01160 w2=998.99324	Alignment (5 BEB sites): 30 104 144 188 462
PTHR32078	-17840.805	p2=0.00916 w2=22.57790	Alignment (3 BEB sites): 231 336 446
PTHR32274	-13890.952	p2=0.04692 w2=999.00000	Alignment (19 BEB sites): 51 79 94 109 115 140 191 331 338 358 382 445 456 495 522 532 571 595 1007
TF314051	-7040.0471	p2=0.03324 w2=111.55364	Alignment (13 BEB sites): 81 110 113 205 208 216 309 313 333 335 345 354 378

TF329423	-2985.979	p2=0.08048 w2=591.08430	Alignment (13 BEB sites): 5 12 23 26 44 87 113 114 116 126 131 133 134
TF331577	-5761.4903	p2=0.16402 w2=1.61038	Alignment (17 BEB sites): 4 25 26 30 46 55 60 118 135 180 209 227 232 261 279 281 289
TF336187	-2753.1939	p2=0.13433 w2=999.00000	Alignment (25 NEB sites): 10 11 15 17 18 19 20 22 26 34 35 37 61 63 69 83 88 89 99 101 107 111 116 123 157
TF336688	-14043.66	p2=0.03418 w2=633.78406	Alignment (8 BEB sites): 15 208 210 289 299 364 418 458
TF338109	-4214.6505	p2=0.10595 w2=15.03484	Alignment (20 BEB sites): 34 50 55 63 68 69 93 117 121 130 133 141 158 159 165 169 170 176 186 189
TF338566	-10615.113	p2=0.03013 w2=4.84463	Alignment (8 BEB sites): 162 181 269 327 377 400 433 581
TF600024	-8519.242	p2=0.03235 w2=998.97021	Alignment (8 BEB sites): 15 238 242 247 366 474 478 496
TF600099	-22470.372	p2=0.01274 w2=109.44624	Alignment (13 BEB sites): 210 222 288 356 524 593 610 617 730 885 941 977 1416
TF600132	-16829.504	p2=0.02208 w2=10.81434	Alignment (9 BEB sites): 129 460 493 494 603 604 671 871 946

TF600156	-12274.807	p2=0.05540 w2=68.77167	Alignment (24 BEB sites): 58 60 79 80 89 91 112 120 123 127 141 143 148 152 165 174 176 186 188 189 227 245 407 467
TF600312	-8925.631	p2=0.03250 w2=276.33164	Alignment (5 BEB sites): 296 306 490 652 664
TF600472	-8093.3066	p2=0.01366 w2=11.10054	Alignment (5 BEB sites): 28 39 85 117 255
TF600863	-36516.836	p2=0.04014 w2=17.42237	Alignment (24 BEB sites): 35 75 178 180 181 187 262 383 388 413 425 511 528 529 843 866 900 902 1032 1195 1231 1308 1318 1400
TF601183	-18234.649	p2=0.01696 w2=391.97073	Alignment (7 BEB sites): 22 259 505 538 547 579 891
TF601676	-19076.405	p2=0.02925 w2=998.98705	Alignment (12 BEB sites): 45 84 132 180 181 189 310 376 468 571 622 887
TF601789	-6698.5022	p2=0.19535 w2=133.13422	Alignment (38 BEB sites): 9 28 30 47 66 69 71 74 81 87 109 110 118 120 124 125 130 136 162 168 176 179 184 185 191 195 196 215 218 219 256 260 261 270 300 311 321 324
TF601864	-9742.8361	p2=0.07084 w2=29.30205	Alignment (33 BEB sites): 10 39 40 70 114 121 126 135 147 150 161 162 166 169 195 202 232 251 252 273 313 352 368 401 439 440 452 463 465 500 531 533 539
TF601915	-19985.63	p2=0.00642 w2=12.76394	Alignment (4 BEB sites): 689 954 1034 1258

TF602000	-21445.373	p2=0.02017 w2=13.72976	Alignment (5 BEB sites): 348 492 664 730 764
TF604566	-21904.751	p2=0.09500 w2=515.80482	Alignment (54 BEB sites): 10 14 19 20 74 79 92 104 105 108 109 114 125 152 163 178 213 214 216 217 228 249 271 277 294 368 427 434 510 512 521 554 588 593 599 609 612 614 620 633 655 663 682 687 704 706 724 750 757 775 805 806 824 856
TF638404	-12558.524	p2=0.02845 w2=15.80374	Alignment (12 BEB sites): 167 204 475 479 493 522 568 584 606 620 633 644
TF644181	-22903.249	p2=0.02348 w2=999.00000	Alignment (4 BEB sites): 115 125 155 473

Supplementary Table 2

Protein IDs	Gene stable ID	Gene name	Gene description
PTHR10035	ENSG00000198821	CD247	CD247 molecule
PTHR10412	ENSG00000115275	MOGS	mannosyl-oligosaccharide glucosidase
PTHR10429	ENSG00000103152	MPG	N-methylpurine DNA glycosylase
PTHR10560	ENSG00000090534	THPO	thrombopoietin
PTHR10643	ENSG00000080986	NDC80	NDC80 kinetochore complex component
PTHR10805	ENSG00000105669	COPE	coatomer protein complex subunit epsilon
PTHR10882	ENSG00000117543	DPH5	diphthamide biosynthesis 5
PTHR11001	ENSG00000242114	MTFP1	mitochondrial fission process 1
PTHR11215	ENSG00000139637	C12orf10	chromosome 12 open reading frame 10
PTHR12045	ENSG00000151360	ALLC	allantoicase
PTHR12066	ENSG00000164362	TERT	telomerase reverse transcriptase
PTHR12069	ENSG00000284282	POLR3E	RNA polymerase III subunit E
PTHR12083	ENSG00000039650	PNKP	polynucleotide kinase 3'-phosphatase
PTHR12455	ENSG00000184967	NOC4L	nucleolar complex associated 4 homolog
PTHR12770	ENSG00000140688	C16orf58	chromosome 16 open reading frame 58
PTHR12863	ENSG00000103089	FA2H	fatty acid 2-hydroxylase
PTHR12872	ENSG00000108784	NAGLU	N-acetyl-alpha-glucosaminidase

PTHR12963	ENSG00000103671	TRIP4	thyroid hormone receptor interactor 4
PTHR12981	ENSG00000162300	ZFPL1	zinc finger protein like 1
PTHR13052	ENSG00000170322	NFRKB	nuclear factor related to kappaB binding protein
PTHR13102	ENSG00000196943	NOP9	NOP9 nucleolar protein
PTHR13113	ENSG00000130159	ECSIT	ECSIT signalling integrator
PTHR13147	ENSG00000179431	FJX1	four-jointed box kinase 1
PTHR13211	ENSG00000141499	WRAP53	WD repeat containing antisense to TP53
PTHR13213	ENSG00000132382	MYBBP1A	MYB binding protein 1a
PTHR13257	ENSG00000108559	NUP88	nucleoporin 88
PTHR13325	ENSG00000012174	MBTPS2	membrane bound transcription factor peptidase, site 2
PTHR13603	ENSG00000184857	TMEM186	transmembrane protein 186
PTHR13779	ENSG00000124535	WRINP1	Werner helicase interacting protein 1
PTHR14115	ENSG00000178522	AMBN	ameloblastin
PTHR14344	ENSG00000178252	WDR6	WD repeat domain 6
PTHR14396	ENSG00000092853	CLSPN	claspin
PTHR14416	ENSG00000108666	C17orf75	chromosome 17 open reading frame 75
PTHR14487	ENSG00000102977	ACD	ACD shelterin complex subunit and telomerase recruitment factor
PTHR14540	ENSG00000146576	C7orf26	chromosome 7 open reading frame 26
PTHR14580	ENSG00000143793	C1orf35	chromosome 1 open reading frame 35

PTHR14913	ENSG00000102871	TRADD	TNFRSF1A associated via death domain
PTHR15171	ENSG00000102119	EMD	emerin
PTHR15182	ENSG00000179841	AKAP5	A-kinase anchoring protein 5
PTHR15289	ENSG00000135451	TROAP	trophinin associated protein
PTHR15294	ENSG00000132952	USPL1	ubiquitin specific peptidase like 1
PTHR15319	ENSG00000103168	TAF1C	TATA-box binding protein associated factor, RNA polymerase I subunit C
PTHR15380	ENSG00000283208	AC001226.2	novel protein
PTHR15380	ENSG00000102805	CLN5	CLN5 intracellular trafficking protein
PTHR15402	ENSG00000101190	TCFL5	transcription factor like 5
PTHR15410	ENSG00000149929	HIRIP3	HIRA interacting protein 3
PTHR15440	ENSG00000102218	RP2	RP2 activator of ARL3 GTPase
PTHR15499	ENSG00000105856	HBP1	HMG-box transcription factor 1
PTHR15499	ENSG00000283847	HBP1	HMG-box transcription factor 1
PTHR15561	ENSG00000241258	CRCP	CGRP receptor component
PTHR15666	ENSG00000170619	COMMD5	COMM domain containing 5
PTHR15668	ENSG00000101997	CCDC22	coiled-coil domain containing 22
PTHR15705	ENSG00000189367	KIAA0408	KIAA0408
PTHR15837	ENSG00000108961	RANGRF	RAN guanine nucleotide release factor

PTHR15914	ENSG00000169877	AHSP	alpha hemoglobin stabilizing protein
PTHR15993	ENSG00000136929	HEMGN	hemogen
PTHR16049	ENSG00000160051	IQCC	IQ motif containing C
PTHR16234	ENSG00000116922	C1orf109	chromosome 1 open reading frame 109
PTHR16236	ENSG00000130748	TMEM160	transmembrane protein 160
PTHR16510	ENSG00000152595	MEPE	matrix extracellular phosphoglycoprotein
PTHR17530	ENSG00000170893	TRH	thyrotropin releasing hormone
PTHR17920	ENSG00000162542	TMCO4	transmembrane and coiled-coil domains 4
PTHR19265	ENSG00000138587	MNS1	meiosis specific nuclear structural 1
PTHR19317	ENSG00000105404	RABAC1	Rab acceptor 1
PTHR19847	ENSG00000100897	DCAF11	DDB1 and CUL4 associated factor 11
PTHR19847	ENSG00000259371	AL136295.3	novel protein
PTHR19847	ENSG00000284796	DCAF11	DDB1 and CUL4 associated factor 11
PTHR19847	ENSG00000284818	AL136295.9	novel protein
PTHR20882	ENSG00000174177	CTU2	cytosolic thiouridylase subunit 2
PTHR20973	ENSG00000169189	NSMCE1	NSE1 homolog, SMC5-SMC6 complex component
PTHR21180	ENSG00000122547	EEPD1	endonuclease/exonuclease/phosphatase family domain containing 1
PTHR21183	ENSG00000136522	MRPL47	mitochondrial ribosomal protein L47
PTHR21193	ENSG00000204237	OXLD1	oxidoreductase like domain containing 1

PTHR21244	ENSG00000062582	MRPS24	mitochondrial ribosomal protein S24
PTHR21362	ENSG00000111644	ACRBP	acrosin binding protein
PTHR21469	ENSG00000110723	EXPH5	exophilin 5
PTHR21494	ENSG00000100325	ASCC2	activating signal cointegrator 1 complex subunit 2
PTHR21695	ENSG00000133250	ZNF414	zinc finger protein 414
PTHR22684	ENSG00000141002	TCF25	transcription factor 25
PTHR22726	ENSG00000162600	OMA1	OMA1 zinc metallopeptidase
PTHR22915	ENSG00000042286	AIFM2	apoptosis inducing factor mitochondria associated 2
PTHR23011	ENSG00000149646	CNBD2	cyclic nucleotide binding domain containing 2
PTHR23051	ENSG00000115084	SLC35F5	solute carrier family 35 member F5
PTHR23052	ENSG00000162779	AXDND1	axonemal dynein light chain domain containing 1
PTHR23216	ENSG00000166197	NOLC1	nucleolar and coiled-body phosphoprotein 1
PTHR23289	ENSG00000014919	COX15	cytochrome c oxidase assembly homolog COX15
PTHR23289	ENSG00000285932	AL133353.2	novel transcript
PTHR23403	ENSG00000118094	TREH	trehalase
PTHR24164	ENSG00000104881	PPP1R13L	protein phosphatase 1 regulatory subunit 13 like
PTHR24394	ENSG00000184677	ZBTB40	zinc finger and BTB domain containing 40
PTHR31230	ENSG00000074842	MYDGF	myeloid derived growth factor
PTHR31514	ENSG00000146147	MLIP	muscular LMNA interacting protein

PTHR31609	ENSG00000161179	YDJC	YdjC chitooligosaccharide deacetylase homolog
PTHR31822	ENSG00000165120	SSMEM1	serine rich single-pass membrane protein 1
PTHR31835	ENSG00000149761	NUDT22	nudix hydrolase 22
PTHR32022	ENSG00000133943	DGLUCY	D-glutamate cyclase
PTHR32078	ENSG00000111554	MDM1	Mdm1 nuclear protein
PTHR32274	ENSG00000145911	N4BP3	NEDD4 binding protein 3
TF314051	ENSG00000188493	C19orf54	chromosome 19 open reading frame 54
TF329423	ENSG00000186973	FAM183A	family with sequence similarity 183 member A
TF331577	ENSG00000205765	C5orf51	chromosome 5 open reading frame 51
TF336187	ENSG00000214128	TMEM213	transmembrane protein 213
TF336688	ENSG00000166246	C16orf71	chromosome 16 open reading frame 71
TF338109	ENSG00000172301	COPRS	coordinator of PRMT5 and differentiation stimulator
TF338566	ENSG00000142698	C1orf94	chromosome 1 open reading frame 94
TF600024	ENSG00000123992	DNPEP	aspartyl aminopeptidase
TF600099	ENSG00000164989	CCDC171	coiled-coil domain containing 171
TF600132	ENSG00000170037	CNTROB	centrobin, centriole duplication and spindle assembly protein
TF600156	ENSG00000163006	CCDC138	coiled-coil domain containing 138
TF600312	ENSG00000197980	LEKR1	leucine, glutamate and lysine rich 1
TF600472	ENSG00000154153	RETREG1	reticulophagy regulator 1

TF600863	ENSG00000165757	JCAD	junctional cadherin 5 associated
TF601183	ENSG00000169064	ZBBX	zinc finger B-box domain containing
TF601676	ENSG00000164304	CAGE1	cancer antigen 1
TF601789	ENSG00000162399	BSND	barttin CLCNK type accessory beta subunit
TF601864	ENSG00000154479	CCDC173	coiled-coil domain containing 173
TF601915	ENSG00000140326	CDAN1	codanin 1
TF602000	ENSG00000165490	DDIAS	DNA damage induced apoptosis suppressor
TF604566	ENSG00000127666	TICAM1	toll like receptor adaptor molecule 1
TF638404	ENSG00000102904	TSNAXIP1	translin associated factor X interacting protein 1
TF644181	ENSG00000177483	RBM44	RNA binding motif protein 44