From DEPARTMENT OF BIOSCIENCES AND NUTRITION
Karolinska Institutet, Stockholm, Sweden

TRANSCRIPTIONAL REGULATION OF THE HUMAN PRE-IMPLANTATION EMBRYO

Elo Madissoon

Stockholm 2016

Published by Karolinska Institutet.
Printed by AJ E-print AB

ISBN 978-91-7676-152-6

Transcriptional regulation of the human pre-implantation embryo

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Elo Madissoon, M.Sc.

Principal Supervisor:
Professor Juha Kere
Karolinska Institutet
Department of Biosciences and Nutrition

Co-supervisor(s):

Dr. Shintaro Katayama
Karolinska Institutet
Department of Biosciences and Nutrition

Dr. Liselotte Vesterlund
Karolinska Institutet
Department of Biosciences and Nutrition

Dr. Virpi Töhönen
Karolinska University Hospital
Reproductive Medicine
Karolinska Institutet
Department of Biosciences and Nutrition

Opponent:
Professor Andras Nagy
University of Toronto
Department of Obstetrics & Gynaecology

Examination Board:

Professor Urban Lendahl
Karolinska Institutet
Department of Cell and Molecular Biology

Docent Elisabet Andersson
Karolinska Institutet
Department of Cell and Molecular Biology

Professor Juha Tapanainen
University of Helsinki
Department of Obstetrics and Gynaecology

To my Mother and Father,

for unconditional love.

ABSTRACT

Millions of couples worldwide have difficulties in conceiving a child. These couples, affected by infertility, suffer from symptoms of stress. The causes of infertility are largely unknown and current available treatment, *in vitro* fertilization has moderate success rates. The *in vitro* fertilization covers pre-implantation stage of the human embryo development. Better understanding of the molecular mechanisms in the early development might help to improve the *in vitro* fertilization methods.

Much of the knowledge on early development has been gained from model organisms: nematode, fruit fly, zebrafish and mouse. Mouse is a commonly used model organism for mammalian pre-implantation development. Global gene expression studies have been performed on mouse and human. Overall, gene expression changes seem to be similar in principle between the organisms. However, the expression of the first transcribed genes in the early development may not be as conserved between species as other genes expressed later in development. Therefore, human pre-implantation development should also be studied on human material for better translation of the results.

Genes with a dynamic expression profile in human pre-implantation development were identified in Paper I. Various criteria such as conservation, expression profile in mouse, relevance in cancer and novelty were applied to choose seventy genes of potential importance in human early development. Expression of those genes was studied in mouse and compared with human orthologues. The results showed differences in the expression profiles between human and mouse.

Paper II found novel regulatory elements and potentially important transcription factors from human early development by single-cell RNA sequencing. About 350 of oocytes and blastomeres from early embryos were studied, the promoters of activated genes were analyzed, and many PRD-like homeodomain genes with first-time evidence of expression were cloned. These genes were suggested to regulate the early development.

Paper III describes the expression pattern, target genes and potential function in early development for the PRD-like homeodomain gene LEUTX. A novel variant of LEUTX was identified and cloned, providing a full homeodomain-containing and functional isoform of the protein. LEUTX was found expressed exclusively in human pre-implantation embryos. The target genes of LEUTX were enriched among the genes activated by human embryo, which strongly indicated regulatory function for LEUTX in the early development.

The other PRD-like homeodomain containing proteins were studied in Paper IV. Expression of those genes was found to be specific for early development. The targets of CPHX1 and DPRX were found to be enriched among the genes activated in the early human embryo. General overlap of the target genes allowed for discussion of their possible functional redundancy.

The thesis offers novel findings for understanding gene expression and regulation in human pre-implantation embryos. The studies identify novel PRD-like homeodomain containing transcription factors that may have a crucial importance in the regulation of gene expression in the human pre-implantation embryos.

LIST OF SCIENTIFIC PAPERS

I. Madissoon E, Töhönen V, Vesterlund L, Katayama S, Unneberg P, Inzunza J, Hovatta O, Kere J. **Differences in gene expression between mouse and human for dynamically regulated genes in early embryo.** PLoS One 2014 Aug 4;9(8):e102949. doi: 10.1371/journal.pone.0102949. eCollection 2014.

II. Töhönen V*, Katayama S*, Vesterlund L, Jouhilahti EM, Sheikhi M, Madissoon E, Filippini-Cattaneo G, Jaconi M, Johnsson A, Bürglin TR, Linnarsson S, Hovatta O, Kere J. **Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development**. Nature Communications 2015 Sep 11;6:8207. doi: 10.1038/ncomms9207.

III. Jouhilahti E*, Madissoon E*, Vesterlund L, Töhönen V, Krjutškov K, Reyes AP, Petropoulos S, Månsson R, Linnarsson S, Bürglin TR, Lanner F, Hovatta O, Katayama S, Kere J. **The human PRD-like homeobox gene LEUTX has a central role in embryo genome activation**. Submitted manuscript.

IV. Madissoon E*, Jouhilahti E*, Vesterlund L, Töhönen V, Krjutškov K, Petropoulos S, Einarsdottir E, Linnarsson S, Lanner F, Månsson R, Hovatta O, Bürglin TR, Katayama S, Kere J. **PRD-like homeodomain transcription factors regulate human Embryo Genome Activation by target gene activation and suppression**. Submitted manuscript.

* Shared first authors

CONTENTS

LIST OF ABBREVIATIONS

bp	Basepair
cDNA	Complementary DNA
DNA	Deoxyribonucleic acid
EGA	Embryonic genome activation
EPI	Epiblast
ES-cell	Embryonic stem cell
FACS	Fluorescence activated cell sorting
FDR	False discovery rate
GV	Germinal vesicle oocyte
Hox	Homeotic gene
ICSI	Intracytoplasmic sperm injection
iPS-cell	Induced pluripotent stem cell
IVF	*In vitro* fertilization
LINE	Long interspersed nuclear elements
MII	Metaphase II oocyte
mRNA	Messenger RNA
PE	Primitive endoderm
PGD	Pre-implantation genetic diagnosis
PRD	PAIRED, a subcategory of the Hox genes
qPCR	Quantitative polymerase chain reaction
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RT	Reverse transcription
SINE	Short interspersed nuclear elements
STRT	Single-cell tagged reverse transcription
TE	Trophectoderm
TFE	Transcript far 5’ end
UMI	Unique molecular identifiers
UTR	Untranslated region

1 INTRODUCTION

1.1 SIGNIFICANCE OF STUDYING EARLY DEVELOPMENT

1.1.1 Impact of infertility on the quality of life

The ability to reproduce is essential for the continuation of every species. The unique genetic material for each individual can be propagated by their progeny in the next generation in a new genetic combination and in the next generation by their progeny and so on. Motivation for parenting, therefore, is considered to arise from instinctive behavior for the survival of the species (Benedek, 1980). Infertility is an individual's or couple's reduced ability to produce offspring. In humans, infertility is a worldwide concern with tens of millions of couples having problems with conceiving a child without medical intervention (Mascarenhas et al., 2012). The inability to follow these instincts in this aspect of life can cause psychological trauma to the couples. For example, infertility problem in women negatively affect various aspects in life: sexuality, mood, anxiety and fears, self-perceived attractiveness and general well-being (Oddens et al., 1999). Feelings of grief and loss are very common among both men and women in infertile couples (Alesi, 2005). Furthermore, stress itself can influence pregnancy, live birth delivery and baby's birth weight in women (Klonoff-Cohen et al., 2001). Conclusively, infertility strongly affects quality of life.

1.1.2 Development of infertility treatment

For long, the only treatments of infertility besides artificial insemination included hormonal stimulation for increased number of eggs (Greenblatt et al., 1961) or surgery for tubal infertility (Pandian et al., 2008). The largest leap in the history of infertility treatment is celebrated by the birth of the first baby born by *in vitro* fertilization (IVF) in 1978 – Louise Brown (Steptoe and Edwards, 1978). The IVF technique was developed by Dr. Robert G. Edwards and Dr. Patrick Steptoe, who in spite of public resistance devoted 10 years of work to optimize follicle aspiration, oocyte maturation, embryo growth conditions and hormonal stimulations before Louise was born (Edwards, 2001). The procedure enables women to produce a surplus number of oocytes that are fertilized by sperm in the laboratory prior to transfer to the uterus.

The importance of this field of medicine was acknowledged by awarding of the Nobel Prize in Physiology and Medicine 2010 to Robert G. Edwards for the development of in vitro fertilization (Nobelprize.org, 2010).

1.1.3 Current status of *in vitro* fertilization

The IVF methods nowadays allow for many different ways to help people: injection of sperm into the oocyte by intracytoplasmic sperm injection (ICSI) (Palermo et al., 1992) (Figure 1 A) when no fertilization occurs spontaneously; selecting a healthy embryo by pre-implantation genetic diagnosis (PGD) by sampling one or two cells of the pre-implantation embryo

(Handyside et al., 1992) (Figure 2 B); oocyte and/or sperm donation from a fertile to infertile person. The rate of successful IVF cycles in humans has not improved much over the years with the live-birth rate of less than 32% (Edwards, 2001; Kupka et al., 2014). This leaves space for improvement of the methods. However, not much is known for the normal physiology of the embryo development with up to 45% of expressed genes with an unknown function in the human oocyte (Zhang et al., 2007). Therefore, the understanding of the molecular mechanisms in the early development is of crucial importance.

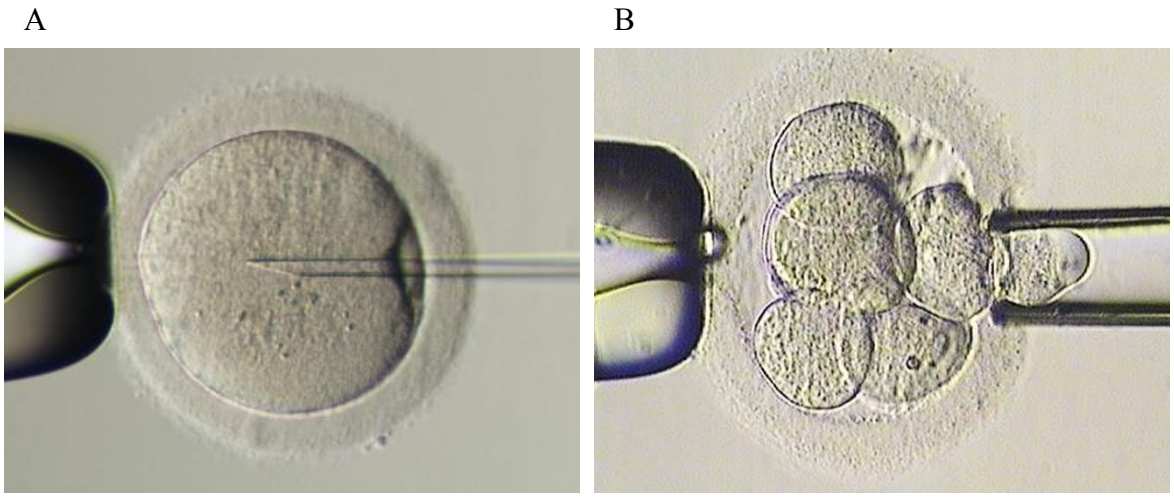

Figure 1. Microscope images of the human oocyte during the ICSI (A) and of the human 3-day old embryo during the PGD (B) procedures. Images were obtained from Virpi Töhönen.

1.2 EARLY DEVELOPMENT

1.2.1 Beginning of a new organism

Beginning of a sexually reproducing organism is the union of a large egg or oocyte cell from female and a small sperm from the male. This forms a fertilized egg or a zygote that will produce a whole organism by series of cellular divisions and differentiation steps of cell lineages. The first cell divisions (cleavage) produce rapidly dividing daughter-cells called blastomeres. The blastomeres formed by cleavage are halved in size after every division and together occupy the volume of an oocyte. Later development includes cell differentiation towards three germ layers while mammals first differentiate cells to form future extraembryonic tissues trophectoderm (TE) as well as primitive endoderm (PE) and epiblast (EPI) cells. The EPI cells form the inner cell mass (ICM) together with PE, and form future three germ layers of the whole organism. Although the embryos from different species have similar body plan in later course of development (Gilbert, 2000), the early embryos directly after fertilization are morphologically quite different (Figure 2 A). The polarity of embryos and asynchronous cell divisions occur already at first cleavage events in the nematode and zebrafish, but there is no apparent differences in the early blastomeres of mammals such as human and mouse.

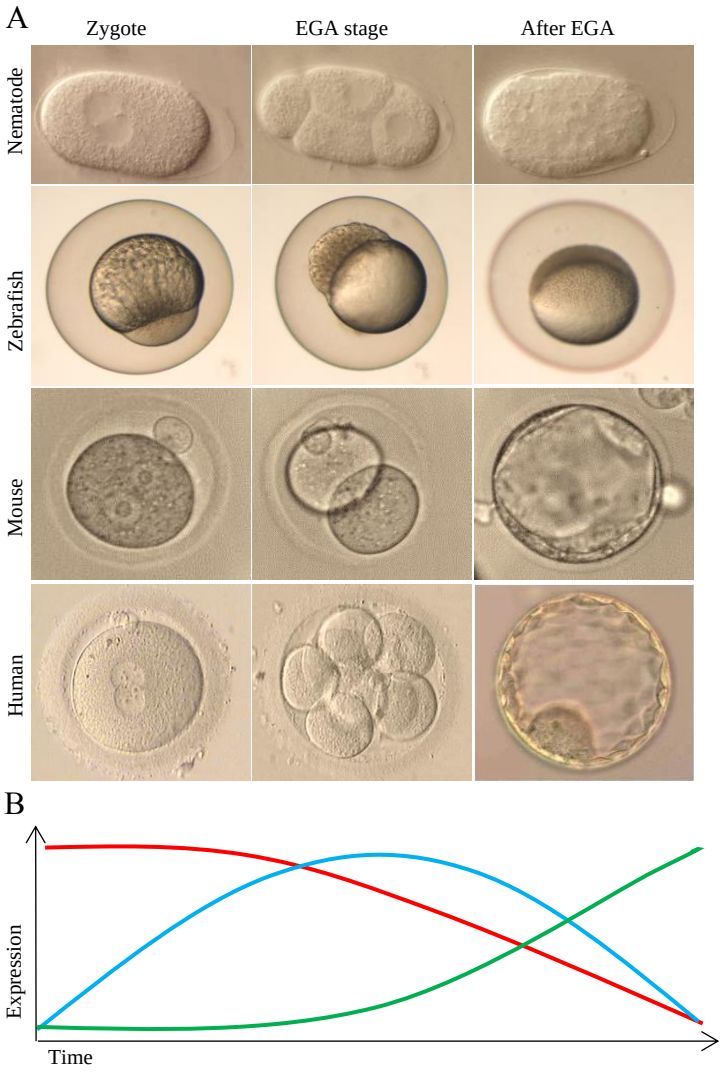

Figure 2. Early developmental stages from nematode, zebrafish, mouse and human for zygote, embryonic genome activation (EGA) stage and after EGA stage (A). Schematic representation of gene expression for the maternal transcript degradation (red), EGA genes (blue) and genes activated later in dvelopment (green) corresponding to the developmental stages on A (B). Mouse EGA stage is represented by 2-cell embryo while human EGA stage is represented by 8-cell embryo. The after EGA stage is represented by blastocyst both for human and mouse. Nematode images were obtained from Johan Henriksson and Thomas Bürglin, Zebrafish images were obtained from Liselotte Vesterlund, human blastocyst image from Virpi Töhönen, and human zygote and 8-cell embryo images from publication (Tohonen et al., 2015).

1.2.2 Transcriptional events in the early development

The early development includes transcription of many genes from DNA to RNA that can lead to the translation of protein. One transcript refers to one RNA copy from a gene on DNA, but one gene can be expressed in many copies of RNA. The whole collection of RNA molecules in a cell is called the transcriptome. The wave of induction and degradation of large number of transcripts in early development has been characterized in many studies. Mammalian development requires differentiation into embryonic TE and ICM cells, which requires further waves of gene expression. These different waves of transcriptional activation and degradation are called differently depending on the study: maternal mRNA degradation, minor zygotic genome activation (ZGA), early genome activation, mid-preimplantation gene activation, major ZGA, blastomere polarization, TE/ICM decision, EPI-PE decision, Morula

compaction, blastocyst cavitation, Major genome activation, minor gene activation, embryo gene transcription, Epi gene expression, TE gene expression and PE gene expression (Bell et al., 2008; Hamatani et al., 2004; Niakan et al., 2012; Wang and Dey, 2006; Zernicka-Goetz et al., 2009). The focus of the thesis will be on the first three days of human pre-implantation including the waves of maternal mRNA degradation and genome activation, here called embryonic genome activation (EGA).

The EGA occurs after the activation of the egg cell after the fertilization by sperm and after the start of cell divisions in embryo (Tadros and Lipshitz, 2009). This is accompanied by the degradation of transcripts present in the oocyte, called maternal transcripts. The EGA timing can vary a lot between the organisms: 1.25 h after fertilization at 4-cell stage for nematode, 2.75 h after fertilization at 512-cell stage in zebrafish (Tadros and Lipshitz, 2009), 22 h after fertilization at 2-cell stage in mouse (Flach et al., 1982; Hamatani et al., 2004) and 2-3 days between 4- and 8-cell stage in humans (Braude et al., 1988; Dobson et al., 2004) (Figure 2 B).

1.2.3 Regulation of gene expression

1.2.3.1 Transcription factors

Control of gene expression is often achieved by specific proteins in the cells, called transcription factors. Transcription factor binds to a specific sequence in DNA and usually triggers the expression of adjacent genes. Transcription factors are important for early development and many are known to have an important function (de Mendoza et al., 2013). Some genes are expressed in the oocyte, called maternal factors, without which the embryo will not develop. The examples include clk-2 gene in nematode (Benard et al., 2001) and *Hsf1* (Metchat et al., 2009) and *Nlrp*5 (Tong et al., 2000) in mouse. Other genes are expressed later in development and can have a role in lineage specification such as *Nanog* (Chambers et al., 2003; Mitsui et al., 2003) in mouse or in the formation of main body plan, such as the homeodomain-containing hox genes in fruit fly (Hughes and Kaufman, 2002).

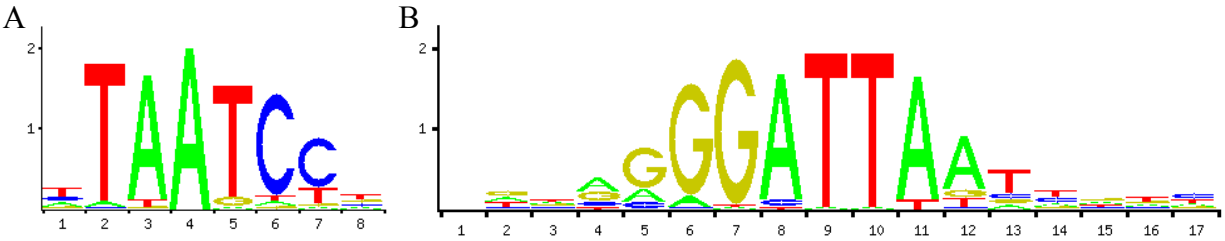

Figure 3. DNA binding motifs for the *OTX2* human (A) and *Otx2* mouse (B) proteins. Images are obtained from the JASPAR database.

Transcription factors act via DNA-binding sequences in the protein. For example the hox genes contain a specific DNA-sequence that codes for homeodomain protein sequence with about 60 amino acids (Burglin and Affolter, 2015). This amino acid sequence itself can bind to a specific base-pair sequence on DNA, which can influence gene expression in proximal genes. For example, the TAATCC sequence in the case of human *OTX2* homeodomain-

containing transcription factor (Jolma et al., 2013; Mathelier et al., 2014) (Figure 3 A).The DNA sequences with specific function is called a motif, such as the homeobox motif coding for homeodomain or the TAATCC motif that the homeodomain protein can bind to. The transcription factor binding site motifs for the same protein can vary between species as shown for human and mouse *Otx2* in Figure 3 B.

1.2.3.2 Homeodomain-containing transcription factors

Homeobox DNA sequence encodes for homeodomain amino acid sequence that forms a globular DNA-recognizing domain on a family of transcription factors (Burglin and Affolter, 2015; D'Elia et al., 2001; Qian et al., 1993). The homeodomain proteins are conserved between animals, having about 100 different homeobox genes in non-vertebrae and about 250 in vertebrates (Burglin and Affolter, 2015). There are many classes and subclasses of homeodomain proteins in human, including PAIRED (PRD) class that has been suggested to function in human germline development (Booth and Holland, 2007; Burglin and Affolter, 2015; Holland et al., 2007).

1.2.3.3 Epigenetics

In addition to transcription factors, epigenetic control of gene expression can be characterized. This includes mechanisms that change the accessibility of DNA for transcription factors. Epigenetic marks on DNA can make the gene accessible or not accessible for transcription. Such epigenetic marks could be histone modifications or methylation of DNA.

DNA is wrapped around histone proteins. Addition of chemical groups to the histones can make DNA more or less compacted, and thus more or less accessible to other proteins, such as transcription factors. DNA in human mature oocyte and sperm is tightly packed and transcriptionally silent. Human oocytes carry histone marks that inhibit transcription while the permissive acetylation marks are added rapidly after fertilization (Fulka et al., 2008). Human sperm DNA is tightly packed with sperm-specific protamine proteins that enable tight packaging of the DNA. The protamines are rapidly replaced by the histones from the oocyte cytoplasm after fertilization (Fulka et al., 2008).

DNA methylation is covalent addition of a methyl group to a C nucleotide that is next to a G nucleotide (CpG). Methylation silences the gene expression locally and can lead to more global changes in the histone marks, which result in further silencing of the whole DNA region. The precursor cells of both oocyte and sperm – primordial germ cells – are de-methylated during the development in fetus (Gkountela et al., 2015; Smallwood and Kelsey, 2012). Only the differentiation into more specialized germ cells allows for global *de novo* methylation of the genome in a sex-specific manner (Gkountela et al., 2015). Once the egg and sperm are ready to combine, their DNA is tightly packed in an epigenetically inaccessible form. The oocyte is loaded with transcripts called maternally expressed genes. Directly after fertilization, global de-methylation occurs for both the maternal and paternal genome, except for small regions that are specific to mother or father (imprinted regions) (Smith et al., 2014). This de-methylation period enables for the massive transcription of genes during the EGA (Howlett and Reik, 1991; Monk et al., 1987). The de-methylation period is shortly after followed by *de novo* methylation of the genome.

The differences in the methylation between mammalian species were noticed already before the spread of the genome-wide studies (Fulka et al., 2008). Later study have shown that although the global methylation patterns between human and mouse early development are similar, the maternally contributed methylation is targeted to species-specific set of CpG island promoters with only 7,5% overlap (Smith et al., 2014). They also showed that mouse has two times more hypomethylated CpG islands in the oocyte compared to human. This finding was further supported by whole-genome bisulfite sequencing study that showed that the maternal genome was de-methylated to a much larger extent in mouse compared to human (Okae et al., 2014).

Studies have also indicated differential methylation for the repeat sequences of the human genome. Smith et al. showed that the regulation of repetitive elements is much more diverse in human than in mouse pre-implantation development (Smith et al., 2014). Further differences between the species have been shown in centromeric satellite repeats that were hypermethylated in human oocytes, but not in mouse (Okae et al., 2014). Analysis of a primate-specific retroviral element HERVK shows that it is actively transcribed in human early development, although silenced everywhere else in the genome (Grow et al., 2015). The activity of primate-specific elements in pre-implantation development shows clear differences between the gene regulation between human and mouse species.

1.2.4 Repeated regions

More than half of the human genome consists of repeat sequences (Lander et al., 2001). These repeated sequences of DNA vary in length and function. For example there are simple sequence repeats such as repeats of an A nucleotide, inactive copies of genes called pseudogenes and long or short interspersed elements (LINE's or SINEs). Alu repeats are a family of SINE's that cover about 10% of the human genome and contain over 25% of the CG sequences that are often enriched in the gene regulatory elements (Lander et al., 2001; Rhead et al., 2010). These Alu elements can play a role in regulating gene expression by acting as enhancers for gene expression (Su et al., 2014). They can also containing binding sites for transcription factors (Polak and Domany, 2006; Tohonen et al., 2015). Alu elements are evolutionarily found only in the primates, rodents have another type of sequence elements called B1 (Tsirigos and Rigoutsos, 2009). Both have diversified from the 7SL RNA sequences, but resemble each other very little. They have been suggested to have distinct functions in regulating gene expression and have been under positive selection.

1.3 PRE-IMPLANTATION DEVELOPMENT IN MAMMALS

1.3.1 First few days of the embryo development

The earliest period in mammalian early development is the pre-implantation phase where the embryo develops without cellular attachment to the mother. This period includes fertilization and cell division accompanied by travelling along the oviduct to the mother's uterus. The fertilized egg divides into two while producing a 2-cell embryo with half sized blastomere cells. These divide further into 4- and 8-cell embryo until the stage of compaction where blastomeres start attaching more strongly to each other and the cell contours become less

visible in the microscope. The cells continue to divide further, start differentiating into different cell types and forming a cavity inside the embryo by the blastocyst stage after 6 days of development in human. Embryos at this stage comprise of two major cell types: outer layer of TE cells that attach to the uterus of the mother, and ICM that gives rise to the fetus (Gilbert, 2000). The ICM cells are important in the laboratory for the ability to give rise to embryonic stem cells (ES-cells) that can be used for studying the mechanism of undifferentiated cells from donated embryos to the research (Evans and Kaufman, 1981; Martin, 1981).

1.3.2 Differences between mouse and human

Most of the knowledge from the mammalian pre-implantation development comes from the mouse model organism. Although many morphological similarities exist, there are differences between mouse and human pre-implantation period. First, the mouse pre-implantation development takes 4 days while it takes 6 days in human (Gilbert, 2000). The differentiation or lineage determination in mouse has been detected already at 4-cell stage, although not even shown at the 8-cell stage in human (Biase et al., 2014; Cockburn and Rossant, 2010). The further differences in the blastocyst stage and in the ES-cells show species-specific differences in surface markers, X-chromosome inactivation, dependency of leukemia inhibitory factor, differentiation potential and gene expression (Blakeley et al., 2015; Ginis et al., 2004). The differences in gene expression in early embryos between species further suggests for different control mechanism between the organisms (Heyn et al., 2014; Jiang et al., 2014; Madissoon et al., 2014). The studies have analyzed global gene expression patterns in different species and found that maternally expressed genes in the oocytes are more similar than the genes expressed in EGA. Furthermore, bovine RNA sequencing study has found hundreds of genes specific for certain stage of pre-implantation embryo development (Jiang et al., 2014). They also demonstrated that the gene expression profiles are more similar between human and cow than between human and mouse.

Further differences in human and mouse early development have been shown with primate-specific endogenous retroviruses and repeat elements (Barbulescu et al., 1999; Grow et al., 2015; Tohonen et al., 2015). Some of these primate-specific elements in the human genome are shown to be transcriptionally active in the human early development, especially during the EGA.

Altogether, in spite of morphological similarities between human and the mouse model organism, many differences have been shown. The motivation arises from improving human pre-implantation development; therefore studies should be performed also in human embryos in addition to other model organisms.

1.4 GENOME-WIDE ANALYSIS FOR STUDYING GENE EXPRESSION

1.4.1 Microarray studies

The development of gene expression microarray technologies allows for measuring the expression of thousands of genes simultaneously in one sample. The method detects the presence and abundance of molecules originating from a gene that bind specifically to a synthesized DNA sequence called a probe. The strength of signal from each probe is correlated to the expression of the gene. Amplification of the samples with small volumes enabled the use of microarray technologies on pre-implantation embryos.

Expression microarray analysis in pre-implantation embryos was first done on mouse as the mammalian model organism with potential for gaining large number of oocytes and embryos (Hamatani et al., 2004; Wang et al., 2004; Zeng et al., 2004). Previously known concepts of maternally degraded transcripts and EGA were confirmed on a large scale as waves of gene expression together with approximate number of expressed transcripts and stage-specific expression of certain genes.

The first experiments using expression microarray on human oocytes enabled identification of oocyte-specific genes (Kocabas et al., 2006; Zhang et al., 2008). The studies compared human oocytes to embryonic stem cells, and detected previously known genes specific to the human oocytes such as *MATER*, *STELLA* and *FIGLA*. The studies also revealed the large proportion of genes with unknown function among the detected genes in the oocytes: 20% out of 7560 (Kocabas et al., 2006) or 45% out of 10183 (Zhang et al., 2007). The whole human pre-implantation development was addressed by two independent studies (Xie et al., 2010; Zhang et al., 2009). Both datasets described major changes in the gene expression during this period of development (Figure 4). A large number of genes were down-regulated after fertilization until the 8-cell stage, and a large number of genes were up-regulated until the 8-cell stage. The same applied for the development from 8-cell stage to the blastocyst stage. However, apparently there was no major difference in gene expression in the earlier stages of development before the 4-cell stage. This describes the major waves of transcriptional process after fertilization, shown schematically on Figures 2 and 4, and can take place during different cell numbers and times of early development in different organisms (Tadros and Lipshitz, 2009).

Microarrays have a major deficiency when it comes to measuring gene expression in the whole genome. Namely, the genes selected for microarrays are based on current knowledge of the genome and expression of genes. Therefore, the discovery of potential new genes or gene variants is impossible. Although single-cell expression studies were performed on human pre-implantation embryos, the method is not suitable for large-scale sensitive single-cell studies (Galan et al., 2010). Another method has been developed in recent years to overcome these shortcomings: RNA sequencing.

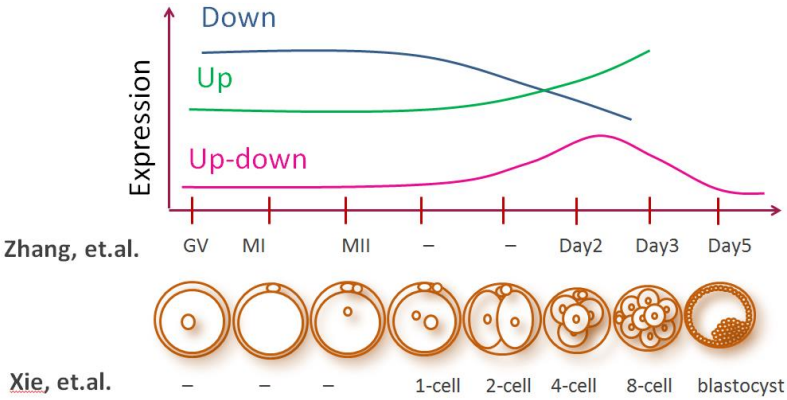

Figure 4. Schematic representation of the transcriptional waves in the example of two microarray studies by Xie et al., 2010 and Zhang et al., 2009.

1.4.2 RNA sequencing studies

The RNA sequencing (RNA-seq) method allows for measuring gene expression everywhere in the genome, not just for the previously known genes as is the case for microarrays. Furthermore, different RNA-seq methods enable to quantitatively measure expression at the beginning or end of a gene, detect different variants of the expressed gene and analyze allelic gene expression.

Application of the RNA-seq methods for pre-implantation embryos required the development of the methods to detect RNA in a very small sample, or a single cell. The first RNA sequencing study in mouse oocytes proved the principle that the novel method can be used to assay for gene expression in a single cell (Tang et al., 2010). A few studies in human and mouse pre-implantation development have analyzed oocytes and single cells from the embryo until the blastocyst stage (Blakeley et al., 2015; Tohonen et al., 2015; Xue et al., 2013; Yan et al., 2013). All the RNA-seq studies confirmed earlier findings from microarray concerning the dynamic expression pattern: maternally expressed genes, EGA genes and the genes activated later in development. The RNA-seq methods detected even more than 20000 gene expressed in single stage and identified a large number of long non-coding RNA-s that were expressed in a stage-specific manner (Yan et al., 2013). Xue et al. discovered that the embryos at 8-cell and morula stages still exhibit monoallelic maternal expression origin (Xue et al., 2013). The studies also looked at similarities and differences between human and mouse. While Xue et al. looked for similarities in the expression of stage-specific genes, then Blakeley et al. looked for differences in the blastocyst lineages. Both studies found what they were looking for, demonstrating that there are fundamental conserved genes as well as species-specific mechanisms in the early embryos.

In conclusion, the whole genome transcriptome analysis of the pre-implantation embryos has revealed a large number of previously unknown genes expressed during this time-period. The expression of these genes is highly dynamic representing maternal transcript degradation, EGA and other transcriptional events that demonstrate the stage-specificity of many transcripts. Both similarities and differences between human and other species can be found. The published datasets can be used for meta-analysis, and re-analysis can provide new knowledge of the pre-implantation development.

2 AIMS OF THE THESIS

The overall aim of the thesis was to expand the knowledge on the transcriptome of the human pre-implantation embryo and to identify genes and gene networks that are involved in human embryo genome activation.

Paper I

The aim of the study was to identify genes relevant in human pre-implantation embryo that could be compared with the mouse.

Paper II

The study aimed to identify the first genes expressed during human embryo genome activation and study mechanisms of fundamental importance in the transcriptional regulation of human pre-implantation development.

Paper III

The aim of the study was to characterize the role of the PRD-like homeodomain gene LEUTX in human pre-implantation development.

Paper IV

This study aimed to study the functions of several PRD-like homeodomain genes that were found to be expressed in early human development.

3 METHODS

This section summarizes the materials and methods used in the thesis. The detailed descriptions can be found in Paper I through Paper IV.

3.1 EXPERIMENTAL METHODS

3.1.1 Study materials

3.1.1.1 Mouse oocytes and embryos

The 4-7 weeks old FBV/N mice were stimulated by 5 IU of Pregnant Mare Serum and 5 IU of human chorionic gonadotropin for about 64 and 20 h before the collection of MII oocytes or zygotes after overnight mating with a male mouse from the same strain. The embryos were cultured in KSOM media (Millipore) under ovoil-100 (Vitrolife) under standard cell culture conditions (5% CO2). Mice were studied in Paper I.

3.1.1.2 Human oocytes and embryos

According to Swedish law, the surplus embryos from IVF procedures can be stored for up to 5 years, and must be thereafter discarded. The embryos used in research were gained from couples who were asked to donate their embryos to research as an alternative to discarding. The oocytes and zygotes used in the study were leftover oocytes or zygotes that did not find usage in the IVF procedure for the couples.
Human oocytes and zygotes were collected individually into 5 ul lysis buffer after removal of zona pellucida by acid Tyrode's solution. The cleavage-stage embryos were frozen for storage, and thawn for culturing into the G1/CCM medium under standard conditions as performed in the IVF Clinic (5% CO2 / 5% O2). The blastomeres were collected after laser-assisted biopsy of the zona pellucida. The whole 8-cell embryos were collected into lysis buffer after removal of zona pellucida. Oocytes and embryos were studied in Papers II-IV.

3.1.1.3 Cell lines

The human embryonic kidney cell line HEK293 was used in the study as an easily transfected cell line suitable for overexpression of the protein for Western blot, and for the luciferase studies where co-transfection of multiple plasmids was necessary (Papers II-IV).
Human embryonic stem cells (hES-cells) were used as a relevant cell line to study the function of embryonic transcription factors (Papers III and IV). The cell line HS401 was used for overexpression experiment, the cell lines HS401, HS980 and H9 were used for qPCR, and cell lines HS980 and HS983a were used for single-cell RNA-seq.

3.1.2 Experimental procedures

3.1.2.1 Quantitative PCR

Quantitative PCR (qPCR) for Paper I on mouse oocytes and embryos was performed using TaqMan Low Density Array Cards (TaqMan) according to the manufacturer's instructions. RNA was extracted by Arcturus PicoPure RNA isolation kit. RNA was converted to cDNA using High Capacity cDNA Reverse Transcription Kit. *Gapdh*, *Actb* and *Psmb6* were used as housekeeping controls.
The qPCR for Papers III and IV was carried out using Fast SYBR Green Real-Time PCR system. The RNA was extracted from HS401, HS980 and H9 and converted to cDNA by SuperScript III First-Strand Synthesis Supermix. The cDNA from the ES-cells and from the 8-cell embryo libraries were used for detection of the signal by real-time PCR using FastStart Universal SYBR Green Master mix. *GAPDH* was used as a housekeeping control. The amplicons were cloned into a TOPO vector, and verified by Sanger sequencing.

3.1.2.2 STRT RNA sequencing

The single-cell STRT RNA sequencing for Paper II was done according to the published protocol (Islam et al., 2012; Kivioja et al., 2012). The method enables a sensitive analysis of transcript start sites in single cells in a strand-specific manner. Only polyA-containing RNA molecules were captured by the cDNA synthesis protocol using poly-T priming followed by template-switching and introduction of barcode for pooling. The library preparation including amplification, fragmentation and adapter ligation was done in one reaction for the whole library. The sequenced reads derived from the very 5' of the transcripts.
The modified STRT RNA-seq method was used for Papers III and IV (Krjutškov et al. submitted). The sample collection procedure for single-cell picking of blastomeres and ES-cells was performed by capillary pipette. The samples for overexpression in hES-cells were collected via FACS-sorting of 75 cells per reaction. The sequencing of STRT libraries was performed on the Illumina platform creating 59-basepair long reads including the 9 bp barcode and 4 bp unique molecular identifier (UMI).

3.1.2.3 Single-cell sequencing by alternative method

An alternative sequencing method was used in order to get access to full-length cDNA template and to confirm our findings by an alternative sequencing method. The single-cell RNA sequencing method published by Tang et al. was used in Papers II-IV (Tang et al., 2010). The method also captures only polyA-tailed RNA, but is not strand-specific and is biased towards the end of the transcript (3'-end). The sequencing was performed on the SOLiD platform for Paper II. The library preparation was stopped before shearing for cloning of the genes in Papers II-IV.

3.1.2.4 Molecular cloning

The modified pFastBac expression vector was used for creating an expression vector with the following construct: CMV enhancer – EF1α promoter – AscI restriction site – PacI restriction site – internal ribosomal entry site element (IRES) – enhanced green fluorescent

protein (eGFP) – WPRE enhancer element (Papers II-IV). Cloning of the genes was achieved by inserting the AscI and PacI restriction sites' sequences to the cloning primers containing the start and stop codons.
The homeobox genes were amplified from the human ovary cDNA (Clontech) or 8-cell embryo from our sequencing library by Tang, et al. by PCR. Different isoforms were obtained by designing primers based on the publicly available gene models and the TFE-s observed from the STRT RNA-seq data in Paper II. The genes were cloned into the TOPO TA and V5-His-TOPO vectors. The genes were subsequently cloned into pFastBac vector by the addition of AscI and PacI restriction sites to the Forward and Reverse primers, respectively. (Papers II-IV)
The promoter region from *ZSCAN4* was cloned into the basic pGL4.11 or minimal-promoter-containing pGL4.25 luciferase vector (Paper II and IV). The identified 36-basepair motif from Paper II was cloned in one and four copies into the minimal-promoter-containing luciferase vector for Paper III.
qPCR fragments were cloned into the pCRII-dual promoter TOPO vector according to the TOPO TA cloning kit (Papers III and IV).

3.1.2.5 Luciferase reporter assay

The HEK293 cells were co-transfected with promoter-containing or empty luciferase vector, renilla luciferase vector for normalization and pFastBac vector containing the gene of interest. The detection of promoter-containing luciferase signal was normalized to the renilla luciferase signal and compared to the control promoterless vectors for every gene in the pFastBac vector (Papers II and III).

3.1.2.6 Immunofluorescence labelling

The 8-cell embryos or hES-cells were washed, fixed and permeabilized prior to treating with the anti-*LEUTX* antibody. After following washes, the secondary Alexa fluor 647 fluorescence-labelled antibody was used. The nuclei were labeled by blue fluorescent DNA-binding dye Hoechst 33342. The images were included in Paper III.

3.1.2.7 Transfection of cells

The HEK293 cells grown in DMEM with 1g/l glucose, L-glutamine, pyruvate and 10% FBS were transfected using the Lipofectamine2000 reagent according to manufacturer's instructions.
The ES-cell line HS401 grown on Laminin-521 in mTeSR1were trypsinized and suspended in DMEM. The transfection solution of 1 ug of pFastBac plasmid DNA and 3 ul of Lipofectamine 2000 were mixed previously in 50 ul DMEM. About 100 000 cells in 50 ul of DMEM were mixed with 50 ul of the transfection solution, and left to settle to the bottom of laminin-coated 96-well Sarstedt culture plate. The media was changed to mTeSR1 after exactly 15 min. This protocol enabled up to 35% of transfection efficiency.

3.1.2.8 Western Blot

The HEK293 cells were transfected to overexpress proteins from vector with gene of interest in frame with the V5 tag (Paper III). The proteins were extracted, denaturated and separated in a 10% Bis-Tris gel. Following transfer to the PDVF membrane, the samples were incubated with V5 antibody conjugated with HRP, and detected as luminescence bands using the ECL detection system.

3.1.2.9 FACS sorting

The transfected HS401 ES-cells were subjected to Fluorescence Activated Cell Sorting (FACS) at 9-11 h after transfection with the homeodomain genes (Papers III and IV). This was necessary due to the low number of transfected cells, and the possibility to compare transfected and non-transfected cells from the same culture well. The pFastBac vector coded for eGFP that provided green fluorescence signal in the transfected cells. Therefore, the cells were sorted according to the green signal for GFP positive cells. 75 GFP positive or 75 GFP negative cells were sorted per reaction in the sequencing library.

3.2 COMPUTATIONAL PROCEDURES

3.2.1 Data analysis

3.2.1.1 Microarray analysis

The publicly available microarray raw data was downloaded, normalized and quantified. The values were normalized independently, and then re-scaled to the same median intensity between all the samples to obtain the gene expression measures. (Paper I)

3.2.1.2 Homologue search in Biomart

Human gene names were used as input for gaining corresponding homologue Ensembl Gene ID-s for mouse. This list of mouse homologue Ensembl Gene ID-s was further used as an input for obtaining the corresponding associated gene names in Paper I.

3.2.1.3 qPCR analysis

Analysis of TaqMan Low Density Array Cards in Paper I was performed by RQ manager version 1.2.2 and DataAssist Software version 3.0. The Ct value of 40 was used for undetected assays. Expression value of qPCR data for Papers III and IV was done according to the following equation: y= Ct(*GAPDH*) – Ct (gene) + 25

3.2.1.4 Analysis of the STRT sequencing data

The pooled reads were assigned to the well of origin by demultiplexing and trimming barcodes from the high-quality reads. Identical reads with same UMI were merged into one if applicable. The reads were aligned to the human genome hg19 and further assembled in order

to obtain dataset-specific identifiers called Transcript 5'-Far Ends (TFEs) that were used for quantification. The TFEs were further annotated to the genes and to various regions around the gene. The quality control for exclusion/inclusion of samples was based on the sequencing depth, number of reads on spike-in molecules, ratio of reads on the first 50 bp of the spike-in molecules, low mapping rates, hierarchical clustering of samples and principal component analysis. Analysis was performed for human oocytes and embryos in Paper II, and for 8-cell blastomeres and hES-cells HS980 and HS983a in Papers III-IV.

3.2.1.5 Prediction of novel transcripts and primer design

The novel transcript start sites were predicted based on the detected TFEs close to the homeodomain genes based on the RNA-seq data from Paper II. The primers were designed prior to the putative start site, and in the putative 3' UTR.

3.2.1.6 Promoter analysis

MEME software was used for the *de novo* motif discovery (Paper II) and MAST software was used for aligning already known motifs to the sequence (Papers II-IV).

3.2.2 Statistical analysis

The differential expression of microarray probesets between the different developmental stages was performed by the Limma package. FDR values for equal to or less than 0.05 were considered significant. (Paper I)
The analysis of publicly available RNA sequencing datasets in Paper I was done by pairwise comparisons and genes with p-val < 0.05 and at least 5-fold differences in expression levels were considered significantly differentially expressed.
The differential expression of TFEs were analyzed by the SAMstrt R package. Significant TFEs for the overexpression studies in ES-cells were used only if they were significant in comparisons with all of the three controls (q-val < 0.05).
The chi-squared test was performed in Paper IV for testing the significance of overlap between genesets.

3.3 ETHICAL STATEMENTS

The use of experimental animals and the study design was approved by the Animal Care Board (Jordbruksverket) under the ethical permits S137-10 and S167-11 in accordance with the Swedish Law.
The studies on human material were reviewed and approved by the ethics review boards in accordance to the laws in Switzerland and Sweden. Donations from patients were obtained with full informed consent under the ethical permits CE2161 (Switzerland), Dnr 2010/937-31/4 and 2012/1765-31/1 (Sweden).

4 RESULTS

4.1 PAPER I

The results of the study showed that there were differences in the gene expression between mouse and human among the genes up-regulated in early genome activation (EGA).

Dynamically expressed genes during human pre-implantation development were identified based on two published sets of expression microarrays. Three clusters were defined based on their dynamic gene expression in the course of human pre-implantation development from zygote to blastocyst: “Up”, “Up-down” and “Down”. The expression level of selected genes was either increasing from oocyte to blastocyst (“Up”), increasing by 4- or 8-cell stage embryo and decreasing by the blastocyst stage (“Up-down”) or only decreasing in the course of development with high expression in the oocytes (“Down”). The classification is shown schematically in Figure 4 along with available embryo stages from the two datasets. The genes in these categories were evaluated based on further criteria that would indicate potential importance in early development: specificity of expression in embryos or pluripotent tissues, function as a transcription factor, association with cancer, novelty and expression of the orthologue in mouse pre-implantation embryo. Seventy genes were selected and used for studying the gene expression in mouse by the qPCR-based method of TaqMan Low Density Array.

Most genes in the cluster “Down” showed similar expression pattern in mouse as in human while being maternally expressed in human oocytes and later down-regulated in the development. However, over half of the mouse orthologues of human genes in clusters “Up” and “Up-down” exhibited the behavior of the maternal genes as being highly expressed in oocytes, and later down-regulated in the course of development. This phenomenon was confirmed by genome-wide approach technique by comparing human and mouse expression microarray and RNA sequencing datasets.

Furthermore, genes from 4 families of interest were assayed, that were previously only known to be expressed in cancer and testis tissues. We demonstrated that they were highly expressed in human pre-implantation development.

4.2 PAPER II

The study described whole-genome transcriptome profiling in human embryos by single-cell sequencing and described for the first time the involvement of PRD-like homeodomain transcription factors and retrotransposon elements in the pre-implantation development.

Altogether 348 oocytes, zygotes and single blastomeres from two- to three-day embryos were subjected to RNA sequencing. The method differs from others by enabling the measurement of the start of the transcript (TFEs) accurately and thus enabling the discovery of the start sites of new genes. Furthermore, the use of spike-in molecules enabled for comparison of unequal RNA contribution in every cell, which was necessary for unevenly-sized blastomeres after cleavage events. Differential gene expression analysis showed up-regulation of total 161 genes between oocyte to 4- or 8-cell stage embryo, describing the first transcriptional events of the zygotic genome during embryonic genome activation (EGA). The DNA sequences around the start sites of the up-regulated genes were used for *de novo* identification of a common motif. Similar 35- and 36-bp motifs were identified that include a binding site of PRD-like homeodomain transcription factors, and often occur in the primate specific Alu retrotransposon elements. The motif proved to induce gene expression by the use of PRD-like homeodomain proteins. Therefore, the PRD-like homeobox genes were looked up in the embryo transcriptome. Altogether, 14 candidate transcription factors were identified with maternal, EGA or mixed expression pattern, and with various degrees of conservation. Seven PRD-like transcription factors were verified: *ARGFX*, *CPHX1*, *CPHX2*, *DPRX*, *DUXA*, *DUXB* and *LEUTX*. All of them except for *ARGFX* were cloned for the first time.

The finding of the novel motif, transcription factors and retrotransposon elements suggested a species-specific gene regulation of the very early events in pre-implantation development.

4.3 PAPER III

This paper described the identification of the novel variant, expression pattern and targets of *LEUTX*, a PRD-like homeobox gene. The results indicated a central role for *LEUTX* as the activator of human embryo genome.

The previous transcriptome analysis in human pre-implantation embryos revealed expression of *LEUTX* gene in the EGA. The current study analyzes various publicly available datasets and shows that the expression of *LEUTX* is restricted to human pre-implantation embryos and to the pluripotent ESC at very low levels. The sequencing study in Paper II also identified a novel start-site for the *LEUTX* gene. The cloning of the gene from human 8-cell embryos revealed a novel variant of the gene with the predicted start-site that encoded a protein with a full homeodomain. This full homeodomain-containing *LEUTX* was overexpressed in hES-cells in order to study the targets induced by *LEUTX*.

Overexpression experiments identified the targets that overlap with EGA genes in human pre-implantation for two independent datasets. This showed that *LEUTX* was inducing the same genes that were activated in the early embryo. The further analysis around the start-sites of the target genes showed enrichment of the previously indicated embryo-specific motif. The activity via the motif was further demonstrated in luciferase reporter experiment that showed the induction of gene expression by *LEUTX*.

Together, these data showed the induction of EGA genes by *LEUTX* and suggested that the activity was mediated by binding of the motif predicted from early embryos.

4.4 PAPER IV

This study characterized the PRD-like homeodomain proteins and considered their role in the human early development.

The PRD-like homeodomain proteins were suggested to have a role in EGA of human early development. This study described the cloning of different variants of human PRD-like homeobox genes many of which lacked evidence of expression before. The cloning of these genes from human 8-cell embryo showed the expression of various novel variants for these genes including the proposed major isoforms with full homeodomain sequence. Our data and analysis of publicly available datasets indicated almost no expression of these genes in any other tissues except for human pre-implantation embryos. This indicated a function of these genes in human early development.

Target genes of the homeodomain proteins were studied by overexpressing the homeobox genes in the hES-cells. The results revealed three interesting findings. First, the homeodomain proteins act either as activators or repressors of gene expression. Second, the overlap of target genes of PRD-like genes with genes expressed in pre-implantation embryos was higher than expected by chance. This indicated the role of the PRD-like genes as regulators of the transcriptional events in the early embryo. Third, the overlap of targets among the homeodomain proteins themselves was higher than expected. The hierarchical clustering of the PRD-like genes based on this similarity of targets allowed for analysis of the homeodomain protein sequence. The analysis showed that the genes with similar targets also had similar residues in the variable positions of the DNA-binding sequence of the homeodomain.

In conclusion, the PRD-like homeodomain proteins might have a role in regulating the pre-implantation development of human pre-implantation development. They can act either as activators or repressors, and share a number of target genes that indicate variable residues in the DNA-binding domain responsible for target gene specificity.

5 DISCUSSION

5.1 SPECIES-SPECIFICITY OF EARLY DEVELOPMENT

Paper I showed the differences in gene expression between human and mouse early development, suggesting that some maternal genes in mouse have orthologues that are EGA genes in human. The observation was also confirmed by a recent sequencing study, where the comparison of human 4-cell specific module had significant overlap with mouse oocyte and zygote modules (p-values 7×10^{-7} and 4×10^{-6} correspondingly) (Yan et al., 2013). These differences could be explained by the time scale differences in the early development. When human embryos develop into blastocyst within 6 days, it requires 4 days for mice. During this time, the similar morphological states need to be covered from first cellular cleavage events to the formation of the blastocyst. Therefore the genes up-regulated during EGA in humans might require maternal expression in mouse in order to have sufficient time for their function. In comparison, the bovine pre-implantation development is more similar in timing to humans. Also, the bovine embryo developmental stage specific gene expression sets were more conserved in human compared to mouse (Jiang et al., 2014).

The findings were further supported by methylation data that show lower methylation levels for mouse maternal genome in the early development. Epigenetic studies have shown differences of DNA methylation levels between mouse and human oocytes where the mouse oocyte was de-methylated to a larger extent (Okae et al., 2014; Smith et al., 2014), suggesting accessibility for quicker gene activation directly after fertilization.

Overall, these findings supported the conclusions in Paper I that show maternal expression pattern for many genes that were expressed only at EGA in human development. Further differences in gene regulation between human and mouse were indicated in Paper II that suggested a fundamental role of primate-specific elements in pre-implantation development.

5.2 ROLE OF ALU ELEMENTS

Paper II suggested the involvement of primate-specific Alu repeats in the human pre-implantation development. The associated Alu elements were found to be enriched upstream of the up-regulated genes. This might indicate the function of Alu elements as regulators of gene expression. Alu elements have been suggested to act as enhancers (Su et al., 2014), that was supported by our findings. The hypomethylation of Alu sequences is needed for efficient transcription (Kochanek et al., 1993). The elements are heavily methylated in somatic tissues, but decrease in methylation in the human sperm (Hellmann-Blumberg et al., 1993; Kochanek et al., 1993). However, this methylation might not be enough, since the global de-methylation of the paternal genome in blastocysts showed that the SINE-VNTR-Alu elements and some other tandem-repeat containing regions were not de-methylated (Okae et al., 2014). It would be interesting to study whether the transcription of genes enabled by Alu elements could be specific to maternal or paternal genome and show the signs of imprinting in the EGA stage.

This could be studied by single-cell or single-embryo whole-genome methylation studies that can distinguish between maternal and paternal genomes.

5.3 ROLE OF THE PRD-LIKE HOMEODOMAIN GENES

The homeodomain-containing proteins are known to be fundamentally important development-specific. Involvement of the homeodomain proteins has been shown in the formation of body axes during later stages of development. For the first time, the studies in the current thesis associated the role of many PRD-like homeodomain proteins with very early events in the human development. Surprisingly, many of these genes appeared recently evolved (Paper II), which contrasts the understanding that developmentally important genes are conserved, since the morphology of the embryos is conserved in mammals (Figure 2). However, the many differences between species as close as human and mouse support the hypothesis of species-specific control mechanisms for early development.

The main novel finding in addition to implication of the Alu elements, was the transcriptional activity and function of many PRD-like homeodomain genes. Most of the genes were previously only annotated based on genomic predictions with no expression evidence, and with incomplete isoform detection. This was in accordance with our studies where we showed that the PRD-like homeodomain genes were expressed exclusively in early embryos, and barely detected in the ES-cells. Furthermore, the genes were found expressed in publicly available RNA-seq datasets in pre-implantation embryos (Xue et al., 2013; Yan et al., 2013), but in no other studied datasets. Overall, our data suggested and other data confirmed the expression of the PRD-like domain genes to be specific to the early development and pluripotent cells. Importantly, these factors included genes that up-regulated essential genes for pluripotency and ES-cells: *DPPA3* was up-regulated by *LEUTX*, *CPHX1*, *CPHX2*, *TPRX* and *OTX2*, and *NANOG* was up-regulated by *LEUTX* (Papers III and IV). This indicated that the homeodomain proteins might contribute to the lineage of ICM cells, and could be needed for the induction of the pluripotent profile of the ES-cells. In that case, the expression of specific homeodomain proteins in early embryos could be the first indication of lineage specification in the human embryo.

6 FUTURE PERSPECTIVES

The whole-genome analysis methods have provided multiple different datasets in human preimplantation development: expression microarrays, RNA-sequencing, bisulfite sequencing, whole-genome sequencing. The overlaps of these datasets have been analyzed in various contexts, but not comprehensively and specifically for human preimplantation development period. The combination and analysis of all currently available data might confirm earlier findings, and find common and reliable trends in gene expression and regulation of the early human embryo. However, more novel methods are already available or under development that could have low detection limits. These methods should be applied to the sparsely available material of human oocytes and embryos, when reaching to the detection levels of a single cell.

6.2 WHOLE GENOME-WIDE METHODS

Currently available whole genome approaches enable studies of the transcriptome by various RNA-seq methods designed for single-cells. However, every method has its benefits and disadvantages. Current study uses STRT RNA-seq that only maps start-sites of the genes, while many other methods are biased towards the end of the transcript. Strand-specificity is beneficial, while providing possibility for transcript isoform-mapping. Therefore, the third-generation sequencing methods with single-molecule detection and longer read length are desirable (Schadt et al., 2010). This would enable the detection of specific isoforms and novel gene variants in an amplification-bias free manner.

A step further, whole genome DNA sequencing from single cells is now also available. Double-stranded breaks in DNA occur during early development (Bohrer et al., 2015), and even cytogenetic analyses have shown chromosomal abnormalities including mosaicism in 23% of the early human embryos (Plachot et al., 1987). The mutation rate, frequency and variation between blastomeres in the early embryo is not known, however understanding of these aspects can help us better explain the behavior of the IVF embryos, and possibly generate hypotheses to assess their culture conditions. Therefore, single-cell genome sequencing on the early embryos might provide better knowledge of the pre-implantation period.

Novel methods for whole genome mapping of RNA and DNA from the same cells have been developed (Dey et al., 2015; Macaulay et al., 2015; Mertes et al., 2015). The application of these methods is especially essential for samples where the number of cells is limited and every cell is different, such as early development and differentiation. The application of these methods in single cells of early embryos would enable for direct comparison of transcription patterns with the DNA variants, and provide hypotheses for causality of observed abnormalities in gene expression.

Whole genome methylation studies from very small material have shown insights for dynamic methylation patterns in pre-implantation embryos (Okae et al., 2014; Smith et al.,

2014). Improvement of these methods should give us higher coverage of the CpG-s and enable mapping to either maternal or paternal genome. This will allow for studies of imprinting of the very early events in the pre-implantation embryo.

There are still no methods for the detection of histone modifications genome-wide in single cells. Chip-seq methods enable to map certain histone modifications to the genome, but single-cell resolution is now only achieved with very low coverage (Rotem et al., 2015). DNA accessibility can also be evaluated by Assay for Transposase-Accesible Chromatin high-throughput sequencing (ATAC-seq), which has been demonstrated on single-cells (Buenrostro et al., 2015). Applying these methods on human embryos in addition to RNA- and DNA sequencing would allow for comprehensive whole genome analysis on human pre-implantation development.

6.3 PRD-LIKE HOMEODOMAIN PROTEINS

In order to pinpoint the role of each PRD-like homeodomain protein in early development, more functional studies are needed. The development of rapid and accurate gene editing techniques allows for knock-out of a specific gene in the early human embryo (Liang et al., 2015). The studies with the knock-out human embryos for the PRD-like homeodomain genes would demonstrate their essential roles for human embryo development. However, due to the similarity of target genes, these proteins could be redundant in function, which means that they are able to compensate for the loss of each other. Furthermore, these studies are complicated due to ethical considerations concerning the human embryos.

The role of PRD-like homeobox genes could be further studied in pluripotent cells. These factors induce expression of genes that are essential for pluripotency such as *NANOG* as target of *CPHX1* (Paper IV) that has the capacity to induce cells from differentiated towards induced pluripotent stem cells (iPS-cells) (Takahashi and Yamanaka, 2006). This might indicate the role of the PRD-like homeodomain proteins in induction of pluripotency and could therefore be used as additional factors for iPS-cell generation. The findings on this field will introduce novel methods influencing the areas of pluripotency, regenerative medicine and reprogramming, which have become more and more important in medicine. The importance of this field was noted recently by Nobel Prize in Physiology or Medicine 2012 to Sir John B. Gurdon and Shinya Yamanaka for the discovery that mature cells can be reprogrammed to become pluripotent (Nobelprize.org, 2015). The homeodomain proteins described in the current study might be key players in this fundamentally important process.

7 ACKNOWLEDGEMENTS

It is hard to overestimate the role of my main supervisor, **Prof. Juha Kere**. I would not be here without him and the decision he made when he picked me out from the sea of many other PhD seekers. He has always been an inspirational leader with fresh ideas and motivating projects. I will not forget how Juha draws the big picture by looking at the data to very small detail. Juha's contribution goes well beyond science. He possesses a rare combination of traits including exceptional kindness and scientific excellence, which sets an example to every PhD student, post-doc and colleague. I have been extremely lucky to work together with such a genuinely good person, who has inspired me in every meeting we have had. I truly hope that we will keep discussing science and life during the long years ahead.

My co-supervisors **Virpi** and **Lisa** have been supportive throughout the years. Virpi is always in a contagiously happy mood that makes everybody else feel good around her. In the same time, her comments have been helpful and constructive, as well as the discussions for which she always finds time in her very busy schedule. Lisa has taught me to be critical in science, and always in doubt of not only mine, but also other people's claims. For that lesson I am really grateful, and also for all the casual chatting and jokes in the lab.

Shintaro, my co-supervisor sets a new standard in scientific analysis. There is so much more that I could learn from him, and I am thankful for what I have learned so far. One day I hope to be able to do what he does with the data from writing the programs to the interpretation of the data and suggestion of experimental follow-up to support the hypothesis.

The thesis work would not have been completed without a list of close collaborators. **Outi Hovatta** has been essential in initiating and continuing the human infertility studies, **Fredrik Lanner** is an inspirational scientist with new ideas in studying human infertility, **Sten Linnarsson** has opened the world of single-cell RNA-seq for our research and **Thomas Bürglin** has supported and educated us in the area of homeobox genes. **Per Unneberg** taught the analysis of microarrays and **Jose Inzunza** taught the mouse model organism and embryo handling techniques. I want to thank **Robert Månsson** for the long hours and good company at FACS, **Kaarel Krjutškov** for STRT RNA-seq analysis and educational lectures in reproduction, **Mona** for her excellent skills in embryo manipulation and **Alvaro** for really pleasant collaboration.

I want to thank **Charles Plessy**, my short-term supervisor during the exchange months in Japan. He has shown me yet another way of leadership and support, as well as taught important aspects of data analysis.

I'm very glad to have worked with **Pauliina** and gotten to know her. I hope one day to be able to follow the example she has set in working and managing many different projects simultaneously while thriving on personal matters in the same time.

Eeva-Mari has been a close colleague and a friend. I will miss the talks, the walks, the runs and especially picking all various edible stuff that we could find around the campus.

There are many people in the lab that helped out with the experiments, analysis and other questions. We have worked hand-in-hand and commented each-others work on daily basis while creating a friendly scientific environment for us all. I want to thank you for these years: **Ingegerd**, **Jingwen**, **Nathalie**, **Myriam**, **Tiina**, **Helena**, **Linda**, **Suvi**, **Aparna**, **Maria**, **Nancy**, **Morana**, **Hans**, **Cilla**, **Lovisa**, **Kristiina**, **Amitha**, **Debora**, **Isabel**, **Mariann**, **Andrea**, **Elisabet**, **Gayathri** and **Ettore**. Some more people have contributed to the lively environment in the past years by sharing thoughts in the lunch-room, scientific seminars and/or leisure time: **Sylvie**, **Jade**, **Minna**, **Sergey**, **Carsten**, **Deffiz**, **Olga**, **Tomas**, **Marcela**, **Gabor**, **Lois**, **Michael**, **JP**, **Sarita**, **Patrick**, **Fredrik, Milica, Arttu**, **Eevi**, **Neivis**, **John**, **Yongtao**, **Emma** and **Johan**.

The years in Stockholm were very educating in other aspects beside science. The local Estonian community has played a fundamental role in my life in here. A few seasons with the Estonian choir introduced the local community and culture to me. Special thanks to **Toomas**, **Henry** and **Hando**. Past three years have been amazing with the folk dance group **Virvel**. This has been a dream that turned into reality, for which I am very grateful to all its members. **Ingrid** has played an essential part in building up this dream, and has been a true friend. **Taavi** has supported me throughout different activities.

My undergraduate teachers and supervisors in the lab have been inspirational. They introduced me to science in general and to the area of stem cells that led me to the PhD. Thank you for that: **Martin**, **Mario**, **Sulev** and **Toivo**.

I gain a lot of inspiration and energy from people around me. I experience thorough discussion, culture and always full cup of friendship from the members of **Amicitia** and **ETV Girls Choir**. I always fill up my energy levels when meeting you: **Janika**, **Margus**, **Kati**, **Riin**, **Liis**, **Kaarin**, **Argo**, **Erkki**, **Kaspar**, **Ülla-Karin**, **Angela**, **Kaie** and **Mari**. Thank you for being there for me throughout the years! Thank you so much, **Kaie**, for introducing me to the wonderful crowd of inspiring young Estonian scientists including **Jaan**, **Juhan**, **Kaur**, **Kristjan** and **Leopold**. Thank you for accepting me to your group, introducing your ideas and views, and for influencing my life beyond science.

My family has shaped me as a person, so that I can be who I am today. My sisters **Lili** and **Maili** motivate me constantly by setting an even higher standard every time I see them. I would not want to have it any other way! My little sister **Lee Ann** has been my greatest scientific fan, and a mirror of my own upbringing. My parents **Aili** and **Ülo** have created safe and loving environment for their family. They surprise me over and over again with portions of attention and support. Independent of my choices, success, mood or well-being, I know they will always be there for me. I could not thank them enough.

I want to thank all the **anonymous couples** undergoing IVF who participated in the study via donation of oocytes and embryos. I sincerely hope that the treatment has helped your families. I want to acknowledge the funding agencies who have made this research possible. First, many thanks to **Karolinska Institutet** for awarding the **Distinguished Professor Award** to Prof. Juha Kere that was used for funding my studies and the experiments. The work was further supported via grants from **Swedish Research Council**. Part of this work was assisted by **Karolinska High Throughput Center**, a core facility at Karolinska Institutet with affiliation to SciLifeLab (https://www.scilifelab.se/facilities/khtc). The computations were performed on resources provided by SNIC through **Uppsala Multidisciplinaty Center for Advanced Computational Science** (UPPMAX). This study was in part performed at the **Live Cell Imaging unit**, Department of Biosciences and Nutrition, Karolinska Institutet, Sweden, supported by grants from the Knut and Alice Wallenberg Foundation, the Swedish Research Council, the Centre for Innovative Medicine and the Jonasson donation to the School of Technology and Health, Royal Institute of Technology, Sweden.

Finally, I would like to thank **the Knut and Alice Wallenberg Foundation** for the continuing support to Prof. Juha Kere for understanding the molecular mechanisms of early development. I wish all of us good luck in the continuous path of research.

8 REFERENCES

Alesi, R. (2005). Infertility and its treatment--an emotional roller coaster. Aust Fam Physician *34*, 135-138.

Barbulescu, M., Turner, G., Seaman, M.I., Deinard, A.S., Kidd, K.K., and Lenz, J. (1999). Many human endogenous retrovirus K (HERV-K) proviruses are unique to humans. Curr Biol *9*, 861-868.

Bell, C.E., Calder, M.D., and Watson, A.J. (2008). Genomic RNA profiling and the programme controlling preimplantation mammalian development. Mol Hum Reprod *14*, 691-701.

Benard, C., McCright, B., Zhang, Y., Felkai, S., Lakowski, B., and Hekimi, S. (2001). The C. elegans maternal-effect gene clk-2 is essential for embryonic development, encodes a protein homologous to yeast Tel2p and affects telomere length. Development *128*, 4045-4055.

Benedek, T. (1980). Parenthood, Its Psychology & Psychopathology (Boston, Little Brown & Co.).

Biase, F.H., Cao, X., and Zhong, S. (2014). Cell fate inclination within 2-cell and 4-cell mouse embryos revealed by single-cell RNA sequencing. Genome research *24*, 1787-1796.

Blakeley, P., Fogarty, N.M., Del Valle, I., Wamaitha, S.E., Hu, T.X., Elder, K., Snell, P., Christie, L., Robson, P., and Niakan, K.K. (2015). Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. Development *142*, 3151-3165.

Bohrer, R.C., Coutinho, A.R., Duggavathi, R., and Bordignon, V. (2015). The Incidence of DNA Double-Strand Breaks Is Higher in Late-Cleaving and Less Developmentally Competent Porcine Embryos. Biol Reprod *93*, 59.

Booth, H.A., and Holland, P.W. (2007). Annotation, nomenclature and evolution of four novel homeobox genes expressed in the human germ line. Gene *387*, 7-14.

Braude, P., Bolton, V., and Moore, S. (1988). Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. Nature *332*, 459-461.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol *109*, 21 29 21-29.

Burglin, T.R., and Affolter, M. (2015). Homeodomain proteins: an update. Chromosoma.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell *113*, 643-655.

Cockburn, K., and Rossant, J. (2010). Making the blastocyst: lessons from the mouse. J Clin Invest *120*, 995-1003.

D'Elia, A.V., Tell, G., Paron, I., Pellizzari, L., Lonigro, R., and Damante, G. (2001). Missense mutations of human homeoboxes: A review. Hum Mutat *18*, 361-374.

de Mendoza, A., Sebe-Pedros, A., Sestak, M.S., Matejcic, M., Torruella, G., Domazet-Loso, T., and Ruiz-Trillo, I. (2013). Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. Proceedings of the National Academy of Sciences of the United States of America *110*, E4858-4866.

Dey, S.S., Kester, L., Spanjaard, B., Bienko, M., and van Oudenaarden, A. (2015). Integrated genome and transcriptome sequencing of the same cell. Nat Biotechnol *33*, 285-289.

Dobson, A.T., Raja, R., Abeyta, M.J., Taylor, T., Shen, S., Haqq, C., and Pera, R.A. (2004). The unique transcriptome through day 3 of human preimplantation development. Hum Mol Genet *13*, 1461-1470.

Edwards, R.G. (2001). The bumpy road to human in vitro fertilization. Nat Med *7*, 1091-1094.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154-156.

Flach, G., Johnson, M.H., Braude, P.R., Taylor, R.A., and Bolton, V.N. (1982). The transition from maternal to embryonic control in the 2-cell mouse embryo. Embo J *1*, 681-686.

Fulka, H., St John, J.C., Fulka, J., and Hozak, P. (2008). Chromatin in early mammalian embryos: achieving the pluripotent state. Differentiation *76*, 3-14.

Galan, A., Montaner, D., Poo, M.E., Valbuena, D., Ruiz, V., Aguilar, C., Dopazo, J., and Simon, C. (2010). Functional genomics of 5- to 8-cell stage human embryos by blastomere single-cell cDNA analysis. PLoS One *5*, e13615.

Gilbert, S.F. (2000). Developmental Biology, 6th Edition (Sunderland (MA), Sinauer Associates).

Ginis, I., Luo, Y., Miura, T., Thies, S., Brandenberger, R., Gerecht-Nir, S., Amit, M., Hoke, A., Carpenter, M.K., Itskovitz-Eldor, J., *et al.* (2004). Differences between human and mouse embryonic stem cells. Dev Biol *269*, 360-380.

Gkountela, S., Zhang, K.X., Shafiq, T.A., Liao, W.W., Hargan-Calvopina, J., Chen, P.Y., and Clark, A.T. (2015). DNA Demethylation Dynamics in the Human Prenatal Germline. Cell *161*, 1425-1436.

Greenblatt, R.B., Barfield, W.E., Jungck, E.C., and Ray, A.W. (1961). Induction of ovulation with MRL/41. Preliminary report. Jama *178*, 101-104.

Grow, E.J., Flynn, R.A., Chavez, S.L., Bayless, N.L., Wossidlo, M., Wesche, D.J., Martin, L., Ware, C.B., Blish, C.A., Chang, H.Y., *et al.* (2015). Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature *522*, 221-225.

Hamatani, T., Carter, M.G., Sharov, A.A., and Ko, M.S. (2004). Dynamics of global gene expression changes during mouse preimplantation development. Dev Cell *6*, 117-131.

Handyside, A.H., Lesko, J.G., Tarin, J.J., Winston, R.M., and Hughes, M.R. (1992). Birth of a normal girl after in vitro fertilization and preimplantation diagnostic testing for cystic fibrosis. N Engl J Med *327*, 905-909.

Hellmann-Blumberg, U., Hintz, M.F., Gatewood, J.M., and Schmid, C.W. (1993). Developmental differences in methylation of human Alu repeats. Mol Cell Biol *13*, 4523-4530.

Heyn, P., Kircher, M., Dahl, A., Kelso, J., Tomancak, P., Kalinka, A.T., and Neugebauer, K.M. (2014). The earliest transcribed zygotic genes are short, newly evolved, and different across species. Cell Rep *6*, 285-292.

Holland, P.W., Booth, H.A., and Bruford, E.A. (2007). Classification and nomenclature of all human homeobox genes. BMC Biol *5*, 47.

Howlett, S.K., and Reik, W. (1991). Methylation levels of maternal and paternal genomes during preimplantation development. Development *113*, 119-127.

Hughes, C.L., and Kaufman, T.C. (2002). Hox genes and the evolution of the arthropod body plan. Evol Dev *4*, 459-499.

Islam, S., Kjallquist, U., Moliner, A., Zajac, P., Fan, J.B., Lonnerberg, P., and Linnarsson, S. (2012). Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. Nat Protoc *7*, 813-828.

Jiang, Z., Sun, J., Dong, H., Luo, O., Zheng, X., Obergfell, C., Tang, Y., Bi, J., O'Neill, R., Ruan, Y., *et al.* (2014). Transcriptional profiles of bovine in vivo pre-implantation development. BMC Genomics *15*, 756.

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., *et al.* (2013). DNA-binding specificities of human transcription factors. Cell *152*, 327-339.

Kivioja, T., Vaharautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S., and Taipale, J. (2012). Counting absolute numbers of molecules using unique molecular identifiers. Nat Methods *9*, 72-74.

Klonoff-Cohen, H., Chu, E., Natarajan, L., and Sieber, W. (2001). A prospective study of stress among women undergoing in vitro fertilization or gamete intrafallopian transfer. Fertility and sterility *76*, 675-687.

Kocabas, A.M., Crosby, J., Ross, P.J., Otu, H.H., Beyhan, Z., Can, H., Tam, W.L., Rosa, G.J., Halgren, R.G., Lim, B., *et al.* (2006). The transcriptome of human oocytes. Proc Natl Acad Sci U S A *103*, 14027-14032.

Kochanek, S., Renz, D., and Doerfler, W. (1993). DNA methylation in the Alu sequences of diploid and haploid primary human cells. Embo J *12*, 1141-1151.

Kupka, M.S., Ferraretti, A.P., de Mouzon, J., Erb, K., D'Hooghe, T., Castilla, J.A., Calhaz-Jorge, C., De Geyter, C., and Goossens, V. (2014). Assisted reproductive technology in Europe, 2010: results generated from European registers by ESHREdagger. Human reproduction *29*, 2099-2113.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Liang, P., Xu, Y., Zhang, X., Ding, C., Huang, R., Zhang, Z., Lv, J., Xie, X., Chen, Y., Li, Y., *et al.* (2015). CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. Protein Cell *6*, 363-372.

Macaulay, I.C., Haerty, W., Kumar, P., Li, Y.I., Hu, T.X., Teng, M.J., Goolam, M., Saurat, N., Coupland, P., Shirley, L.M., *et al.* (2015). G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat Methods *12*, 519-522.

Madissoon, E., Tohonen, V., Vesterlund, L., Katayama, S., Unneberg, P., Inzunza, J., Hovatta, O., and Kere, J. (2014). Differences in gene expression between mouse and human for dynamically regulated genes in early embryo. PLoS One *9*, e102949.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proceedings of the National Academy of Sciences of the United States of America *78*, 7634-7638.

Mascarenhas, M.N., Flaxman, S.R., Boerma, T., Vanderpoel, S., and Stevens, G.A. (2012). National, regional, and global trends in infertility prevalence since 1990: a systematic analysis of 277 health surveys. PLoS Med *9*, e1001356.

Mathelier, A., Zhao, X., Zhang, A.W., Parcy, F., Worsley-Hunt, R., Arenillas, D.J., Buchman, S., Chen, C.Y., Chou, A., Ienasescu, H., *et al.* (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res *42*, D142-147.

Mertes, F., Lichtner, B., Kuhl, H., Blattner, M., Otte, J., Wruck, W., Timmermann, B., Lehrach, H., and Adjaye, J. (2015). Combined ultra-low input mRNA and whole-genome sequencing of human embryonic stem cells. BMC Genomics *16*, 925.

Metchat, A., Akerfelt, M., Bierkamp, C., Delsinne, V., Sistonen, L., Alexandre, H., and Christians, E.S. (2009). Mammalian heat shock factor 1 is essential for oocyte meiosis and directly regulates Hsp90alpha expression. J Biol Chem *284*, 9521-9528.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631-642.

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

Niakan, K.K., Han, J., Pedersen, R.A., Simon, C., and Pera, R.A. (2012). Human pre-implantation embryo development. Development *139*, 829-841.

Nobelprize.org (2010). The Nobel Prize in Physiology or Medicine 2010. In Nobel Media AB (Stockholm).

Nobelprize.org (2015). The Nobel Prize in Physiology or Medicine 2012. In Nobel Media AB.

Oddens, B.J., den Tonkelaar, I., and Nieuwenhuyse, H. (1999). Psychosocial experiences in women facing fertility problems--a comparative survey. Human reproduction *14*, 255-261.

Okae, H., Chiba, H., Hiura, H., Hamada, H., Sato, A., Utsunomiya, T., Kikuchi, H., Yoshida, H., Tanaka, A., Suyama, M., *et al.* (2014). Genome-wide analysis of DNA methylation dynamics during early human development. PLoS Genet *10*, e1004868.

Palermo, G., Joris, H., Devroey, P., and Van Steirteghem, A.C. (1992). Pregnancies after intracytoplasmic injection of single spermatozoon into an oocyte. Lancet *340*, 17-18.

Pandian, Z., Akande, V.A., Harrild, K., and Bhattacharya, S. (2008). Surgery for tubal infertility. Cochrane Database Syst Rev, CD006415.

Plachot, M., Junca, A.M., Mandelbaum, J., de Grouchy, J., Salat-Baroux, J., and Cohen, J. (1987). Chromosome investigations in early life. II. Human preimplantation embryos. Human reproduction *2*, 29-35.

Polak, P., and Domany, E. (2006). Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. BMC Genomics *7*, 133.

Qian, Y.Q., Otting, G., Billeter, M., Muller, M., Gehring, W., and Wuthrich, K. (1993). Nuclear magnetic resonance spectroscopy of a DNA complex with the uniformly 13C-labeled Antennapedia homeodomain and structure determination of the DNA-bound homeodomain. J Mol Biol *234*, 1070-1083.

Rhead, B., Karolchik, D., Kuhn, R.M., Hinrichs, A.S., Zweig, A.S., Fujita, P.A., Diekhans, M., Smith, K.E., Rosenbloom, K.R., Raney, B.J., *et al.* (2010). The UCSC Genome Browser database: update 2010. Nucleic Acids Res *38*, D613-619.

Rotem, A., Ram, O., Shoresh, N., Sperling, R.A., Goren, A., Weitz, D.A., and Bernstein, B.E. (2015). Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. Nat Biotechnol *33*, 1165-1172.

Schadt, E.E., Turner, S., and Kasarskis, A. (2010). A window into third-generation sequencing. Hum Mol Genet *19*, R227-240.

Smallwood, S.A., and Kelsey, G. (2012). De novo DNA methylation: a germ cell perspective. Trends Genet *28*, 33-42.

Smith, Z.D., Chan, M.M., Humm, K.C., Karnik, R., Mekhoubad, S., Regev, A., Eggan, K., and Meissner, A. (2014). DNA methylation dynamics of the human preimplantation embryo. Nature *511*, 611-615.

Steptoe, P.C., and Edwards, R.G. (1978). Birth after the reimplantation of a human embryo. Lancet *2*, 366.

Su, M., Han, D., Boyd-Kirkup, J., Yu, X., and Han, J.D. (2014). Evolution of Alu elements toward enhancers. Cell Rep *7*, 376-385.

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

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

Tang, F., Barbacioru, C., Nordman, E., Li, B., Xu, N., Bashkirov, V.I., Lao, K., and Surani, M.A. (2010). RNA-Seq analysis to capture the transcriptome landscape of a single cell. Nat Protoc *5*, 516-535.

Tohonen, V., Katayama, S., Vesterlund, L., Jouhilahti, E.M., Sheikhi, M., Madissoon, E., Filippini-Cattaneo, G., Jaconi, M., Johnsson, A., Burglin, T.R., *et al.* (2015). Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development. Nat Commun *6*, 8207.

Tong, Z.B., Gold, L., Pfeifer, K.E., Dorward, H., Lee, E., Bondy, C.A., Dean, J., and Nelson, L.M. (2000). Mater, a maternal effect gene required for early embryonic development in mice. Nat Genet *26*, 267-268.

Tsirigos, A., and Rigoutsos, I. (2009). Alu and b1 repeats have been selectively retained in the upstream and intronic regions of genes of specific functional classes. PLoS Comput Biol *5*, e1000610.

Wang, H., and Dey, S.K. (2006). Roadmap to embryo implantation: clues from mouse models. Nat Rev Genet *7*, 185-199.

Wang, Q.T., Piotrowska, K., Ciemerych, M.A., Milenkovic, L., Scott, M.P., Davis, R.W., and Zernicka-Goetz, M. (2004). A genome-wide study of gene activity reveals developmental signaling pathways in the preimplantation mouse embryo. Dev Cell *6*, 133-144.

Xie, D., Chen, C.C., Ptaszek, L.M., Xiao, S., Cao, X., Fang, F., Ng, H.H., Lewin, H.A., Cowan, C., and Zhong, S. (2010). Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. Genome Res.

Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C.Y., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y.E., *et al.* (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. Nature.

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., *et al.* (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat Struct Mol Biol.

Zeng, F., Baldwin, D.A., and Schultz, R.M. (2004). Transcript profiling during preimplantation mouse development. Dev Biol *272*, 483-496.

Zernicka-Goetz, M., Morris, S.A., and Bruce, A.W. (2009). Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. Nat Rev Genet *10*, 467-477.

Zhang, P., Dixon, M., Zucchelli, M., Hambiliki, F., Levkov, L., Hovatta, O., and Kere, J. (2008). Expression analysis of the NLRP gene family suggests a role in human preimplantation development. PLoS One *3*, e2755.

Zhang, P., Kerkela, E., Skottman, H., Levkov, L., Kivinen, K., Lahesmaa, R., Hovatta, O., and Kere, J. (2007). Distinct sets of developmentally regulated genes that are expressed by human oocytes and human embryonic stem cells. Fertil Steril *87*, 677-690.

Zhang, P., Zucchelli, M., Bruce, S., Hambiliki, F., Stavreus-Evers, A., Levkov, L., Skottman, H., Kerkela, E., Kere, J., and Hovatta, O. (2009). Transcriptome profiling of human pre-implantation development. PLoS One *4*, e7844.