**Univerzita Karlova v Praze**

**Přírodovědecká fakulta**

**Studijní program: Biologie**
**Studijní obor: Buněčná a vývojová biologie**


Bc. Martin Moravec

# Analýza pluripotentního programu genové exprese v časných embryích a embryonálních kmenových buňkách

# Analysis of pluripotent gene expression program in early embryos and embryonic stem cells

Diplomová práce

Školitel: Mgr. Petr Svoboda, PhD

Praha 2012

**Prohlášení**

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

..... .....

Datum/Date Signature/Podpis

Acknowledgement

I would like to thank to all my mentors and teachers throughout my early scientific career, for their support, belief and their guidance. Especially, I would like to devote my thanks to Matyáš Flemr and Dr. Petr Svoboda and the whole team of laboratory of Epigenetic Regulations for friendly atmosphere. Also, my thanks belong to Vedran Franke, Kristian Vlahoviček and Sara Sumic for introducing me to the field of bioinformatics and for giving me a chance to spend a summer in Zagreb. My special gratitude belongs to Sara Sumic for sharing her expertise in Affymetrix microarrays. I would like to thank Vendula Rusňáková for collaboration and expertise in high-throughput experiments.

# Contents

# 1 Abstract

Pluripotency is an ability of a cell to differentiate into any cell type. It naturally forms during early mammalian development and it is accompanied by global reprogramming of gene expression. The process of natural pluripotency establishment remains poorly understood. To get further insights into this process, I studied gene expression changes during mouse oocyte-to-zygote transition. In this model system, the fertilized oocyte undergoes reprogramming resulting in formation of pluripotent blastomeres, which give a rise to the embryo. The goal of my thesis was to analyse transcriptional activation during early development and to develop a method for convenient monitoring of expression of numerous genes in oocytes, early embryos and embryonic stem cells. The method employs high-throughput quantitative real-time PCR and allows for measuring expression of 48 genes, which serve as markers for maternal mRNA degradation, activation of the pluripotent program, and differentiation into germ lineages. I show that the assay allows for monitoring transcriptome dynamics during oocyte-to-zygote transition and generates data comparable with microaray platforms. In addition, our bioinformatic screening identified novel oocyte-specific and zygotic non-coding RNAs.

**Key words:** pluripotency, oocyte-to-zygote transition, embryonic stem cells, real-time PCR, oocyte, zygote, zygotic genome activation

# 2 Abstrakt

Pluripotence je schopnost buňky diferencovat do jakéhokoliv buněčného typu. Formuje se během časného embryonálního vývoje u savců a její vznik je spojen s reprogramací genové exprese na globální úrovni. Proces přirozeného vzniku pluripotence není stále zcela pochopen. Pro získání nového pohledu na události, které vedou ke vzniku pluripotence u savců, studovali jsme změny v genové expresi během oocyt-zygotického přechodu u myši. V tomto modelovém systému, oplodněné vajíčko podstoupí reprogramaci, která vede k vytvoření pluripotentních blastomer. Tyto blastomery zakládají samotné embryo. Cílem mé diplomové práce bylo analyzovat aktivaci transkripce během časného vývoje a vyvinout metodu pro monitorování exprese genů v oocytech, časných embryích a embryonálních kmenových buňkách. Metoda využívá kvantitativní PCR a umožnuje změřit expresi až 48 vybraných genů, které slouží jako markery pro maternální degradaci, aktivaci pluripotentního programu a diferenciaci do zárodečných linií. Dále ukazujeme, že náš systém monitoruje dynamiku transkriptomu během oocyt-zygotického přechodu, a získané výsledky jsou srovnatelné s daty naměřenými pomocí jiných metod. Díky našemu bioinformatickému přístupu jsme navíc identifikovali nové oocyt-specifické a zygotické nekódující RNA.

**Klíčová slova:** pluripotence, oocyt-zygotický přechod, embryonální kmenové buňky, real-time PCR, oocyt, zygota, aktivace genomu

## Aims of the thesis

The main aim of the thesis was to get new insights into the zygotic genome activation and establishment of pluripotency during early development. My research involved the following tasks:

- Identification of novel oocyte-specific and zygotic non-coding RNAs
- Development a quantitative PCR array system for single-cell/single embryo phenotyping during oocyte-to-zygote transition
- Development a quantitative PCR assay for monitoring of lineage commitment and pluripotency in embryonic stem cells

# 3 List of abbrevations

| | |
|---|---|
| 1C | one-cell |
| 2C | two-cell |
| 8C | eight-cell |
| ATRA | all-trans retinoic acid |
| BrUTP | brome uridine triphosphate |
| cDNA | complementary DNA |
| DMEM | Dubelco modified medium |
| dNTP | deoxyribonucleotides |
| dpc | days post-coitum |
| EGF | epidermal growth factor |
| ESCs | embryonic stem cells |
| FCS | fetal calf serum |
| FSH | follicule stimulating hormone |
| GV | germinal vesicle |
| hCG | chorionic gonadotropin |
| IBMX | 3-isobutyl-1-methylxanthine |
| ICM | inner cell mass |
| IL | interleukin |
| iPSCs | induced pluripotent stemcells |
| LH | luteinisation hormone |
| LIF | leukemia inhibitory factor |
| lincRNA | long intergenic RNA |
| MEF | mouse embryonic fibroblasts |
| mESC | mouse embryonic stem cells |
| MII | metaphase II |
| miRNA | micro RNA |
| mRNA | messenger RNA |
| ncRNA | non-coding RNA |
| NGS | Next generation sequencing |
| OZT | oocyte-to-zygote transition |
| PCA | principal component analysis |
| PE | primitive endoderm |
| PGC | primordial germ cell |
| piRNAs | piwi interacting RNA |
| PMSG | pregnant mare serum gonadotropin |
| qPCR | quantitative PCR |
| RT-PCR | real-time polymerase chain reaction |
| SEM | standard error of the mean |
| shRNA | short-hairpin RNA |
| siRNA | short interfering RNA |
| TE | trophoectoderm |
| TGF-β | tumor growth factor |
| TSS | transcription start site |
| U | unit |
| ZGA | zygotic genome activation |

# 4 Theoretical background

During early mammalian development two forms of high developmental potential appear: totipotency and pluripotency. Totipotent cells have an ability to produce all differentiated cells in the entire organism including extraembryonic tissues. Pluripotent cells can differentiate into any cell type of the three germ layers of the embryo: endoderm, ectoderm and mesoderm. The actual word pluripotency comes from Latin, where *plurimus* stands for very many and *potents* means to have power. Pluripotency and totipotency are gained early in the mammalian development and their establishment are not fully understood.

Mouse preimplantation development is a valuable model for studying totipotency and pluripotency. Embryonic development starts upon fertilisation of the female gamete (oocyte), which transforms into the totipotent zygote. Further cleavage of the zygote will give rise to a fraction of pluripotent cells in the inner cell mass (ICM) of the blastocyst. Mammalian blastocyst is formed usually 3.5 days post-fertilisation (dpc) and consists of three distinct cell types: epiblast, primitive endoderm and trophoectoderm. The pluripotent epiblast gives rise to the three major embryonic germ layers and contributes to the body plan establishment, whereas primitive endoderm and trophoectoderm form extra-embryonic tissues.

The transition between fully differentiated mammalian oocyte and totipotent zygote, and the pluripotency establishment in the embryo are complex and encompass changes in gene expression, epigenetic modifications and activation of diverse signaling networks. In my thesis, I used high-throughput technologies and bioinformatic approaches to explore the earliest events of pluripotency formation during mouse oocyte-to-zygote transition (OZT) and in embryonic stem cells (ESCs).

## 4.1 Mammalian oogenesis

*Omne vivum ex ovo*, first postulated by William Harvey in 17$^{th}$ century, states that the oocyte is the most important cell that gives rise to everything living. Mammalian oocytes develop from primordial germ cells (PGCs), which are formed early in development. Soon after embryo implants in the uterus (dpc 7), PGCs are found between epiblast and allantois as a small population of cells. They are characterized by expression of several markers such as *Blimp1*, *Fragilis* (Ohinata et al, 2005; Saitou et al, 2003). PGCs subsequently migrate to the genital ridge and enter the meiosis, thus becoming primary oocytes. Primary oocytes in the mouse are arrested at prophase of the first meiotic division until receiving the stimuli to grow and resume meiosis. Secondary oocytes develop from the primary oocytes by completing the first meiotic division. In the mouse, secondary oocytes become, however, arrested at metaphase of the second meiotic division (designated by MII) until fertilization (Sorensen & Wassarman, 1976).

Development of mouse oocytes takes place in a spherical structure, called a follicle (Figure 1). Follicle cells surround the oocyte and control growth and resumption of meiosis in primary oocytes via secreting small signaling molecules. The transition from primary to secondary oocytes is controlled by hormones. Follicle-stimulating hormone (FSH) is a gonadotropin hormone, which plays an important role during oogenesis and folliculogenesis (Weil et al, 1999). FSH is secreted from hypothalamus after female reaches puberty and positively regulates synthesis of a receptor for luteinisation hormone (LH). It also promotes growing phase of the preantral follicle. In defined time periods, rapid increase of concentration of LH binds its receptors on mural granulosa cells and mediates expression of EGF-like molecules. EGF-like signaling has two major consequences on the preantral follicle. First, it promotes expansion of cumulus cells surrounding the oocyte and, second, it promotes resumption of meiosis of the oocyte [reviewed in (Edson et al, 2009)].

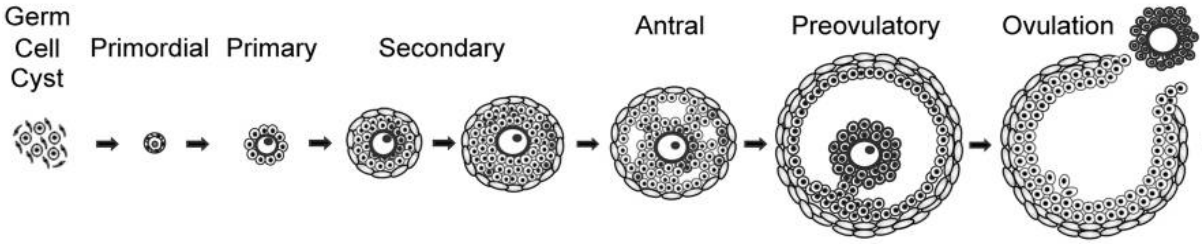


**Figure 1 Schematic representation of mammalian follicular development**

Growth and development of mammalian oocytes is tightly coupled with follicular development. During development, follicles change morphology and composition of cells from primordial germ cells to fully fertilization-competent cells and are accompanied by accumulation of mRNA molecules and transcription factors, which will support early events of embryo development. Up to the preovulatory follicle, the primary oocyte remains arrested at the prophase of the first meiotic division. Figure adopted from (Edson et al, 2009)

## 4.2 Mammalian transcriptome in oocytes and early embryos

Oocytes must accumulate factors sustaining zygotic genome activation and early development. Numerous factors, including proteins, RNAs are stored in the oocyte, which will be used at different stages of the oocyte-to-zygote transition [reviewed in (Stitzel & Seydoux, 2007)]. A specific storage mechanism is represented by so-called dormant maternal mRNAs. Dormant maternal mRNAs accumulate during oocyte growth but they are not translated [reviewed in (Vassalli & Stutz, 1995)]. The molecular mechanism of dormancy involves cytoplasmic polyadenylation, which is takes place upon resumption of meiosis (Huarte et al, 1987; Stutz et al, 1998). Dormant maternal mRNAs facilitate important biological functions, including cell-cycle regulation (Oh et al, 1997), replication of pronucleus (Murai et al, 2010), fertilisation (Colledge et al, 1994), etc.

While maternal mRNAs support oocyte-to-zygote transition and early development, maternal mRNAs are also removed during these processes as well. There are three major waves of maternal mRNA degradation, the first wave already takes place already during meiotic maturation (Piko & Clegg, 1982).

## 4.3 Zygotic genome activation (ZGA) in early embryos

The mouse zygote forms upon fertilization, when a sperm enters an oocyte. The mouse 1-cell embryo contains maternal and paternal pronuclei, which do not fuse. These pronuclei replicate their genomes and enter the first mitosis (Krishna & Generoso, 1977). Blastomeres of the 2-cell stage contain the first diploid nuclei.

The transition between 1-cell and 2-cell embryos is accompanied by global changes in chromatin structure and initiation of the first transcription (Ahmed et al, 2010). Initiation of transcription is important for two reasons. First, mouse oocyte has a limited amount of stored mRNAs. Second, transcription is required for further cleavage of the embryo. Detailed analysis of transcriptome by microarrays shows that at 2-cell stage embryos contain transcripts of 10000 genes. Treatment by α-amanitin, potent inhibitor of polymerase II, shows that the actual number of genes transcribed at 2-cell stage embryos drop to 2600 genes (Figure 2)(Zeng et al, 2004). Importance of these genes was demonstrated by treatment with α-amanitin, after which embryos became blocked at the 2-cell stage (Braude, 1979; Flach et al, 1982). Microarray data suggests that transcriptional activation of the zygotic genome (ZGA) might not be as opportunistic as previously envisioned, but it is precisely organized and timed event (Hamatani et al, 2004; Zeng & Schultz, 2005). Interestingly, BrUTP incorporation experiments show that first RNA synthesis occurs already at the 1-cell stage (Aoki et al, 1997). Biological roles of most of the genes, which become active during the first wave of transcription are largely unknown. *Brg1* was the first identified regulator of ZGA and its depletion led to developmental arrest of 2-cell embryos, suggesting that *Brg1* facilitate ZGA (Bultman et al, 2006). *Brg1* is a part of histone remodeling complex and its ablation reduced dimethyl at lysine 4 on histone 3 (H3K4me2) marks in early embryos. This study indicates, ZGA is coupled with remodeling of the chromatin (Bultman et al, 2006).

Besides the major wave of transcription at 2-cell stage embryos, there are two other waves of transcription in mouse development, designated as second and third waves, occurring at 8-cell stage and 16-cell stage (Figure 2). During each wave of transcription, thousands of genes become active, which are involved in the morula and blastocyst formation (Braude, 1979; Hamatani et al, 2004).

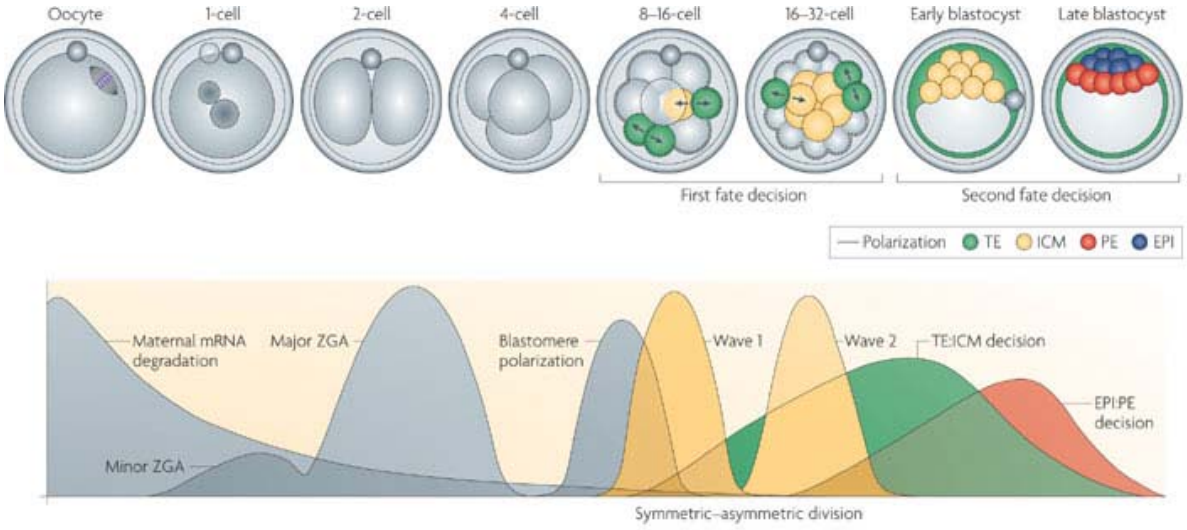


**Figure 2 Oocyte-to-zygote transition**

Transcriptome of early embryos and oocytes changes during preimplantation development. Maternal mRNAs are selectively degraded while zygotic mRNAs accumulate. Three independent waves of maternal mRNA degradation have been recognized. During the first wave of degradation, which occurs after resumption of meiosis, one third of maternal transcripts becomes degraded. The second and the third waves of degradation occur after fertilization and zygotic genome activation, respectively. Zygotic genome starts being transcribed at the 2-cell stage and the amount of mRNA in embryos is gradually increasing during development. There are three major waves of transcription occurring consecutively at the 2-cell stage, the 8-cell stage and the 16-cell stage. Novel transcribed genes in the embryo regulate embryo fate determination and further development. Figure adopted from (Zernicka-Goetz et al, 2009)

## 4.4 Preimplantation embryo development

Following ZGA at the 2-cell stage, an embryo cleaves to the 4-cell stage and then to the 8-cell stage. At the 8-cell stage, the embryo undergoes $Ca^{2+}$-dependent compaction and forms a 16-cell morula (Ziomek & Johnson, 1980). The morula stage is characterized by increased contacts between cells. The next cleavage of morula blastomeres produces two distinct populations – inner cells (ICM) and outer cells (known as the trophoectoderm) (Johnson & Ziomek, 1981; Pedersen et al, 1986). This formation of the two distinct cell types is also called the first cell fate decision. Besides first cell fate specification, cells in the ICM of the blastocyst undergo second specification, during which primitive endoderm forms on the surface of the ICM, and epiblast appears in the deeper layer of the ICM. Upon implantation into the uterus,

epiblast cells differentiate into the three germ layers and determine the future body plan of the embryo (Nichols et al, 1998).

The cell fate specification is controlled by activity of genes. The formation of trophoectoderm is induced by expression of *Cdx2* and *Tead4 (Niwa et al, 2005)*. In contrast, formation of the ICM is controlled by *Nanog*, *Oct4* and *Sox2* (Avilion et al, 2003; Mitsui et al, 2003; Nichols et al, 1998; Strumpf et al, 2005).

During second cell specification, activity of genes dictates formation of two distinct populations of cells within ICM. *Gata6* and *Gata4* dictate primitive endoderm, whereas deeper ICM cells express pluripotent genes *Oct4*, *Nanog* and *Sox2* (Figure 3) (Koutsourakis et al, 1999).

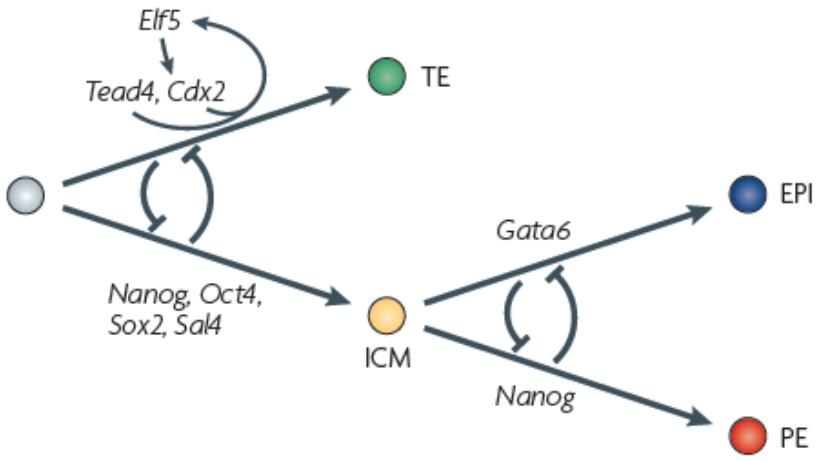


**Figure 3 Cell fate determination in early embryos**

Cell fate determination in early embryos is facilitated by activity of specific genes. *Tead4* and *Cdx2* are important for trophoectoderm formation. *Nanog*, *Oct4* and *Sox2* are implicated in formation of the ICM, which serves as pool for embryo development. Within the inner cell mass expression of *Gata6* and *Nanog* directs primitive endoderm and epiblast lineage, respectively. TE= trophoectoderm, EPI= epiblast, PE= primitive endoderm, ICM= inner cell mass. Scheme adopted from (Zernicka-Goetz, Morris et al. 2009)

### 4.4.1 Differentation into germ lineages

During normal development, soon after a blastocyst implants in a uterus, gastrulation initiates development of the bodyplan of an embryo. To track lineage commitment during differentiation of the epiblast, one can monitor expression lineage specific markers. In principle, it is expected that some genes are on a top of the

differentiation hierarchy and initiate the differentiation program. This hypothesis is supported by knockout studies, in which embryos lacking one copy of the gene Gsc gene exhibited defects in musculature defects in embryo development. To trace differentiation into ectodermal lineage *Nes*, *Pax3*, *Sox3* and *Crabp2* genes are widely used as markers (Bergsland et al, 2011; Goulding et al, 1991). For mesoderm well-conserved genes such as *Brachyury* (also known as *T)*, *Tbx6* and *Gsc* can be used (Hoffmann et al, 2002; Yamada et al, 1995). When monitoring endoderm lineage, *Sox17*, *Gata6* and *Gata4* are reliable markers (Figure 4) (Guo et al, 2010; Soudais et al, 1995).

## 4.5 Embryonic stem cells

Embryonic stem cells are derivatives of the ICM, which can be cultured *in vitro* under defined conditions (Evans & Kaufman, 1981). ESCs require a defined set of interleukins, growth and signaling factors for their proliferation. Among all, leukemia inhibitory factor (*LIF*) is a regulator of pluripotency in murine embryonic stem cells (Cartwright et al, 2005). *LIF* belongs to the *IL6* family of interleukins and regulates expression of *Stat3*. STAT3 acts as a transcription factor, which in turn triggers transcription of *c-Myc*. Importance of *LIF* for ESCs was demonstrated by a study, in which LIF removal from media positively regulated differentiation of ESCs (Cartwright et al, 2005).

Established ESC lines are used as an *in vitro* model for studying pluripotency maintenance and differentiation. *In vitro*, one of the most potent ectodermal inducers of ESC differentiation is *all-trans* retinoic acid (ATRA). ATRA treatment induces differentiation of ES cells into neural cell types. Therefore, ATRA has been widely used for testing the quality and ability of ES cells to differentiate (Simeone et al, 1990). In addition, these prepared cell lines are used as a tool for producing genetically manipulated mice, where a gene of interest can be artificially introduced or inactivated in ESCs. These modified ESCs can be transferred into a blastocyst of a pseudopregnant female and contribute to chimeras [reviewed in(Mansour, 1990)].

### 4.5.1 Core pluripotency network

While the first ESCs were obtained almost three decades ago, establishment and maintenance of the pluripotency program is not completely understood (Evans & Kaufman, 1981). What is known, however, is that pluripotency program can be initiated by expression of genes termed as the core pluripotency factors (Figure 4). Among them *Oct4*, *Sox2*, *Nanog*, *Stat3*, *c-Myc*, *Klf4*, *Esrrb* and *Zfx* are the most prominent members and are thought to be the major executors of pluripotency program (Boyer et al, 2006).

Detailed transcriptome analysis indicates that pluripotency network is being built in a stepwise manner during development. Pluripotency is established and maintained by genes transcribed from the 2-cell stage onward during preimplantation development. *Nanog* is a zygotic gene, which first transcribed at 8-cell stage. *Sox2* and *Oct4*, important regulators of pluripotency, are maternally provided factors but whose transcription appears at 8-cell stage (Hamatani et al, 2004). The expression of *Oct4*, *Sox2* and *Nanog* dictates the first cell fate determination. Importance of *Oct4* was supported by knockout studies, in which ablation of *Oct4* induced trophoectoderm formation in ESCs (Velkey & O'Shea, 2003). Consistent with this proposal is the fact that *Oct4*$^{-/-}$ failed to develop ICM (Nichols et al, 1998). Interestingly, chromatin immunoprecipitation experiment followed by massive parallel sequencing (ChIP-Seq) revealed that *Oct4*, *Nanog* and *Sox2* often co-occupy their own promoters and *Sox2* and *Oct4* operate as a part of the same complex suggesting that they target same genes (Chen et al, 2008; Loh et al, 2006; Rodda et al, 2005).

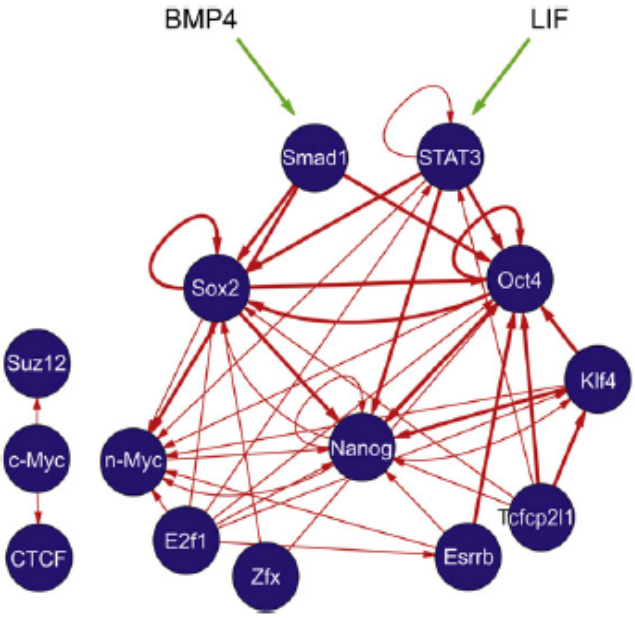


**Figure 4 Core pluripotency network**

The pluripotency program of is propagated by transcription factors which operate in a network. These genes activate a large number of genes. *Sox2*, *Nanog* and *Oct4* are essential key players in self-renewal of embryonic stem cells and pluripotency program. Scheme adopted from (Chen, Xu et al. 2008)

## 4.6 Induced pluripotent stem cells (iPS)

In 2006, Yamanaka and his colleagues performed a key experiment showing that pluripotency can be induced in somatic cells by a set of transcription factors (Takahashi et al, 2006). They used mouse embryonic fibroblasts (MEFs) and a lentiviral system expressing *Oct4*, *Sox2*, *c-Myc* and *Klf4*. Reprogrammed cells exhibited common features of pluripotent cells, such as surface markers, a specific signature of gene expression, and an ability to generate teratocarcinomas (Takahashi et al, 2006). Recent studies suggested that reprogramming of somatic cells is a stochastic process and can be rapidly improved via additional molecules (Hanna et al, 2009). Yamanaka and his colleagues recently tested a library of human transcription factors for the ability to replace *Oct4*, *Sox2*, *cMyc* and *Nanog* (Maekawa et al, 2011). They identified 18 novel transcription factors, which significantly increased reprogramming efficiency. Among those, maternal gene GLIS1 was identified as an important reprogramming regulator whose expression promotes iPS formation (Maekawa et al, 2011). There is evidence that reprogramming efficiency can be increased by a set of additional molecules (Yu et

al, 2007; Zhang et al, 2011). Besides transcription factors and inhibitors, other molecules such as non-coding RNAs seem to be important for reprogramming *in vitro* (Yang et al, 2011)

## 4.7 Non-coding RNAs & lincRNAs

In addition to protein coding genes, non-coding RNAs were identified as important regulators of gene expression in mammals. NcRNAs were reported to interfere with OZT in mouse(Murchison et al, 2007). Three clasees of small ncRNAs were found in mouse oocytes- Piwi-interacting RNAs (piRNAs), microRNAs (miRNAs) and short-interfering RNAs (siRNAs). PiRNAs have genome protective function, siRNAs and miRNAs interact with the mRNA and facilitate RNAi effect [reviewed in (Ohnishi et al, 2010)]. Interestingly, some ncRNAs, such as miRNAs, were shown to be non-functional during OZT (Ma et al, 2010). Recently, long ncRNAs were proposed as key regulators of biological processes. The role of long ncRNAs during OZT is, however, unknown.

LincRNAs are long (usually 2-5 kb) non-coding RNAs generated from intergenic regions. Most of lincRNA transcripts are polyadenylated at the 3' end and usually possess a cap at the 5' end (Guttman et al, 2009). LincRNAs were originally identified based on distribution of epigenetic modifications typical for transcriptional units. Promoter regions of transcribed lincRNAs bear mark of tri-methylation of lysine at position 4 on histone 3 (H3K4me3). The transcribed region, known as gene body, of lincRNAs contains methylated lysines at position 36 on histone 3 (H3K36me3). These epigenetic modifications thus can serve as a clue for identification of transcribed regions producing lincRNAs. H3K4me3 and H3K36me3 modifications are also shared with actively transcribed protein coding genes. Thus, some lincRNAs have been originally annotated as mRNAs. The major difference between mRNAs and lincRNAs is in their protein coding potential. While mRNAs serve as template for ribosomes and encode proteins, lincRNAs produce no proteins at all (Guttman et al, 2009). Despite the effort, the precise mechanism of function of most lincRNAs remains unclear (Figure 5). LincRNAs apparently bind diverse proteins including chromatin modifiers and transcription factors, and interact with chromatin at specific genomic loci. LincRNAs have been reported to regulate gene cellular processes, such as X-inactivation in

mammals (Brown et al, 1991), alternative splicing (Tripathi et al, 2010), DNA damage response (Huarte et al, 2010), etc. In addition, long non-coding RNAs have been found associated with cancer (Niinuma et al, 2012) and neurological disorders (Qureshi et al, 2010). In addition, lincRNAs can act as decoy molecules buffering key transcription factors (Azzalin et al, 2007).

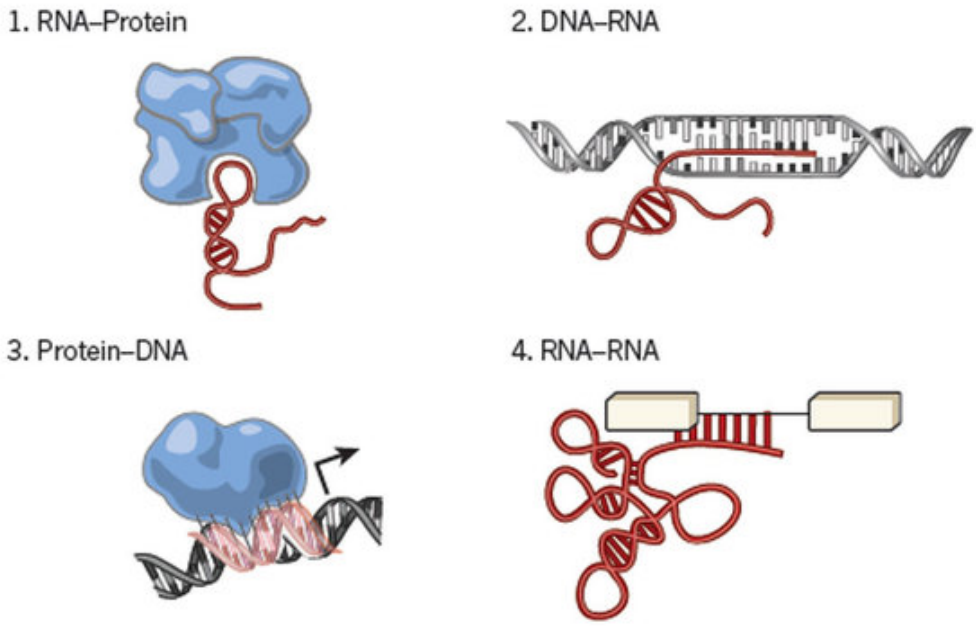


**Figure 5 Hypothetical models of long non-coding RNA action**

Four mechanisms how long ncRNAs can regulate cellular processes were proposed. Recent studies suggest that long ncRNAs can change binding properties of proteins and therefore modulate protein:protein interaction (1). The second potential mode of action assumes that long non-coding RNAs directly interact with DNA molecules and prevents binding of transcription factors (2). The third model suggests that long non-coding RNAs serve as protein binding platforms for transcription factors, enzymes, and DNA (3). The fourth mode of action proposes that long noncoding RNAs can directly interact with mRNAs or other non-coding RNAs (4). Figure adopted from (Guttman & Rinn, 2012)

### 4.7.1 The role of lincRNAs in the pluripotency program

Loss-of-function studies in ESCs revealed that some lincRNAs act as positive regulators of the pluripotent program and as negative regulators of differentiation (Guttman, et al. 2011). What is striking is the impact of lincRNAs on pluripotency and self-renewal program in ESCs in knockdown experiments. While inactivation of key transcription factors, such as *Oct4*, *Sox2* and *Nanog*, influenced on average hundreds of

protein coding genes, knockdown of specific lincRNAs influenced tens of protein coding genes (Guttman, et al. 2011). In addition, some lincRNAs have been associated with ectodermal, endodermal and mesodermal lineages (Guttman et al, 2011). Another study suggested that some lincRNAs strongly contribute to undifferentiated state of skin cells as their acute depletion resulted in burst of differentiation (Kretz et al, 2012). The role of lincRNAs during OZT, however, has not been addressed yet.

# 5 High-throughput methods for gene expression analysis

Recent progress in molecular biology has been accelerated by development of high-throughput tools for analysis of gene expression. The main advantage of these tools is their ability to monitor whole transcriptomes of samples. Today, there are three main high-throughput tools for transcriptome monitoring: microarrays, next generation sequencing and qPCR arrays.

## 5.1 Microarrays

Analysis of gene expression has been accelerated by Affymetrix who developed technology for quantification of transcripts in biological systems in 1990's. Affymetrix used sets of short 25-nt oligonucleotides (probes) against all mouse or human known transcripts. To increase specificity of detection the target, Affymetrix used ten to twelve probes against the same transcript. This set of probes is called a probeset. Affymetrix probes in the probeset target a defined region in a transcript. Probes organized into probesets were directly sythetized on a glass plate, giving rise to a microarray chip.

To quantify gene expression on array, one must isolate RNA molecules and convert them to fluorescently labeled cDNA. Subsequently, labeled cDNA is hybridized to probes on an Affymetrix chip (Figure 6). Gene expression is then calculated from intensity of fluorescence of individual probes in a probeset. In general, the more abundant a transcript is, the higher amount of intensity of fluorescence will be obtained [reviewed in (Ahmed, 2006)].

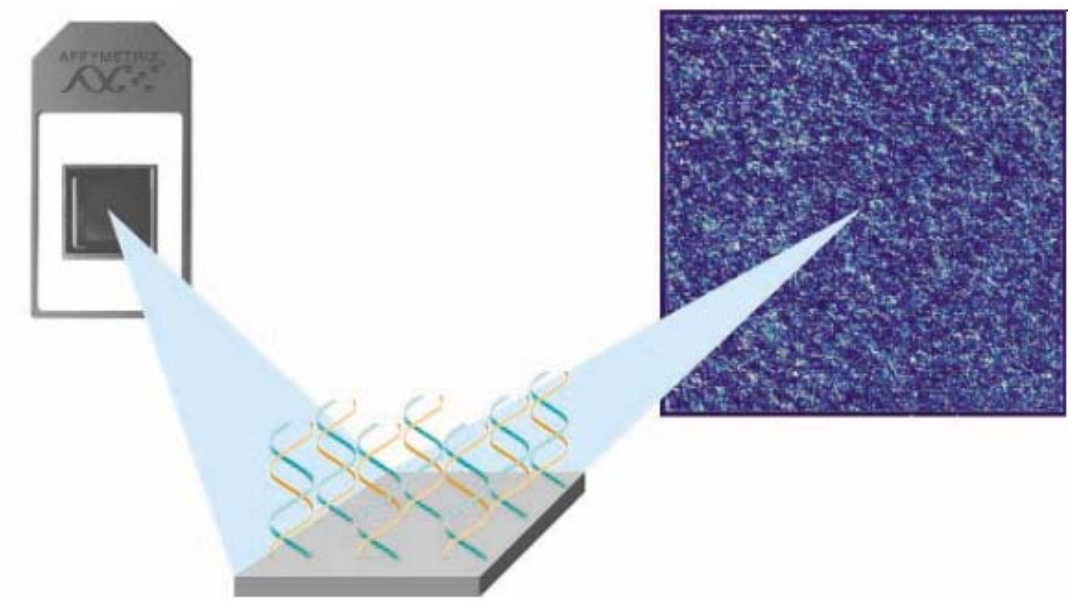

**Figure 6 Affymetrix technology**

A scheme of an Affymetrix array. Short nucleotides are immobilized on glass slide. Interaction between probes and fluorescence-labeled transcript generates a signal, which is scanned and quantified. The amount of fluorescence is proportional to the amount of transcripts in the tested sample. Figure adopted from (Affymetrix manufacturer website)

## 5.2 Next generation sequencing (NGS)

Next generation sequencing (NGS) technology was introduced in 2006 and it has become a powerful tool for transcriptome analysis. NGS platforms allow for sequencing the whole transcriptome in a single experiment. Today, there are three major providers of NGS technology: Illumina, Roche and Applied Biosystems. Each of these NGS technologies has certain advantages and disadvantages (Werner, 2010).

During NGS experiment, RNA is fragmented and ligated with adaptors. These fragments are amplified and immobilized on a glass chip or beads. Adaptor sequences serve as templates for sequencing primers, which initiate sequencing reaction. Next, fragments (also called tags or reads) are sequenced. There are three main methods how fragments can be sequenced. Roche modified pyrosequencing method, which uses luciferase to generate light for the detection of individual nucleotides added to nascent

DNA [reviewed in (Schuster, 2008)]. Illumina employs reversible dye-terminators to sequence the fragments. Applied Biosystems developed a sequencing method, which is based on ligation and fluorescence-labeled nucleotides [reviewed in (Mardis, 2008)].

The outcome from NGS experiment is extremely rich, as it contains tens of millions of short sequences, which must be processed. To analyse the data, one must apply complex algorithms, which align sequenced fragments against a reference sequence. The results are more accurate when compared to microarray data mainly because NGS allows for quantification of the whole transcriptome. In contrast, microarrays quantify only expression only of transcripts, against which probes are present. Therefore, NGS also reveals information of rare or novel RNAs present in the sample. On the other hand, volume of the data limits the speed of analysis. Since NGS is a relatively new method, mathematical and statistical methods are required for accurate data analysis. Unfortunately they are still under development [reviewed in (Rogers & Venter, 2005)].

## 5.3 Quantitative PCR arrays

Quantitative PCR is a method, by which one can measure abundance of a nucleic acid in a sample. In principle, it employs a fluorescent molecule, which binds specifically the amplified DNA and which emits a signal during each cycle of PCR. As the amount of DNA in the sample geometrically increases during the PCR, the total amount of fluorescence increases exponentially. The quantification of transcripts is based on the, so called, $C_T$ value, which is defined by a cycle, in which fluorescence curve reaches a particular baseline (Figure 7) (Nolan et al, 2006). Relative changes in the amount of transcripts can be estimated from $C_T$ value difference. Recently, quantitative microfluidic PCR arrays became a powerful high-throughput tool for quantification of gene expression. Fluidigm™ developed a platform allowing quantification of large amount of transcripts during one experiment by reducing total volume of PCR reaction (Spurgeon et al, 2008). A great advantage of Fluidigm™ is that it developed system how to perform qPCR reactions on a chip.

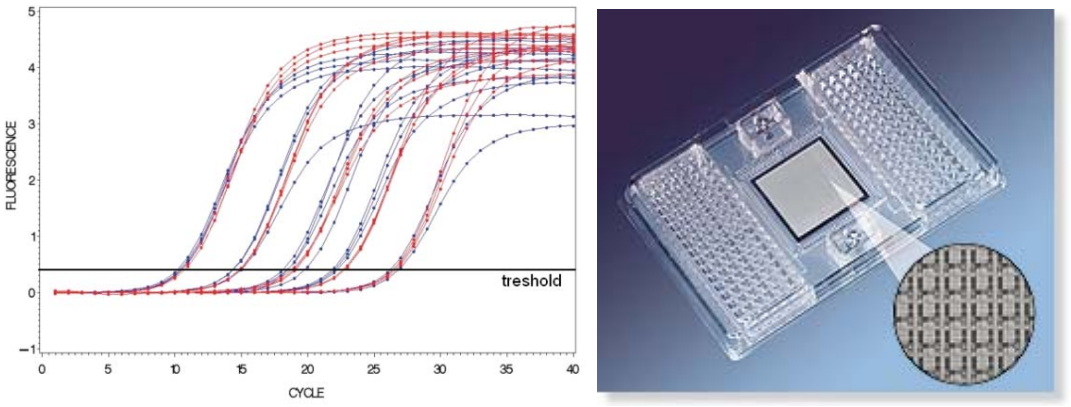


**Figure 7 Output of the qPCR experiment**

Left panel represents quantitative PCR method uses fluorescence-based reporter, which binds specifically the double-stranded DNA. During each cycle reporter emits amount of fluorescence, which is scanned and quantified. Right panel represents qPCR array with nanoliter chambers. Figure adopted from (Bar et al, 2012)

# 6 Methods

## 6.1 List of instruments

Centrifuge (Biofuge)
Centrifuge (Eppendorf)
$CO_2$ incubator (Jouan)
DeltaVision (Applied Precision)
Flow box (Jouan)
Fluidigm cycler (Biomark)
Gel Dox XR+ (BioRad)
Horizontal electrophoresis (BioRad)
Light cycler 384 (Roche)
LightCycler 480 System (Roche Applied Science)
Mx3000P cycler (Stratagene)
ND-1000 Spectrophotometer (NanoDrop)
Odyssey Infrared Imaging System (LI-COR)
PCR cycler MJ Mini Gradient (BioRad)
Power supplies (BioRad)
Rocker (Biosan)
Rotator (Biosan)
Shaker 37°C (Lab. Companion)
Thermo shaker (Biosan)
Vertical electrophoresis (BioRad)
Vortex (Scientific Industries)

## 6.2 Primer design

Primers were designed based on Ensembl (version 58) sequences using default settings of Universal probe library primer design software (Roche, www.universalprobelibrary.com). Primers were designed for melting temperature $T_m$ = 60°C. Primers were compared with the collection of cDNAs deposited in Ensembl database to assure that correct primer sequences were selected. Several primers were further modified by increasing GC content of the 3' end of the primer. Primers were

obtained from Sigma. Primers were tested by a real-time PCR using various templates. Quality of the primers was determined from amplification and melting curves. Primers, which made dimers and/or had low efficiency of amplification were replaced.

**Oocyte-to-zygotic set**

| Primer | Sequence | Product | UPL probe |
|---|---|---|---|
| Actb_qpcr_fwd | taaggccaaccgtgaaaagat | 109 nt | #64, cat.no. 04688635001 |
| Actb_qpcr_rev | ggtacgaccagaggcatacag | 109 nt | #64, cat.no. 04688635001 |
| BMP15_qpcr_fwd | acacagtaaggcctcccaga | 75 nt | #72, cat.no. 04688953001 |
| BMP15_qpcr_rev | atgctacctggtttgatgctagag | 75 nt | #72, cat.no. 04688953001 |
| Cbx2_qpcr_fwd | ggccgaggaaacacacag | 75 nt | #88, cat.no. 04689135001 |
| Cbx2_qpcr_rev | atttggatggcgcatctg | 75 nt | #88, cat.no. 04689135001 |
| Cdh1_qpcr_fwd | gccaccagatgatgataccc | 89 nt | #10, cat.no. 04685091001 |
| Cdh1_qpcr_rev | gctggctcaaatcaaagtcc | 89 nt | #10, cat.no. 04685091001 |
| Cdkn1a_qpcr_fwd | agatccacagcgatatccagac | 103 nt | #21, cat.no. 04686942001 |
| Cdkn1a_qpcr_rev | aagagacaacggcacactttg | 103 nt | #21, cat.no. 04686942001 |
| cMyc_qpcr_fwd | ctagtgctgcatgaggagacac | 90 nt | #77, cat.no. 04689003001 |
| cMyc_qpcr_rev | cacagacaccacatcaatttcttc | 90 nt | #77, cat.no. 04689003001 |
| Dazl_qpcr_fwd | tgatattttgcccaatgaatgtt | 88 nt | #78, cat.no. 04689011001 |
| Dazl_qpcr_rev | tatgcttcggtccacagactt | 88 nt | #78, cat.no. 04689011001 |
| Dcp1a_qpcr_fwd | ccttccattatcctcagcaagt | 76 nt | #1, cat.no. 04684974001 |
| Dcp1a_qpcr_rev | tgaggaagctggagtcattct | 76 nt | #1, cat.no. 04684974001 |
| Dnmt3b_qpcr_fwd | ccagggccttctttcagg | 90 nt | #94, cat.no. 04692110001 |
| Dnmt3b_qpcr_rev | cgataatgcactcctataccc | 90 nt | #94, cat.no. 04692110001 |
| E2F2_qpcr_fwd | gcgcatctatgacatcacca | 100 nt | #67, cat.no. 04688660001 |
| E2F2_qpcr_rev | gtcttcaaatagttccctgcctac | 100 nt | #67, cat.no. 04688660001 |
| E2F5_qpcr_fwd | ctgcaattgctttcatggtg | 92 nt | #11, cat.no. 04685105001 |
| E2F5_qpcr_rev | cattctgtcccatttctggaata | 92 nt | #11, cat.no. 04685105001 |
| E2F7_qpcr_fwd | tgttacgtgagacatccggta | 73 nt | #13, cat.no. 04685121001 |
| E2F7_qpcr_rev | ggatgctcttgggagtcg | 73 nt | #13, cat.no. 04685121001 |
| Eif1a_qpcr_fwd | gccagaaccgaagtactattttgt | 97 nt | #27, cat.no. 04687582001 |
| Eif1a_qpcr_rev | caactgggacactgtgaatatagaa | 97 nt | #27, cat.no. 04687582001 |
| Eif3l_qpcr_fwd | gatggcgaattccagtcg | 73 nt | #10, cat.no. 04685091001 |
| Eif3l_qpcr_rev | tggtgtctgcaatatggatca | 73 nt | #10, cat.no. 04685091001 |
| Esrrb_qpcr_fwd | cgattcatgaaatgcctcaaa | 68 nt | #89, cat.no. 04689143001 |
| Esrrb_qpcr_rev | cctcctcgaactcggtca | 68 nt | #89, cat.no. 04689143001 |
| Gapdh_qpcr_fwd | cggtgctgagtatgtcgtgg | 130 nt | #29, cat.no. 04687612001 |
| Gapdh_qpcr_rev | tcacacccatcacaaacatgg | 130 nt | #29, cat.no. 04687612001 |
| GDF9_qpcr_fwd | ctacaataccgtccggctct | 99 nt | #104, cat.no. 04692225001 |
| GDF9_qpcr_rev | ttaaacagcaggtccaccatc | 99 nt | #104, cat.no. 04692225001 |
| GLB_qpcr_fwd | cgtggagaggatgttcttgg | 67 nt | #77, cat.no. 04689003001 |
| GLB_qpcr_rev | gtgggtgaagtcgaagtgg | 67 nt | #77, cat.no. 04689003001 |
| Hdac1_qpcr_fwd | tggtctctaccgaaaaatggag | 78 nt | #73, cat.no. 04688961001 |

| | | | |
|---|---|---|---|
| Hdac1_qpcr_rev | tcatcactgtggtacttggtca | 78 nt | #73, cat.no. 04688961001 |
| Hmga2_qpcr_fwd | aaaacaagagcccctctaaagc | 105 nt | #34, cat.no. 04687671001 |
| Hmga2_qpcr_rev | tcttctgaacgacttgttgtgg | 105 nt | #34, cat.no. 04687671001 |
| Hprt1_qpcr_fwd | cctcagaccgctttttgc | 74 nt | #95, cat.no. 04692128001 |
| Hprt1_qpcr_rev | cctggttcatcatcgctaatc | 74 nt | #95, cat.no. 04692128001 |
| IAP_qpcr_fwd | cgaggtgttctccactccat | 85 nt | #13, cat.no. 04685121001 |
| IAP_qpcr_rev | acgtgtcactccctgattgg | 85 nt | #13, cat.no. 04685121001 |
| Igf2bp1_qpcr_fwd | gatgagaacgaccaagtcattg | 76 nt | #20, cat.no. 04686934001 |
| Igf2bp1_qpcr_rev | ctcggatcttccgctgag | 76 nt | #20, cat.no. 04686934001 |
| Klf4_qpcr_fwd | agtcccctctctccattatcaag | 84 nt | #82, cat.no. 04689054001 |
| Klf4_qpcr_rev | gaccttcttcccctctttgg | 84 nt | #82, cat.no. 04689054001 |
| Lhx8_qpcr_fwd | gagctcggaccagcttca | 67 nt | #1, cat.no. 04684974001 |
| Lhx8_qpcr_rev | ttgttgtcctgagcgaactg | 67 nt | #1, cat.no. 04684974001 |
| Lin28_qpcr_fwd | aagaacatgcagaagcgaagat | 73 nt | #94, cat.no. 04692110001 |
| Lin28_qpcr_rev | ccttggcatgatggtctagc | 73 nt | #94, cat.no. 04692110001 |
| Mos_qpcr_fwd | tgagcaagacgtttgtaagatca | 95 nt | #32, cat.no. 04687655001 |
| Mos_qpcr_rev | tgccccctatgtggtgag | 95 nt | #32, cat.no. 04687655001 |
| MT_qpcr_fwd | atgtcttggggaggactgtg | 306 nt | #45, cat.no. 04688058001 |
| MT_qpcr_rev | aaccagcatcaatagtcccagt | 306 nt | #45, cat.no. 04688058001 |
| Muerv_qpcr_fwd | tattatttgtgtcaagttgacaagg | 150 nt | #97, cat.no. 04692144001 |
| Muerv_qpcr_rev | cctccagataagggtcactgg | 150nt | #97, cat.no. 04692144001 |
| Nanog_qpcr_fwd | tacctcagcctccagcagat | 82 nt | #25, cat.no. 04686993001 |
| Nanog_qpcr_rev | ggttttgaaaccaggtcttaacc | 82 nt | #25, cat.no. 04686993001 |
| Nlrp5_qpcr_fwd | gcagacatcagaaaccttacaatc | 91 nt | #34, cat.no. 04687671001 |
| Nlrp5_qpcr_rev | ggccttgtagtcttgtaagtcacc | 91 nt | #34, cat.no. 04687671001 |
| Nobox_qpcr_fwd | aaagacccgaaccctgtacc | 67 nt | #11, cat.no. 04685105001 |
| Nobox_qpcr_rev | gtggtcttcctgaaatatcctctc | 67 nt | #11, cat.no. 04685105001 |
| Obox1_qpcr_fwd | ccttgaagacttttgacacatcag | 95 nt | #78, cat.no. 04689011001 |
| Obox1_qpcr_rev | aaggttggactcgtcaaggac | 95 nt | #78, cat.no. 04689011001 |
| Oog1_qpcr_fwd | ggtgatctgtctccattgtcc | 111 nt | #6, cat.no. 04685032001 |
| Oog1_qpcr_rev | tccctcagtagactctgaattgc | 111 nt | #6, cat.no. 04685032001 |
| Plat_qpcr_fwd | cctcatgggcaagagttacac | 114 nt | #9, cat.no. 04685075001 |
| Plat_qpcr_rev | atcacatggcaccaaggtct | 114 nt | #9, cat.no. 04685075001 |
| Pou5f1_qpcr_fwd | gttggagaaggtggaaccaa | 75 nt | #95, cat.no. 04692128001 |
| Pou5f1_qpcr_rev | gcaaactgttctagctccttctg | 75 nt | #95, cat.no. 04692128001 |
| Ppil3_qpcr_fwd | ttcgaggtgttgtatctatggcta | 131 nt | #5, cat.no. 04685024001 |
| Ppil3_qpcr_rev | ctccagaccatctattacctttcc | 131 nt | #5, cat.no. 04685024001 |
| Prb1(107)_qpcr_fwd | gcggcaactacagcctagag | 66 nt | #20, cat.no. 04686934001 |
| Prb1(107)_qpcr_rev | ggcaagcaacatataaagagca | 66 nt | #20, cat.no. 04686934001 |
| Rbl2(130)_qpcr_fwd | agattgggagacatggatttatct | 63 nt | #50, cat.no. 04688112001 |
| Rbl2(130)_qpcr_rev | caagagtgacctgtggaatgc | 63 nt | #50, cat.no. 04688112001 |
| Rpl18a_qpcr_fwd | cgcatgatccgaaagatga | 79 nt | #72, cat.no. 04688953001 |
| Rpl18a_qpcr_rev | cagaatccgcacatcatctgt | 87 nt | #72, cat.no. 04688953001 |
| Sox2_qpcr_fwd | acagctacgcgcacatga | 99 nt | #19, cat.no. 04686926001 |
| Sox2_qpcr_rev | ggtagcccagctgctcct | 99 nt | #19, cat.no. 04686926001 |

| | | | |
|---|---|---|---|
| Spin1_qpcr_fwd | ctcctcgatgactacaaagaagg | 123 nt | #2, cat.no. 04684982001 |
| Spin1_qpcr_rev | ggcatattccacttgcttgc | 123 nt | #2, cat.no. 04684982001 |
| Trim71_qpcr_fwd | ttctccattctctcggtgttc | 91 nt | #29, cat.no. 04687612001 |
| Trim71_qpcr_rev | cagagcaggtgtcacagtagagat | 91 nt | #29, cat.no. 04687612001 |
| Trp53_qpcr_fwd | gagtatctggaagacaggcagac | 170 nt | #47, cat.no. 04688074001 |
| Trp53_qpcr_rev | ccagaaggttcccactgga | 170 nt | #47, cat.no. 04688074001 |
| YY1_qpcr_fwd | agaactcacctcctgattattctga | 128 nt | #79, cat.no. 04689020001 |
| YY1_qpcr_rev | aatttttcttggcttcattctgg | 128 nt | #79, cat.no. 04689020001 |
| Zar1_qpcr_fwd | catgtcctgccgcagaga | 95 nt | #15, cat.no. 04685148001 |
| Zar1_qpcr_rev | ccgtacttctgctctaagaactgg | 95 nt | #15, cat.no. 04685148001 |
| ZP3_qpcr_fwd | ctctccagttcacggtggat | 73 nt | #50, cat.no. 04688112001 |
| ZP3_qpcr_rev | agatggcaggtgatgtagagc | 73 nt | #50, cat.no. 04688112001 |
| Ccnb1_qpcr_fwd | tgcattttgctccttctcaa | 126 nt | #45, cat.no. 04688058001 |
| Ccnb1_qpcr_rev | caggaagcagggagtcttca | 126 nt | #45, cat.no. 04688058001 |

**Pluripotency&Differentiation set**

| **Primer** | **Sequence** | **Product** | **UPL probe** |
|---|---|---|---|
| Actb_qpcr_fwd | taaggccaaccgtgaaaagat | 109 nt | #64, cat.no. 04688635001 |
| Actb_qpcr_rev | ggtacgaccagaggcatacag | 109 nt | #64, cat.no. 04688635001 |
| Bax_qpcr_fwd | acactggacttcctccgtga | 84 nt | #83, cat.no. 04689062001 |
| Bax_qpcr_fwd | acactggacttcctccgtga | 84 nt | #83, cat.no. 04689062001 |
| Bax_qpcr_rev | ggtcccgaagtaggagagga | 84 nt | #83, cat.no. 04689062001 |
| Bax_qpcr_rev | ggtcccgaagtaggagagga | 84 nt | #83, cat.no. 04689062001 |
| Bcl2_qpcr_fwd | agtacctgaaccggcatctg | 77 nt | #75, cat.no. 04688988001 |
| Bcl2_qpcr_rev | ggggccatatagttccacaaa | 77 nt | #75, cat.no. 04688988001 |
| Bim_qpcr_fwd | cgagttcaacgaaacttacacaag | 108 nt | #41, cat.no. 04688007001 |
| Bim_qpcr_rev | agacggaagataaagcgtaacagt | 108 nt | #41, cat.no. 04688007001 |
| Brach.T_qpcr_fwd | gataactggtctagcctcggagt | 107 nt | #27, cat.no. 04687582001 |
| Brach.T_qpcr_fwd | gataactggtctagcctcggagt | 107 nt | #27, cat.no. 04687582001 |
| Brach.T_qpcr_rev | acagaccagagactgggatactg | 107 nt | #27, cat.no. 04687582001 |
| Brach.T_qpcr_rev | acagaccagagactgggatactg | 107 nt | #27, cat.no. 04687582001 |
| Cdkn1a_qpcr_fwd | agatccacagcgatatccagac | 103 nt | #21, cat.no. 04686942001 |
| Cdkn1a_qpcr_rev | aagagacaacggcacactttg | 103 nt | #21, cat.no. 04686942001 |
| Cdkn2a_qpcr_rev | atctggagcagcatggagtc | 131 nt | #70, cat.no. 04688937001 |
| Cdkn2a_qpcr_rev | ggggtacgaccgaaagagtt | 131 nt | #70, cat.no. 04688937001 |
| Cdx2_qpcr_fwd | atacatcaccatcaggaggaaaag | 85 nt | #34, cat.no. 04687671001 |
| Cdx2_qpcr_rev | gcggttctgaaaccaaatttta | 85 nt | #34, cat.no. 04687671001 |
| Cebpa_qpcr_fwd | ccaaactgagactcttcactaacg | 72 nt | #12, cat.no. 04685113001 |
| Cebpa_qpcr_rev | tccctaaaccaaaaagaatgagag | 72 nt | #12, cat.no. 04685113001 |
| cMyc_qpcr_fwd | ctagtgctgcatgaggagacac | 90 nt | #77, cat.no. 04689003001 |
| cMyc_qpcr_rev | cacagacaccacatcaatttcttc | 90 nt | #77, cat.no. 04689003001 |
| Crabp2_qpcr_fwd | cacggagattaacttcaagatcg | 89 nt | #68, cat.no. 04688678001 |
| Crabp2_qpcr_rev | cactctcccatttcaccaaac | 89 nt | #68, cat.no. 04688678001 |

| | | | |
|---|---|---|---|
| Creb3l_qpcr_fwd | cttatccttctgccaccaaga | 101 nt | #75, cat.no. 04688988001 |
| Creb3l_qpcr_rev | ttagcaggttcctggatctcac | 101 nt | #75, cat.no. 04688988001 |
| Dppa3_qpcr_fwd | aagcaatcttgttccgagcta | 88 nt | #29, cat.no. 04687612001 |
| Dppa3_qpcr_rev | ccttcattgggtcgactttc | 88 nt | #29, cat.no. 04687612001 |
| Esrrb_qpcr_fwd | cgattcatgaaatgcctcaaa | 68 nt | #89, cat.no. 04689143001 |
| Esrrb_qpcr_rev | cctcctcgaactcggtca | 68 nt | #89, cat.no. 04689143001 |
| FGF5_qpcr_fwd | aaaacctggtgcaccctaga | 65 nt | #29, cat.no. 04687612001 |
| FGF5_qpcr_rev | catcacattcccgaattaagc | 65 nt | #29, cat.no. 04687612001 |
| Fgfr2_qpcr_fwd | cactctgcatggttgacagttc | 102 nt | #60, cat.no. 04688589001 |
| Fqfr2_qpcr_rev | gaagacccctatgcagtaaatagc | 102 nt | #60, cat.no. 04688589001 |
| Gapdh_qpcr_fwd | cggtgctgagtatgtcgtgg | 130 nt | #29, cat.no. 04687612001 |
| Gapdh_qpcr_rev | tcacacccatcacaaacatgg | 130 nt | #29, cat.no. 04687612001 |
| Gata3_qpcr_fwd | cttatcaagcccaagcgaag | 76 nt | #108, cat.no. 04692276001 |
| Gata3_qpcr_rev | tggtggtggtctgacagttc | 76 nt | #108, cat.no. 04692276001 |
| Gata4_qpcr_fwd | gcccaagaacctgaataaatctaa | 103 nt | #18, cat.no. 04686918001 |
| Gata4_qpcr_rev | gctagtggcattgctggagt | 103 nt | #18, cat.no. 04686918001 |
| Gata6_qpcr_fwd | ggtctctacagcaagatgaatgg | 104 nt | #40, cat.no. 04687990001 |
| Gata6_qpcr_fwd | ggtctctacagcaagatgaatgg | 104 nt | #40, cat.no. 04687990001 |
| Gata6_qpcr_rev | gtgtgacagttggcacagga | 104 nt | #40, cat.no. 04687990001 |
| Gata6_qpcr_rev | gtgtgacagttggcacagga | 104 nt | #40, cat.no. 04687990001 |
| Gsc_qpcr_fwd | gagacgaagtacccagacgtg | 119 nt | #32, cat.no. 04687655001 |
| Gsc_qpcr_rev | cgcttctgtcgtctccactt | 119 nt | #32, cat.no. 04687655001 |
| Hprt1_qpcr_fwd | cctcagaccgctttttgc | 90 nt | #95, cat.no. 04692128001 |
| Hprt1_qpcr_rev | cctggttcatcatcgctaatc | 90 nt | #95, cat.no. 04692128001 |
| Klf4_qpcr_fwd | agtcccctctctccattatcaag | 84 nt | #82, cat.no. 04689054001 |
| Klf4_qpcr_rev | gaccttcttcccctctttgg | 84 nt | #82, cat.no. 04689054001 |
| Lefty1_qpcr_fwd | ctgcccttatcgattctaggc | 91 nt | #97, cat.no. 04692144001 |
| Lefty1_qpcr_rev | agctgctgccagaagttcac | 91 nt | #97, cat.no. 04692144001 |
| Lin28_qpcr_fwd | aagaacatgcagaagcgaagat | 73 nt | #94, cat.no. 04692110001 |
| Lin28_qpcr_rev | ccttggcatgatggtctagc | 73 nt | #94, cat.no. 04692110001 |
| Msc_qpcr_fwd | agctttccaaactggacacg | 135 nt | #11, cat.no. 04685105001 |
| Msc_qpcr_rev | gtccagagaccacgaatgg | 135 nt | #11, cat.no. 04685105001 |
| Nanog_qpcr_fwd | tacctcagcctccagcagat | 82 nt | #25, cat.no. 04686993001 |
| Nanog_qpcr_rev | ggttttgaaaccaggtcttaacc | 82 nt | #25, cat.no. 04686993001 |
| Nes_qpcr_fwd | ctgcaggccactgaaaagtt | 73 nt | #1, cat.no. 04684974001 |
| Nes_qpcr_rev | tctgactctgtagaccctgcttc | 73 nt | #1, cat.no. 04684974001 |
| Pdgfra_qpcr_fwd | gtcgttgacctgcagtgga | 61 nt | #80, cat.no. 04689038001 |
| Pdgfra_qpcr_rev | ccagcatggtgatacctttgt | 61 nt | #80, cat.no. 04689038001 |
| Pecam1_qpcr_fwd | actcacgctggtgctctatg | 62 nt | #64, cat.no. 04688635001 |
| Pecam1_qpcr_rev | tgctgttgatggtgaaggag | 62 nt | #64, cat.no. 04688635001 |
| Pou5f1_qpcr_fwd | gttggagaaggtggaaccaa | 75 nt | #95, cat.no. 04692128001 |
| Pou5f1_qpcr_rev | gcaaactgttctagctccttctg | 75 nt | #95, cat.no. 04692128001 |
| Rbl2(130)_qpcr_fwd | agattgggagacatggatttatct | 63 nt | #50, cat.no. 04688112001 |
| Rbl2(130)_qpcr_rev | caagagtgacctgtggaatgc | 63 nt | #50, cat.no. 04688112001 |
| Sox13_qpcr_fwd | atgtggaagctaaggatgtcaaag | 74 nt | #102, cat.no. 04692209001 |

| | | | |
|---|---|---|---|
| Sox13_qpcr_rev | gatcatgaccaaaagctggagt | 74 nt | #102, cat.no. 04692209001 |
| Sox17_qpcr_fwd | aacgcagagctaagcaagatg | 129 nt | #53, cat.no. 04688503001 |
| Sox17_qpcr_rev | gtacttgtagttggggtggtcct | 129 nt | #53, cat.no. 04688503001 |
| Sox2_qpcr_fwd | acagctacgcgcacatga | 99 nt | #19, cat.no. 04686926001 |
| Sox2_qpcr_rev | ggtagcccagctgctcct | 99 nt | #19, cat.no. 04686926001 |
| Sox3_qpcr_fwd | gaccgttgccttgtaccg | 62 nt | #101, cat.no. 04692195001 |
| Sox3_qpcr_rev | aaaaccccgacagttacgg | 62 nt | #101, cat.no. 04692195001 |
| Stat3_qpcr_fwd | agtttggaaataacggtgaaggt | 71 nt | #18, cat.no. 04686918001 |
| Stat3_qpcr_rev | catgtcaaacgtgagcgact | 71 nt | #18, cat.no. 04686918001 |
| Tbx6_qpcr_fwd | aggaactgtggaaggaattcag | 93 nt | #9, cat.no. 04685075001 |
| Tbx6_qpcr_rev | tgactgatactcggcaagca | 93 nt | #9, cat.no. 04685075001 |
| Tcfap2a_qpcr_fwd | caagtacgaagactgcgagga | 97 nt | #104, cat.no. 04692225001 |
| Tcfap2a_qpcr_rev | gctggtgtagggagattgacc | 97 nt | #104, cat.no. 04692225001 |
| Tead4_qpcr_fwd | ctctacgaaggtctgctcatttg | 74 nt | #22, cat.no. 04686969001 |
| Tead4_qpcr_rev | cattctcatagcgggcatactc | 74 nt | #22, cat.no. 04686969001 |
| Trp53_qpcr_fwd | gagtatctggaagacaggcagac | 170 nt | #47, cat.no. 04688074001 |
| Trp53_qpcr_fwd | atgcccatgctacagaggag | 78 nt | #94, cat.no. 04692110001 |
| Trp53_qpcr_rev | ccagaaggttcccactgga | 170 nt | #47, cat.no. 04688074001 |
| Trp53_qpcr_rev | aagtagactggcccttcttggt | 78 nt | #94, cat.no. 04692110001 |
| VNP_qpcr_fwd | agaagcgcgatcacatgg | 63 nt | #67, cat.no. 04688660001 |
| VNP_qpcr_rev | ccatgccgagagtgatcc | 63 nt | #67, cat.no. 04688660001 |
| Zfp42_qpcr_fwd | ggatttccttttttaaatccttcg | 78 nt | #69, cat.no. 04688686001 |
| Zfp42_qpcr_rev | gaactcgcttccagaacctg | 78 nt | #69, cat.no. 04688686001 |

Red color highlights primers, which are not spanning exon:exon junction. Black color represents spanning exon:exon junction.

## 6.3 Sample collection

Fully-grown GV oocytes were obtained from sacrificed mice (*C57B16xBalb-c)* by puncturing antral follicles with a needle. Oocytes were collected in M2 medium (Sigma) containing 0.2 mM isobutylmethylxanthine (IBMX; Sigma) to prevent resumption of meiosis. To obtain MII oocytes and embryos, female mice 14-16 weeks of age, were superovulated with 5 U of pregnant mare serum gonadotropin (PMSG, Intervet) followed by stimulation with 5 U chorionic gonadotropin (hCG). MII oocytes were collected 16 hours post hCG injection by tearing oviduct ampulla. To isolate 1-cell embryos, superovulated female mice were mated with *C57B16* males overnight. Isolation was performed 24-26 hours post hCG in M2 medium containing 3 mg/ml hyaluronidase (Sigma) to remove cumulus cells. 2-cell, 4-cell and 8-cell stage embryos were isolated by tearing ampulla or flushing uterus and were collected 48 hours, 60 hours, and 72 hours post hCG injection, respectively. All samples were isolated in pre-

warmed M2 medium (Sigma) and washed 3-times in PBS before transfer to an eppendorf tube.

Samples were transferred into a mix of 3.7 µl of nuclease-free water (Fermentas), 0.3 µl of Ribolock RNase inhibitor (Fermentas), and 1 µl of 1000x diluted rabbit globin RNA (Sigma), which served as an external normalization control. Samples were immediately stored at -80°C. All animal experiments were approved by the Institutional Animal use and Care Committee and were consistent with the Czech law.

## 6.4 Culture conditions

Human HEK293 and mouse P19 embryocarcinoma cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) (Invitrogen) supplemented with 10% fetal calf serum (FCS) (Sigma), penicillin (100 U/mL, Invitrogen), and streptomycin (100 µg/mL, Invitrogen). Cells were cultured at 37 °C and 5% $CO_2$ atmosphere. P19 cells were further supplemented with 50 µM β-mercaptorthanol (Invitrogen).

Mouse embryonic stem cells (mESC) were cultured in 2i-LIF medium (Silva et al, 2008). DMEM was supplemented with 15% ESC-compatible FCS (Invitrogen), penicillin (100 U/mL), streptomycin (100 µg/mL), 50 µM β-mercaptoethanol, 2 mM L-glutamine (Invitrogen), and 100 µM non-essential amino acids (Invitrogen). To keep ES cells in undifferentiated state, inhibitors of mitogen-activated protein kinase/extracellular signal-regulated kinase (1 µM PD0325901) and glycogen synthase kinase 3β (3 µM CHIR99021) (both Selleck Chemicals) were used.

## 6.5 RNA isolation

RNA was isolated with RNA Blue (Top-Bio) according to the manufacturer's instructions. Briefly, P19, HEK293 and mouse ES cells were washed in PBS three times and lysed in 0.5 ml of RNA Blue reagent. Samples were shaked and left on bench at room temperature for 10 minutes. Subsequently, 200 µl of bromochloropropane (Sigma) was added and samples were centrifugated at 13000 rpm for 10 minutes at 4°C. The aqueous phase was transferred into a new 0.5 ml tube and RNA was precipitated with 250 µl of isopropanol. Samples were freezed in -20°C for overnight. Next day, samples were centrifuged at 13000 rpm for 30 minutes at 4°C and precipitated RNA was washed

with 75% ethanol. The final pellet of RNA was resuspended in 12 µl of nuclease-free water (Fermentas). Isolated RNA was either used directly for reverse transcription or stored at -80°C freezer.

## 6.6 cDNA synthesis

The amount of RNA isolated from population of mESCs, HEK293 and p19 cells was measured by NanoDrop (Thermo Fisher). For reverse transcription, 2 µg of total RNA were used as a template. Reverse transcription reaction included 4 µl M-MuLV buffer (Fermentas), 1 µl random hexamer primer (Fermentas), 4 µl of 10 mM dNTPs, 0.5 µl of RiboLock (Fermentas), and 1 µl of (20 U) Revert Aid reverse transcriptase (Fermentas), and nuclease-free water (Fermentas) up to the total volume of 20 µl. cDNA synthesis was performed according to the conditions described above.

For single-cell qPCR, cDNA was prepared as follows: individual oocytes or embryos were lysed by heating to 85°C for 10 minutes. Samples were placed on ice and 0.5 µl random hexamer primers (Fermentas) was added to prime reverse transcription. cDNA synthesis was performed using Revert Aid Reverse Transcription enzyme (Fermentas) according to the manufacturer's protocol. Briefly, 2 µl M-MuLV buffer (Fermentas), 2 µl 10 mM dNTPs and 0.5 µl (10 U) Revert Aid reverse transcriptase (Fermentas) were added to each sample. Samples were incubated at room temperature for 10 minutes. Reverse transcription was performed at 42°C for 50 minutes. Reverse transcriptase was inactivated by heating samples at 70°C for 10 minutes, after inactivation samples were immediately centrifuged and stored at -20°C.

## 6.7 Preamplification

To yield an optimal amount of template for real-time PCR arrays, preamplification PCR was performed. Each preamplification reaction consisted of 20 µl of 2x Maxima qPCR SYBR GREEN Mix (Fermentas), 2 µl of preamplification primer mix consisting of 48 primers (500 nM each), 4 µl cDNA from reverse transcription and 14 µl of sterile nuclease-free water (Fermentas). PCR was performed as follows: 10 minutes of initial denaturation at 95°C to activate DNA polymerase, and 18 cycles of 95°C for 15 seconds (denaturation), 57°C for 4 minutes (annealing), and 72°C for 30 seconds (elongation).

To use Fluidigm™ for monitoring of expression in ESCs, samples were preamplified by 14 cycles of PCR. The preamplification mix constisted of 2 µl of cDNA, 10 µl of SYBR GREEN Mix (Fermentas), 2 µl of preamplification primer mix consisting of 48 primers (500 nM each) and 6 µl of water. PCR was performed as follows: 10 minutes of initial denaturation at 95°C to activate DNA polymerase, and 14 cycles of 95°C for 15 seconds (denaturation), 57°C for 4 minutes (annealing), and 72°C for 30 seconds (elongation). For preamplification a Bio-Rad T100 cycler was used. Preamplified samples were freezed in minus 20°C immediately to inactivate the polymerase.

## 6.8 Fluidigm™ experiment

Microfluidic Fluidigm array was used according to the manufacturer's protocol. Briefly, a primer mix consisted of 3 µl of 8 µM primers mixed with 3 µl of loading buffer (Fluidigm). A reaction mix consisted of 2.6 µl 10-times diluted preamplified sample, 3 µl of iQ mastermix (Bio-Rad) and of 0.1 µl 10-times diluted ROX fluorescent internal standard (Invitrogen). Primer and sample mixs were vortexed and 5 µl of each was loaded on PCR array. PCR was run for 35 cycles of 95°C for 15 seconds (denaturation), 59°C for 20s (annealing), and 72°C for 30s (elongation).

## 6.9 Quantitative PCR

Primer optimizing experiments on mESC were performed using a Roche Light cycler 384. The reaction was performed in total volume of 10 µl and consisted of 5 µl of 2x Maxima qPCR SYBR GREEN Mix (Fermentas), 2 µl of a 2 µM primer mix (consisted of reverse and forward primer), and 3 µl of cDNA. For the qPCR reaction, 3 µl of 40x diluted cDNA were used as a template. The template was obtained by reverse transcription of RNA, which was isolated from HEK293, NIH3T3 and mESC cells. The PCR was performed as follows: 95°C for 5 minutes to activate hot start polymerase, 95°C for 15 s for denaturing of template, 60°C for 10 s for primer annealing, synthesis phase was run at 72°C for 10s. Melting curve analysis and $C_T$ values were calculated using the original Roche software.

Expression of lincRNAs during OZT was assessed using a Stratagene Mx3000P lightcycler. The total volume of reaction was 10 µl. The reaction was performed in total

volume of 10 µl and consisted of 5µl of 2x Maxima qPCR SYBR GREEN Mix (Fermentas), 2 µl of 2 µM primer mix (consisted of reverse and forward primer), 0.5 µl of cDNA from reverse transcription of individual GV, MII, 1C and 2C and 2.5 µl of water (Fermentas). The experiment was performed as follows: 95°C for 10 minutes to activate hot start polymerase, 95°C for 15 s for denaturing of template, 60°C for 20 s for primer annealing and synthesis phase was run at 72°C for 30s. Fluorescence was measured at the end of synthetic phase during each cycle of the PCR. Melting curve analysis was performed for detailed analysis of specificity. $C_T$ values were calculated by the original software of the Mx3000P.

## 6.10 qPCR data processing

$C_T$ value corresponds to the cycle, in which amount of fluorescence crosses an artificially defined treshold level. In Fluidigm™ Biomark experiment, final $C_T$ values were calculated using the original Biomark software (Fluidigm). In lincRNA and mESC experiments $C_T$ values were calculated by MxPro software (Stratagene). In mESC experiment, $C_T$ values were calculated by Light cycler 384 software (Roche). Expression of mESCs was normalized to expression of housekeeping gene *Actb*. Raw data ($C_T$ values) were used for visualization of maternal mRNA degradation. Heatmap analysis of early embryos and oocytes was performed using GeneX software (MultiD), which calculates similarity level between samples. Clustering analysis of embryonic stem cells was performed in MultiExperimentViewer software (MEV4) using Manhattan clustering method. Principal component analysis (PCA) was selected as a statistical method for single-cell datavisualization. PCA of early oocytes and early embryos was performed in R environment by using princomp and biplot command line. To perform PCA of maternal and zygotic genes Ade4TkGUI package was used, PCA of embryonic stem cells was performed in GeneX (MultiD).

## 6.11 Identification of ncRNAs

Affymetrix MOE430 microarray platform was chosen for analysis because it should detect many ncRNAs. The reason is that the Affymetrix MOE430 microarray design was based on transcripts annotated in the Unigene database, the first attempt to

systematically annotate mammalian transcriptomes (Miller et al, 1997) (http://www.ncbi.nlm.nih.gov/unigene). The Unique database clustered cloned transcripts and anonymous expressed short tags (ESTs). Thus, long ncRNAs would be included in Unigene if they were cloned. These transcripts were poorly annotated and stored in the database regardless to their protein coding coding potential. Affymetrix MOE430 arrays were used for ncRNAs for identification of oocyte-specific and zygotic ncRNAs (Zeng & Schultz, 2005). Next, Ensembl (version 55) genome annotation was used to associate each individual probeset with targeted corresponding transcript.

Ensembl (version 53) database was used as an external reference for Affymetrix probes. Analysis was performed in R environment. Probes corresponding to intronic, intergenic, exonic and intronic regions in antisense orientation by full length of the probe were used for the analysis. Probes, which corresponded to exonic sequences in sense orientation were removed from original Affymetrix data (Affybatch) by using script written by Dr. Jenny Drnevich (University of Illinois). Probes, which flanked two regions or probesets, which had only one probe left after filtering, were removed from the analysis.

## 6.12 Microarray data analysis

Affymetrix microarray technology assigns expression based on intensity of fluorescence, which is generated after the probe interacts with the transcript. The intesity of the probeset is calculated by amount of intensity from all individual probes in the probeset. Affymetrix gene expression microarrays use probesets as main working units. Specificity of each probeset is determined by 12 oligonucleotides, called probes, each which is 25 nt long. Analysis of microarray data was performed in R environment by using bioconductor software. Microarray raw data (CEL files) were loaded in R. Data were normalized using GCRMA, signals from probesets and environment was corrected by using GC robust multiarray averaging (GCRMA) algorithm from GCRMA package to calibrate microarrays. AFFY package was used for loading and processing of the probesets. Rowttest algorithm from GENEFILTER package was used to determine differentially expressed genes and ncRNAs. Rowttest algorithm calculated intensity of differentially expressed probesets using comparison of between GV oocytes

and 2C embryos, and between 2C embryos and. 2C embryos treated with α-amanitin. Package ANNOTATE was used for annotation of Affymetrix probesets. Selection of zygotic/maternal genes in analysis of promoter regions was performed via Present/Absent/Marginal calling using MAS.5 package, because it statistically defines significant intensity of the probeset. Probesets, which were classified as present or absent at least by three of the replicates were taken for the analysis. Probesets classified as marginal were not included in the analysis. All necessary packages were downloaded and manipulated from bioconductor website (www.bioconductor.org).

# 7 Results

## 7.1 Identification of non-coding RNAs during oocyte-to-zygote transition (OZT)

OZT is a highly organized process. Activity, processing, and localization of mRNAs of protein coding genes were shown to have positive and negative effects on OZT (Chen et al, 2011; Mutter et al, 1988; Oh et al, 1997). Since the role of long ncRNAs in OZT is unknown, I decided to explore which long ncRNAs might contribute to OZT. I took the advantage of Affymetrix MOE430 arrays to identify maternal and ZGA long ncRNAs.

I expected to detect many ncRNAs, because Affymetrix designed probes based on clustered EST sequences deposited in the Unigene database (Miller et al, 1997). These sequences were included in Unigene regardless of their protein coding potential. Therefore it was likely that some of the probes would detect lincRNAs and long ncRNAs (refer to the 6.12 section). Complete reannotation of Affymetrix MOE430 microarray probes based on the latest transcriptome annotation (Ensembl, version 55) revealed that 14491 probesets on the array detect ncRNAs. As expected, most (~ 70%) of the ncRNAs are detected by the MOE430B platform of all ncRNAs (Table 1). This is due to the fact that MOE430A was largely used for detecting known protein coding genes.

| Array | All probesets | ncRNAs probesets |
|---|---|---|
| MOE430A | 22690 | 2270 |
| MOE430B | 22575 | 12221 |

**Table 1 Distribution of ncRNAs on Affymetrix MOE 430 platforms**

Affymetrix MOE430A chip contains 22690 probesets; 2270 of those detect ncRNAs. MOE430B contains 22575 probesets a significant fraction of those (12221) detect ncRNAs.

Next, I focused on identifying ncRNAs, differentially expressed between GV oocytes and 2-cell embryos, as these could play a role in regulation of zygotic genome activation. I focused on two types of ncRNAs: maternal and ZGA. Maternal ncRNAs are expressed in the oocyte and become degraded in the embryo. ZGA ncRNAs are expressed in early embryos. Altogether, I identified 143 maternal ncRNAs, whose intensity declined 2-fold between GV oocyte and 2C embryos. I also found that 223 of ZGA ncRNAs were detected in 2C embryos and were sensitive to α-amanitin, suggesting that these ncRNAs are transcribed from the zygotic genome. Intensity for these ncRNAs declined at least 2-fold after the treatment with α-amanitin (Figure 8).

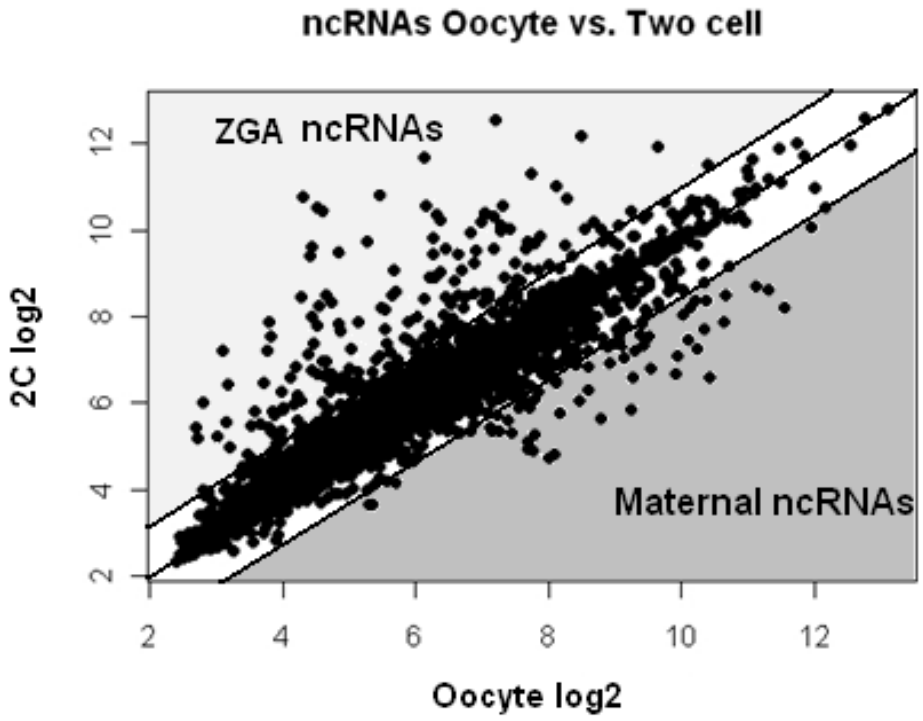


**Figure 8 Oocytes and embryos contain many ncRNAs**

Affymetrix detects many ncRNAs. Plot of probeset hybridisationintensity of ncRNAs between in germinal vesicle oocytes (GV) and 2-cell embryos (2C) reveals that 2-cell embryos generate more ncRNAs. Each dot represents hybridization intensity of one ncRNA.

### 7.1.1 Maternal and ZGA ncRNAs during OZT

Further characterization of ncRNAs showed that most of the maternal ncRNAs are generated from intergenic regions (65%), whereas the least represented class of ncRNAs are derived from intronic sequences in antisense orientation (5%) (Figure 9).

Majority ZGA ncRNAs is generated from intergenic (52%) and intron (42%) sequences. A small population of ncRNAs comes from exon (4%) and intronic (3%) sequences in antisense orientation (Figure 9). Interestingly, the comparison of maternal and ZGA profiles indicates that the absolute relative amounts of intronic sense ncRNAs raises 20% in the oocytes to 41% in the 2-cell embryos. This could be caused by higher rate of transcription of protein coding genes in 2-cell embryos, which generates more nascent transcripts, which would also be detected by these probesets. Next, I selected two maternal ncRNAs and validated their maternal status by qPCR. Both selected ncRNAs were detected in ovarian oocytes (GV), ovulated oocytes (MII) and are present in the zygote (1C), but they are absent in 2-cell embryos, suggesting that these ncRNAs are truly maternal (Figure 10). Taken together, I have identified 143 oocyte-specific and 221 ZGA long ncRNAs, which can potentially interfere with OZT and can contribute to the ZGA. The role of these ncRNAs during OZT and in the zygotic genome activation will be subject for future research.

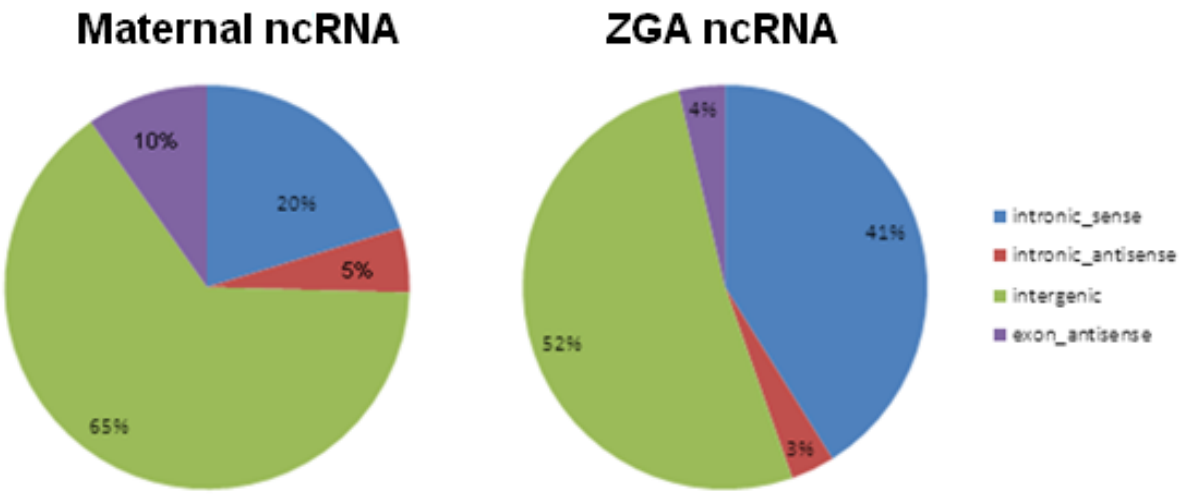


**Figure 9 Distribution of maternal and ZGA ncRNAs**

Analysis of ncRNA distribution in the genome probesets reveals that most of the maternal ncRNAs are generated from intergenic regions (65%), intronic regions (20%), exon antisense (10%) and intronic antisense (5%), respectively. Most of the ZGA ncRNAs are generated from intergenic regions (52%), intronic regions (41%). Only a small fraction of ncRNAs is generated from intronic antisense (3%) and exon antisense (4%) regions of known protein coding genes.

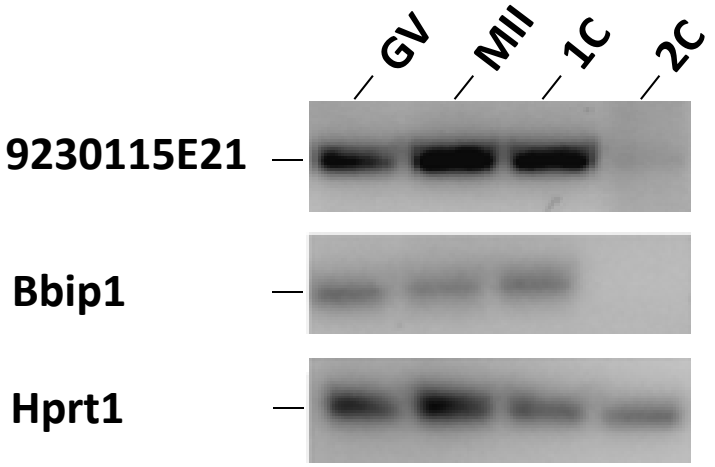


**Figure 10 Validation of maternal ncRNA expression**

Real-time PCR confirms that two lincRNAs, *Bbip1* and *9230115E21* are maternal ncRNAs as they are present in GV, MII, 1C and disappear in 2C suggesting that these are maternally expressed ncRNAs. *Hprt1* was used as endogenous control and should be expressed in all samples. GV stands for germinal vesicle, MII stands for metaphase II oocyte, 1C stands for 1-cell embryo and 2C stands for 2-cell embryo.

## 7.2 Monitoring of gene expression during oocyte-to-zygote transition

Oocyte-to-zygote transition is a complex process with a dynamic transcript turnover. Our laboratory needed an assay for simultaneous gene expression analysis in individual oocytes and early embryos. Such assay would provide several benefits. First, it would allow to monitor transcriptome dynamics during OZT. Second, it would allow to study genes of interest, including ncRNAs, during OZT. Third, this assay could be used for rapid phenotyping of experimentally manipulated preimplantation embryos. To develop the assay, I took the advantage of qPCR-based platform Fluidigm™, which allows for analyzing expression of 48 genes in 48 samples in a single experiment. Development of the assay consisted of three main steps: i) selection of genes, ii) optimalization of primers, iii) optimalization of the preamplification step.

### 7.2.1 Gene selection

To develop a method for simple and rapid phenotyping of cells during OZT, I selected marker genes for different phases of OZT, such as maternal mRNA degradation, zygotic genome activation and pluripotency establishment during OZT. I focused on genes, which would be informative markers of one of the three categories. To select marker genes, I combined literature search with analysis of microarray data.

I selected several maternal genes, which exhibited phenotype and had significant consequences on embryo development: *Spin1*, *Zar1*, *Zp3*, *Nlrp5*, *Mos* and *Nobox*. *Zar1* is a maternal factor, which is crucial for oocyte-to-zygote transition (Wu et al, 2003). *Zar*$^{-/-}$ females are infertile and embryos generated from these females are arrested at the 1-cell stage (Wu et al, 2003). ZP3 is a gene expresed exclusively in mammalian oocytes. ZP3forms a receptor for interaction with the sperm and its depletion leads to 100 % infertility (Litscher et al, 2009). *Nlrp5* (*Mater*) is one of the first identified maternal effect genes in mammals (Tong et al, 2000). Female knockout of this gene are infertile and embryos arrest in the preimplantation development (Tong et al, 2000).

The best characterized ZGA markers are *Eif1a* and two repetetive elements *MuERV-L* and *IAP*, whose expression increase several fold during ZGA (Davis & Schultz, 2000; Kigami et al, 2003; Svoboda et al, 2004). Other genes, such as *Hdac1*, *YY1* and were selected based on published microarray data, generously provided by Zeng and her colleagues (Zeng & Schultz, 2005). For the complete list of genes included in the assay and their characterisation please refer to the Appendix.

### 7.2.2 Primer testing & Preamplification

I tested the quality of primers using three independent templates (mESCs, NIH3T3, and ovarian templates), in which pluripotency, differentiation and maternal genes should be expressed (Figure 11). In addition to primer quality, I had to address the problem of sufficient template amounts. Despite numerous advantages, which Fluidigm™ system offers, it requires certain quality and quantity of the sample, which can be a challenge when it comes to single oocytes and single embryos, which contain up to hundreds of picograms (Piko & Clegg, 1982). To overcome this problem, I had to include a preamplification step, which would produce sufficient template to perform Fluidigm™ array experiment. I developed a protocol, which allows for an amplification

of 48 templates during one reaction by using a mix of forward and reverse primers of all tested genes.

To test the protocol, I used three different templates– cDNA from embryonic stem cells, p19 cells and oocytes, and tested effects of preamplification. Selection of templates was based on genes in the assay. Oocyte template was selected to test monitoring of maternal genes. Templates from mESCs and p19 are used for testing the preamplification of pluripotent and zygotic transcripts in the assay. A small aliquot of each template for the preamplification reaction was used. Preamplification was primed by a primer mix consisting of all 48 genes (refer to the 8.7 section), the nature of amplification had therefore a similar pattern as a usual qPCR. The preamplified sample was then diluted ten times to reduce probability of primer dimers in the following qPCR. The overall behaviour of preamplification was tested for individual genes by qPCR method (Figure 12).

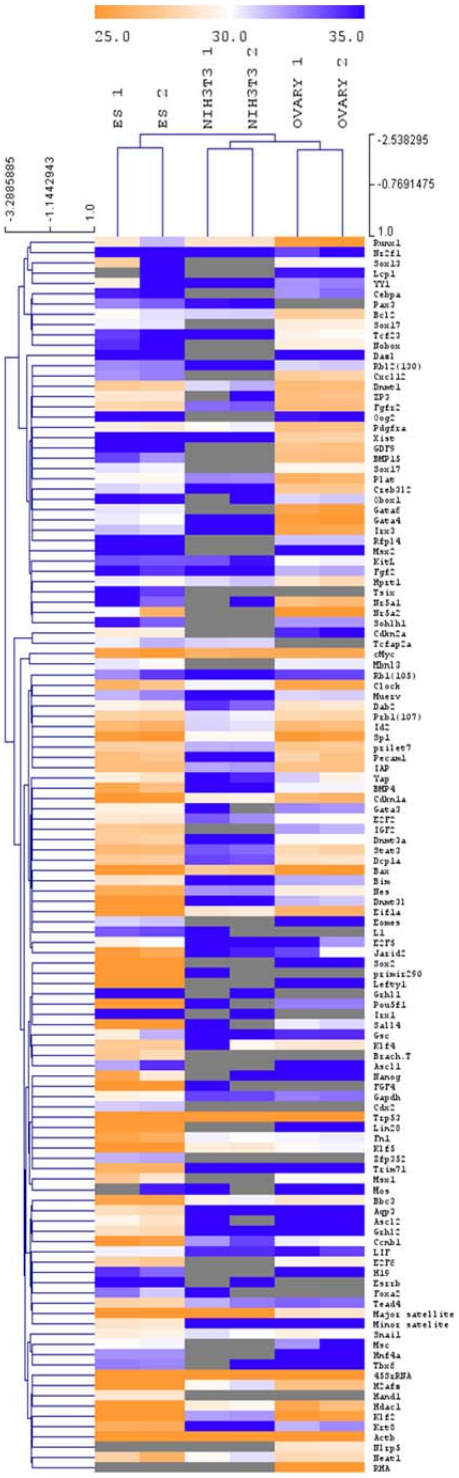


**Figure 11 Primer test**

A heatmap generated from raw expression values of the genes used in the assay. Three different cDNAs were selected for testing. NIH3T3 cDNA was used as a negative control for pluripotency and some differentiation genes. ES cDNA was used for detecting pluripotency markers. Ovarian cDNA was used for testing the expression of maternal genes and some differentiation genes. Raw expression of the genes is indicated by the color gradient on the top of the heatmaps. Highly expressed genes are represented by orange color, whereas low-expressed genes are marked by blue color. Grey color represents genes, whose

expression was not detected as expected from their described roles. Analysis demonstrated that some genes are expressed in mES cells and are absent in NIH3T3.

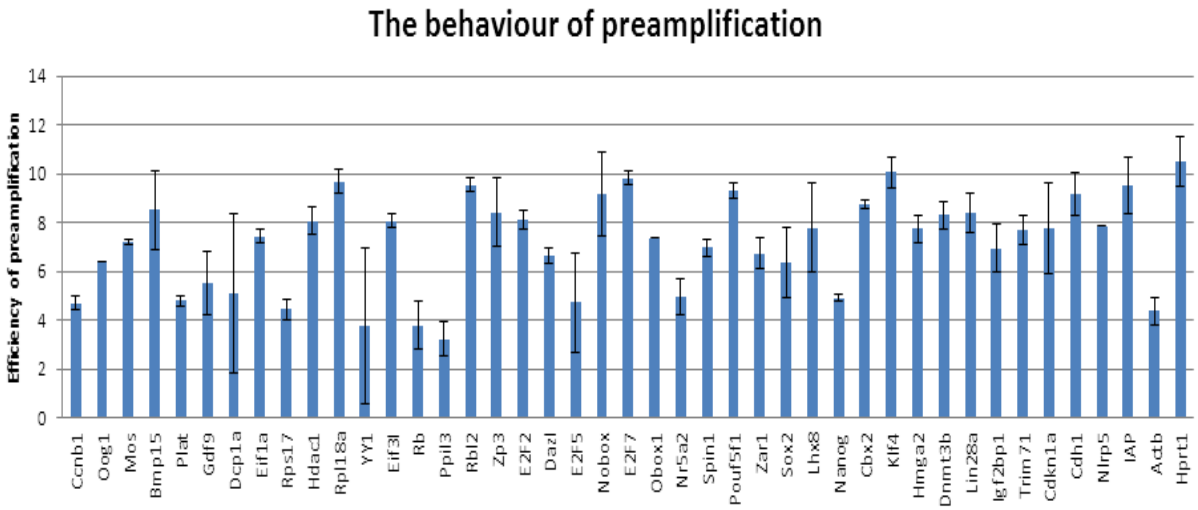


**Figure 12 Efficiency of preamplification tested on p19, mESC and oocyte templates**

Preamplification was tested on 43 genes, included in the assay. The graph shows that the preamplification has different behavior for different genes, demonstrated by height of blue bars. The height of blue bar refers to efficiency of preamplification reaction and is calculated as a ΔCt difference between the mean of nonamplified and preamplified p19, mESCs and oocyte templates. Genes, for which preamplification efficiently increased number of template of molecules are demonstrated by higher bars. Error bars = standard error of the mean (SEM).

Ideally, preamplification would show a constant increase of the amount of template molecules for all genes. This effect would be manifested by identical heights of blue bars. I observed that preamplification introduces certain amount of noise to the experiment, which varies between genes and which is most likely given by efficiency of primers (demonstrated by different heights of bars). However, there is also a second source of noise, which is caused by variability among templates, demonstrated by different length of error bars. Overall, the preamplification protocol has worked sufficiently well for the Fluidigm™ use.

Upon assessing preamplification noise per each gene in the assay, I proceeded to generating data by Fluidigm™ 48.48 PCR array. For the first experiment, I used 12 GV oocytes, 11 MII oocytes, 6 1-cell embryos, 11 2-cell embryos and 5 8-cell embryos. Preamplification was performed according to the optimized protocol and final results are shown in the Figure 13. One sample of 8-cell embryo, one sample of 2-cell embryo and two samples of GV oocytes were excluded from analysis because of poor quality.

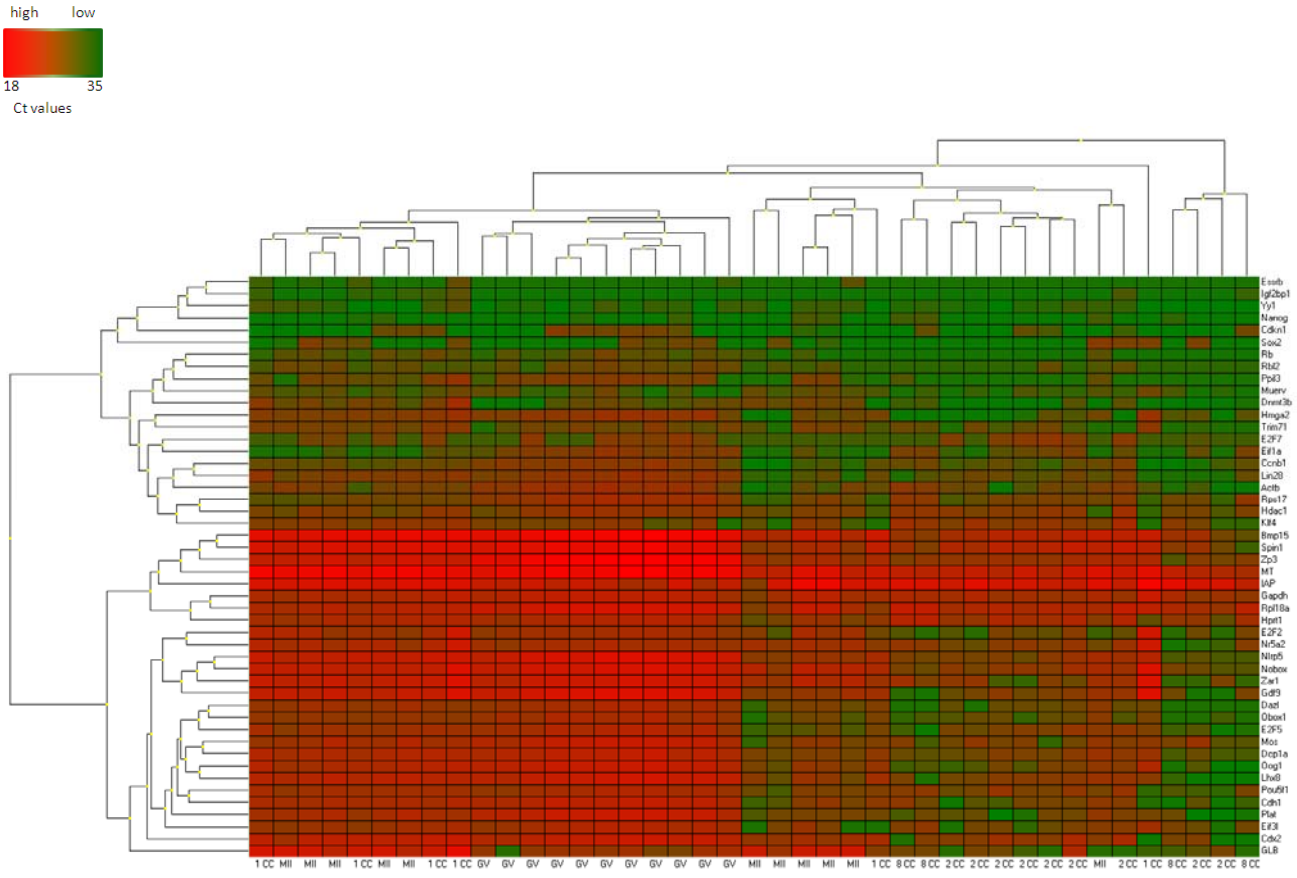


**Figure 13 Analysis of gene expression in individual mouse oocytes and embryos**

Heatmap display of Fluidigm 48.48 array analysis of gene expression in oocytes and early embryos. Data from individual samples are shown in columns, whereas rows represent individual gene expression. Level of expression is indicated by colors. Highly expressed genes are marked by red color, whereas low-expressed genes are shown by green color. The clustering analysis divides dataset into two groups. One is defined by presence of GV oocytes while 8-cell (8CC) and 2-cell (2CC) embryos are formed in the other one. Ovulated oocytes (MII) and 1-cell embryos (1CC) were found in both of the groups.

Clustering analysis separated the samples into two distinct groups – the first one consisted of all GV oocyte samples and many fertilized (1CC) and unfertilized eggs (MII). The second cluster consisted of early embryos and ovulated oocytes, demonstrating that transcriptome change between GV oocytes and embryos. This is likely caused by differences in maternal and embryonic transcriptome. The highest transcriptome similarity was found for 1-cell embryos and MII oocytes. As expected, maternal genes were found abundantly expressed in oocytes and declined in embryos, suggesting that the assay is suitable for maternal mRNA degradation (Figure 14). However, I observed that the assay poorly detected the activation of zygotic genes.

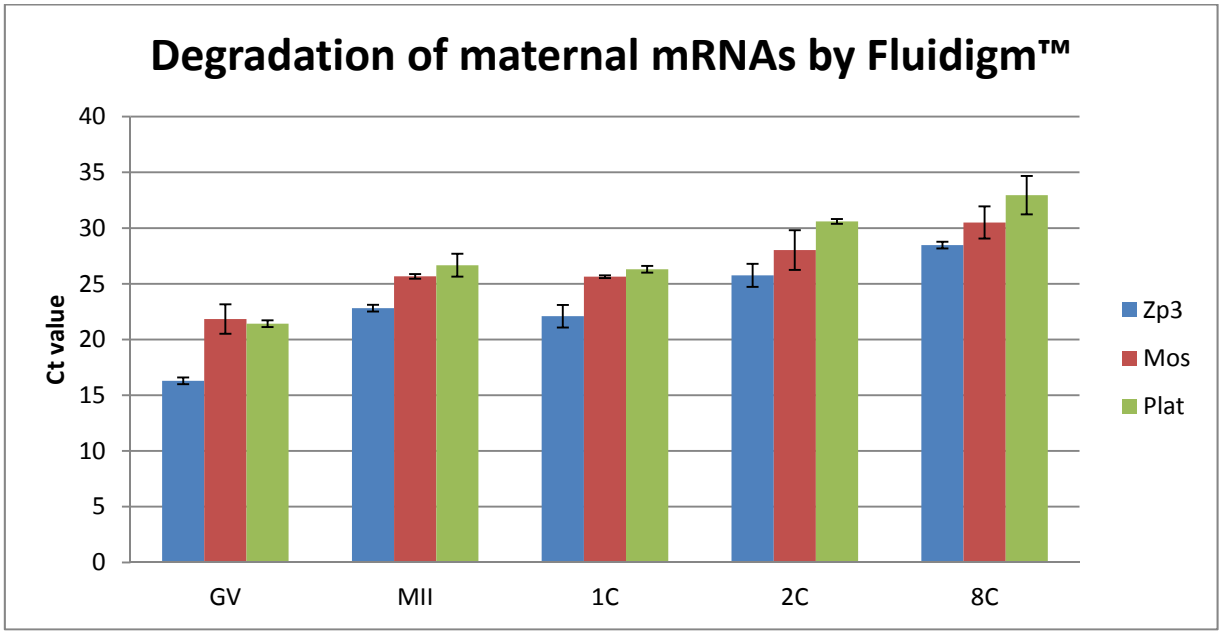


**Figure 14 Maternal mRNA degradation revealed by selected markers**

Three typical maternal transcripts were selected to illustrate dynamics of mRNA degradation during OZT. Individual stages of OZT are shown on x axis, whereas y axis represents raw expression ($C_T$ values) of the genes. *Zp3*, *Mos* and *tPlat* are abundantly present in GV premature oocytes (they have low $C_T$ values), their expression declines in 2C and 8C embryos ($C_T$ values increases). Surprisingly, the assay shows significant differences in expression between GV and MII oocytes for *ZP3* and *tPlat*, which is in contrast with the microarray data. Please note that $C_T$ values refer to geometrical amplification of the template, thus the lower the $C_T$ is, the more abundant the corresponding transcript is. Error bars = SEM

To further analyze the results, one can compare profiles of individual stages during OZT (Figure 15). As expected, the more distant the samples were in development, the higher difference in profiles was obtained. This demonstrates that the assay monitors dynamic changes of the transcriptome during OZT where zygotic genome activation is superimposed on maternal mRNA degradation. Notably, MII oocytes and 1-cell embryos were the most similar samples, suggesting that fertilization has minimal impact on degradation of marker genes.

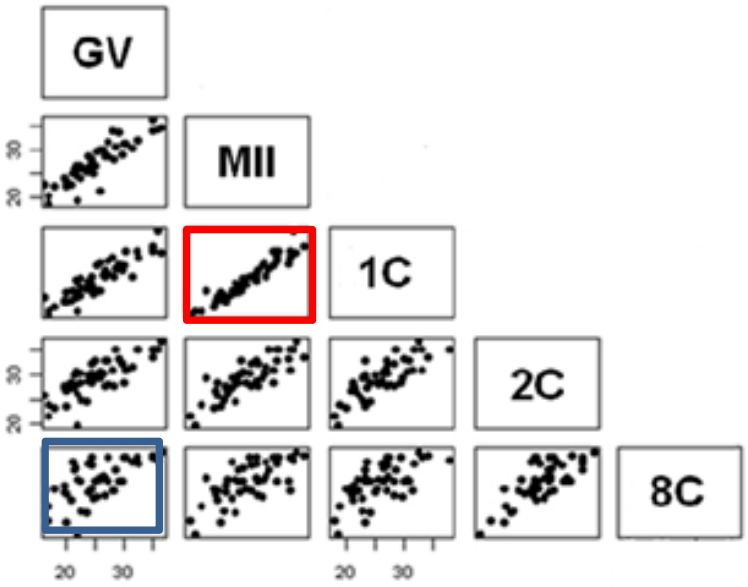


**Figure 15 Transcriptome remodeling during OZT**

Plot of the averaged raw data from different stages shows that Fluidigm 48.48 assay reveals transcriptome remodeling during OZT. Similarity of the samples is manifested as diagonal distribution of expression data. Most similar were MII and 1C samples (marked by red). Likewise, distance between the most developmentally distant samples, GV oocytes and 8C embryos (marked by blue), is also expected because maternal mRNAs are degraded and transcription of embryonic genome is already initiated at 8-cell stage.

Next, I performed statistical clustering of the data to have a more precise measure of similarity and difference among the data. I used principal component analysis (PCA), which was introduced as an alternative method for visualization of high-throughput data. By this method one can distribute samples according to the major variables in the

dataset and visualize the results in a user-friendly way (Figure 16). Plotting of the results showed three major groups of expression data in the dataset. 2-cell and 8-cell embryos represented one group, GV oocyte samples clustered separately. MII oocytes showed another cluster, creating a third group suggesting that these cells are in the transition process and have different signature from embryos and ovarian GV oocytes (Figure 16). These data show that a profile of 48 genes from samples yields enough information to allow for its classification into one of several biologically relevant categories. These results are consistent with data from recent study, which shows that embryos and oocytes have distinct transcriptome signatures (Tang et al, 2011).

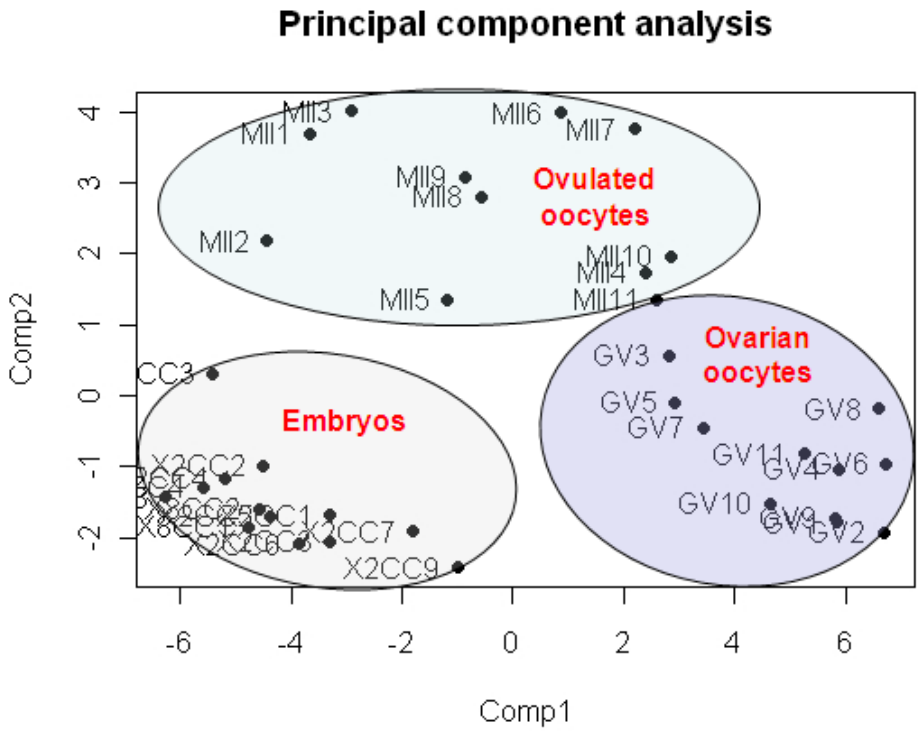


**Figure 16 PCA reveals distinct signatures in embryos and oocytes**

PCA scoring suggests that there are three distinct patterns in the dataset. The most significant difference along the first experiment was between ovarian oocytes and cleaved embryos. This shows that transcriptomes of ovarian oocytes exhibited ovarian oocytes which scored in a separate cluster, whereas 2-cell and 8-cell embryos resulted in clustered together in a distinct group suggesting that transcriptome between ovarian oocytes and early embryos have major differences. Ovulated oocytes (MII) scored in between ovarian oocytes and embryos indicating that these are already in transition process. Interestingly, ovulated oocytes exhibited a high level of variability. 1-cell embryos were excluded from the analysis.

Next, I compared my results with data from microarrays. I used the data from Affymetrix MOE430 microarrays, which monitor gene expression during OZT (Zeng et al., 2004). Relative gene expression of maternal genes was calculated from GV oocyte and 2-cell embryo data. The 2-cell embryos were selected, because at this stage maternal mRNA should be largely degraded. The results show good concordance of Fluidigm™ and Affymetrix data (Figure 17). Most of the genes (represented by dots in the Figure 17) scored in the same quadrant of the plot, suggesting that Fluidigm™ can be a method-of-choice for maternal mRNA degradation. Some genes showed discrepancy between Fluidigm™ and Affymetrix microarrays. This discrepancy might be caused by different priming of the reverse transcription in both experiments. Reverse transcription in Fludigm™ experiment was primed by random hexamer primers, whereas oligo(dT) were used for priming in Affymetrix experiment. These differences in cDNA priming may cause discrepancies because deadenylation will effect oligo(dT) primed mRNAs during reverse transcription.

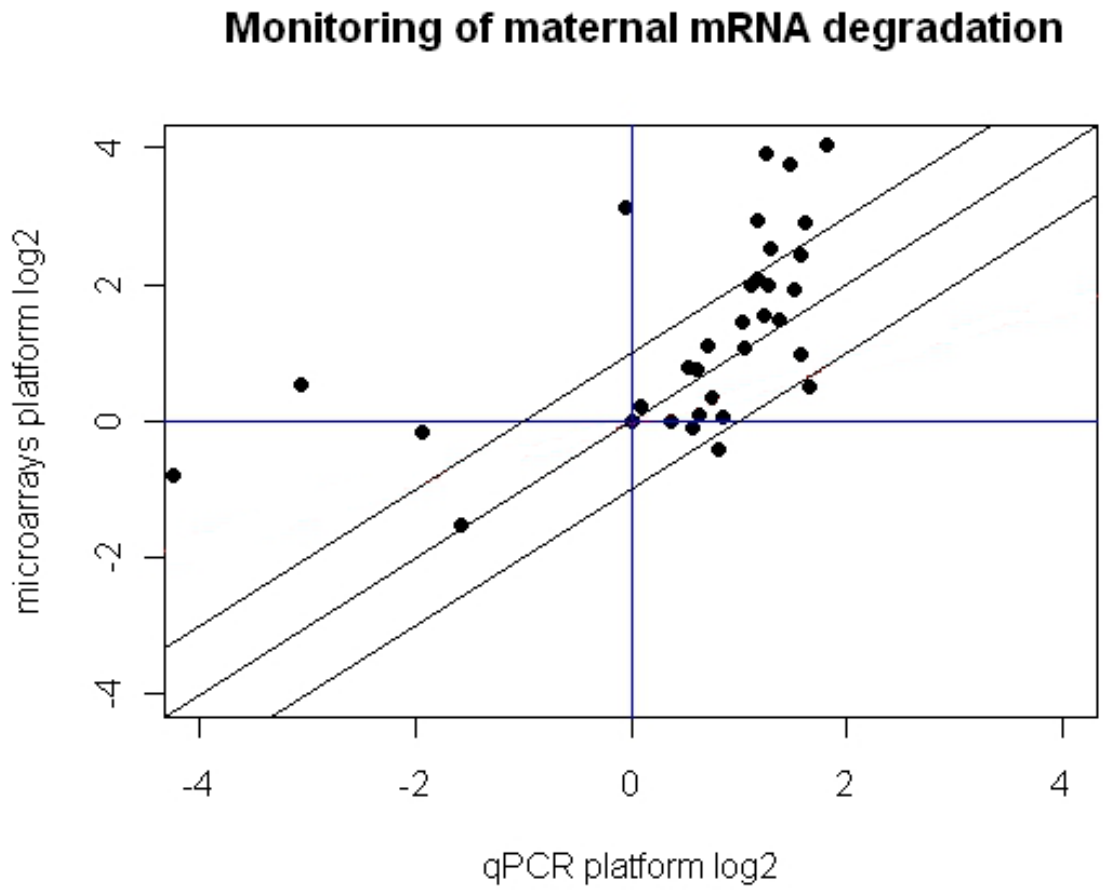


**Figure 17 Maternal mRNA degradation revealed by microarrays and qPCR arrays**

Comparison of monitoring of expression of maternal genes by qPCR Fluidigm™ platform and Affymetrix microarrays on genes. Results show a general concordance as well as a clear difference in dynamic ranges of mRNA levels measured by both platforms. Since most of the genes (represented by dots) locate into upper-right and lower-left quadrants data generated by Fluidigm™ are comparable with microarray platform for monitoring of maternal degradation. The fold-change of maternal genes from Fluidigm™ was normalized to *Gapdh* expression. Similarly, fold-change of Affymetrix data was calculated by normalization of intensity of probeset hybridization intensity to basal *Gapdh* intensity.

## 7.3 Analysis of pluripotency and differentiation in ESCs

ESCs can be cultured *in vitro* in undifferentiated state. Today, they are being used for studying pluripotency and differentiation by RNAi and by small inhibitors. To analyze gene expression in ESCs on large-scale, one can use microarray platform or NGS. These experiments are, however, costly. Therefore, our laboratory was looking for a simpler and cheaper alternative. I decided to develop a qPCR array assay for analysis of transcriptome in ESCs. The assay would use 48 diagnostic markers for pluripotency and differentiation analyzed on Fluidigm 48.48 array. Development of the assay consisted of: i) gene selection and ii) primer testing.

To develop the assay for gene expression in ESCs, I had to select relevant pluripotency and differentiation markers. I used the markers, which were already used in the literature (Guo et al, 2010; Chen et al, 2008). Guo and his colleagues identified and tested marker genes associated with pluripotency establishment and differentiation during preimplantation development in mouse (Guo et al., 2010). As the assay is based on qPCR, I designed and tested the primers for the selected marker genes (refer to the section 8.2). Once I finalized the panel of marker genes, I continued to test reliability of the assay by analyzing expression of 48 genes in ESCs. We took the advantage of a collaboration with the group of Domingos Henrique from Portugal, who provided samples of ESCs cultured under different conditions (Table 2).

| Sample | Condition |
|---|---|
| 1 and 11 | Cultured in GMEM and LIF, expressing endogenous NANOG |
| 2 and 12 | Cultured in GMEM and LIF, low expression of NANOG |
| 3 and 13 | Cultured in GMEM and LIF, intermediate expression of NANOG |
| 4 and 14 | Cultured in GMEM and LIF, high expression of NANOG |
| 5 and 15 | Cultured in GMEM in the absence of LIF for 48h |
| 6 and 16 | Cultured in GMEM and LIF |
| 7 and 17 | Cultured with ERK inhibitors 1:10 |
| 8 and 18 | Cultured with ERK inhibitors 1:5 |
| 9 and 19 | Cultured with ERK inhibitors 1:2 |
| 10 and 20 | Cultured in iStem medium (2i inhibitors) |

**Table 2**

An overview of ECSc cultured under different conditions

We were particularly interested in quality assessment of samples cultured in the the iStem medium, in the samples cultured with presence/absence of LIF, and in samples with low/high protein level of NANOG. To explore whether the assay monitors differences in these samples, we performed heatmap analysis (Figure 18). The clustering analysis divided samples into three groups. We observed that samples with low and high protein expression of NANOG resulted in two distinct clusters. Samples with high NANOG (number 4) clustered together with samples cultured in iStem medium (number 10), indicating that higher concentration of NANOG promotes pluripotency program in ESCs. We observed significant transcriptome changes in samples cultured in the absence of LIF (number 5). In these samples expression of *Stat3*, *Esrrb*, *Klf4*, *Nanog*, *Rex1* was reduced in comparison with samples cultured with LIF (number 6). Conversely, LIF removal promoted expression of *Fgf5*, *Bcl2*, *Tcfap2a* and *Gata3*. This suggests that LIF removal itself has negative effects on pluripotency and initiates differentiation program.

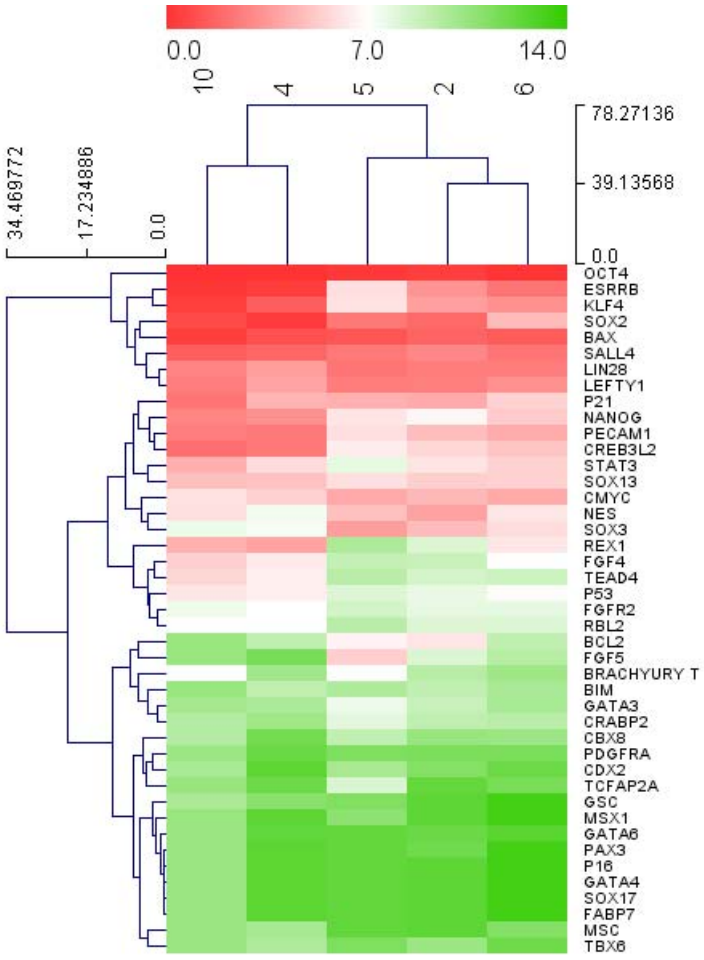


**Figure 18 Transcriptome analysis in ESCs cultured under different conditions**

The Fluidigm™ assay accurately monitors differences in the expression of pluripotency and differentiation marker genes in samples cultured under different conditions. The heatmap shows expression based on color gradient. Green color represents low expressed genes, whereas red color represents genes with high expression. ESCs cultured in iStem medium (number 10) and cells with high protein level of NANOG (number 4) resulted separated from embryonic stem cells cultured in the presence/absence LIF in medium (number 6 and 5, respectively) and cells, expressing low protein level of NANOG (number 2). The analysis shows that LIF removal promotes expression of differentiation genes, such as *Brachury T* and *Gata3*, which is demonstrated by change of color towards the red color. In contrast, LIF removal suppresses expression of pluripotency genes, such as *Nanog*, *Klf4* and *Stat3*. This change is manifested as a color change towards the green color. The expression of all genes was normalized to the endogenous *Actb*. The expression was calculated as the mean from two biological samples.

Results from PCA from the first experiment showed that biological duplicates (here denoted by one digit and two digit pairs, e.g. 5 and 15) shared highly similar pattern of gene expression (Figure 19). Interestingly, the distribution of the points in the figure reveals that samples, which were cultured in the absence of LIF (S5 and S15) have distinct signature from samples, which were cultured in the presence of LIF (S6 and S16). This suggests LIF removal have positive effect on initiation of differentiation. In contrast, samples cultured in a pluripotent medium iStem (S10 and S20) ended in the opposite direction than S5 and S15, indicating that iStem medium keeps cells in a truly pluripotent state. Apparently, the distribution of samples along the x axis can serve for monitoring of pluripotency in analyzed samples. Samples, which resulted on the right in PCA are more pluripotent that samples on the left. We conclude that the assay works and is prepared for monitoring the quality of ESCs and culturing conditions of ESCs.

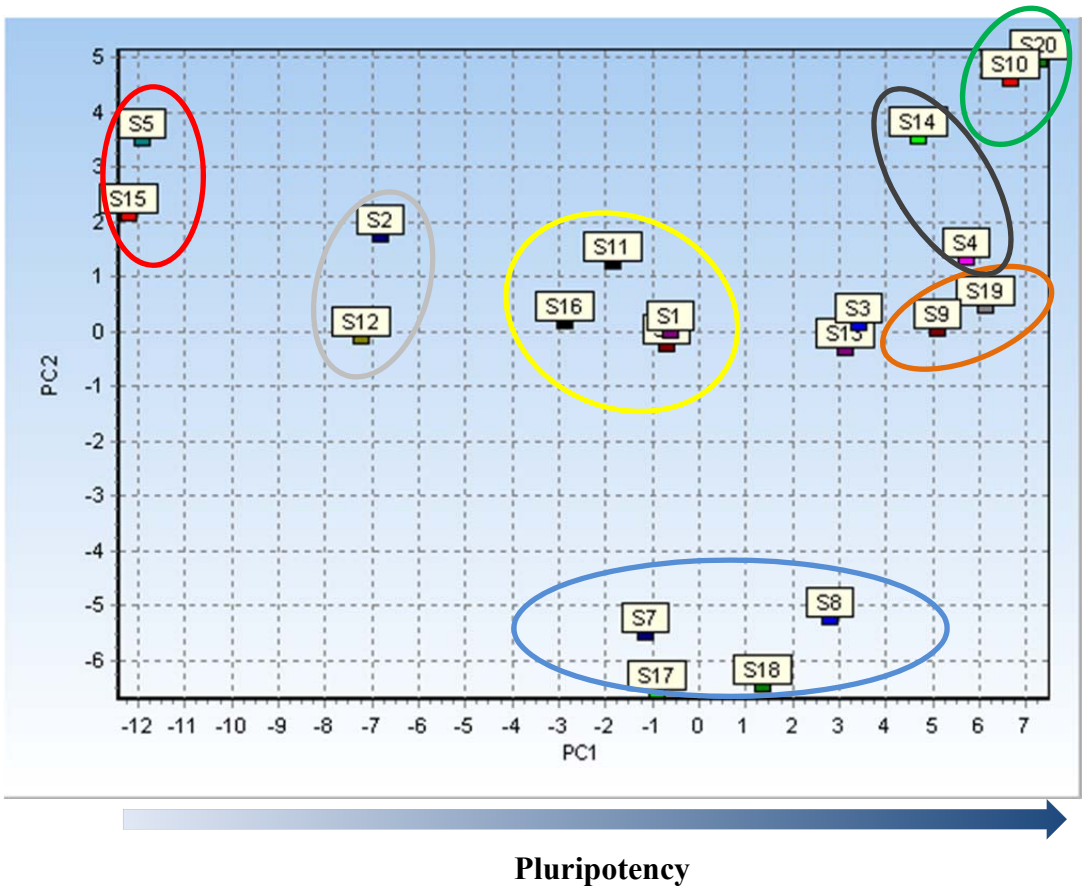


**Figure 19 PCA reveals differences in ESCs cultured under different conditions**

Results from Fluidigm assay reveals distribution along PC1 reflects pluripotent potential of analyzed cells. Cells, which resulted on the right side are more pluripotent than cells, which resulted on the left. Cells cultured in the absence of LIF (S5,S15) resulted in the opposite direction than ESCs cultured with LIF (S6, S16).

# 8 Discussion:

During my thesis I applied high-throughput methods to analyze early events during activation of the pluripotent zygotic program during mammalian development. My thesis was divided into two parts: identification of ncRNAs during OZT and analysis of transcriptome dynamics during OZT and in ESCs.

## 8.1 ncRNAs during OZT

Taken together, I have identified 143 putative maternal ncRNAs and 223 ZGA ncRNAs by reannotation of Affymetrix MOE430 microarrays. The identification of ncRNAs opens questions and directions for the future research. First, it is necessary to evaluate how much of the data are computational artifacts. The easiest way how to validate the data would be via qPCR method. So far, I verified two differentially expressed maternal lincRNAs.

Second, it is important to elucidate how much information is encoded by these ncRNAs and what is the actual function of these ncRNAs. While the function of long ncRNAs during OZT is unknown, I speculate that they can interfere with OZT by i) promoting of degradation of maternal mRNAs, ii) contributing to the activation of zygotic program. There is evidence that some human lincRNAs can interact and promote degradation of a mRNA (Gong & Maquat, 2011). It is unknown whether a similar mechanism operates in mice for degradation of maternal mRNAs during OZT.

Microarray analysis revealed that 2-cell embryos contain 2607 active genes, which are transcribed from mouse genome (Zeng & Schultz, 2005). My data suggest that 223 (9%) of these transcripts do not encode any protein, suggesting that these ncRNAs might contribute to ZGA. There is increasing evidence that ncRNAs, such as lincRNAs, are implicated in initiation of transcription of other genes (Orom et al, 2010; Sessa et al, 2007). In addition, the knockout studies in embryonic stem cells indicate that some lincRNA can play an essential role in the circuitry controlling pluripotency and differentiation (Guttman et al, 2011). While embryonic stem cells are developmentally related to 2-cell embryos, what is the actual function of long ncRNAs on ZGA remains elusive. To address this question it would be necessary to apply loss-of-function approach. The simplest strategy would be based on obtaining specific siRNAs and injecting them into oocyte or 1-cell embryo. It would also be worth of

further investigation of the target genes, which might be activated by lincRNAs. To address this question, one could compare nucleotide sequences of lincRNAs and regions upstream of lincRNA TSS. In addition, it is becoming evident that mammalian cells contain large fraction of ncRNAs, which come from permissive transcription and from splicing events (Valen et al, 2011). Therefore it is important to assess how many ncRNAs are transcriptional noise.

There is evidence that some long ncRNAs promote generation of human iPS cells (Loewer et al, 2010). Thus, it would be interesting to test what is the direct effect of maternal lincRNAs on reprogramming of mouse somatic cells into iPSCs. One could test this ability by transfecting somatic cells, such as MEFs, by vector bearing lincRNAs sequences under conditions of iPSC production.

## 8.2 Gene expression analysis during OZT and mESCs

My experimental work largely focused on development of assay for rapid phenotyping of mouse oocytes and embryos by using a novel real-time PCR-based platform Fluidigm™. The assay uses 48 markers for maternal mRNA degradation, zygotic genome activation and pluripotency establishment and could be used for distinguishing between individual stages during OZT. Based on the results, I conclude that the assay allows for monitoring a signature of maternal mRNA degradation in individual cells. My results are consistent with the data from a study, in which authors employed NGS and PCA for cells during OZT (Tang et al, 2011). However, it seems that the Fluidigm™ assay is suboptimal for detailed analysis of ZGA and pluripotency establishment in the early embryos. One of the possible explanation might be a low amount of RNA in 2-cell embryos, which is significantly reduced after waves of maternal degradation (Piko & Clegg, 1982). Such a low amount of mRNA would require stronger preamplification, which introduces more noise into the results. Moreover, increased preamplification simultaneously disturbs the analysis of highly-expressed genes which become out-of-range for detection by the Fluidigm™. Therefore, the system is suboptimal for detailed analysis of low-expressed genes.

Another possible explanation why the assay does not monitor ZGA properly could be suboptimal selection of ZGA markers. Our panel of ZGA genes contained well known markers of ZGA, such as *murine endogenous retrovirus-L* (*MuERV-L*), *intracisternal A particle* (*IAP*) (Kigami et al, 2003; Svoboda et al, 2004), *Yin Yan 1*

(*YY1*), *eukaryotic translation initiation factor 1A* (*Eif1a*) and *histon deacetylase 1* (*Hdac1*). According to our data, *MuERV-L* and *IAP* did not turn out to be reliable markers of ZGA, as they seemed to be expressed across all samples, which is in straight contrast with published data (Kigami et al, 2003; Svoboda et al, 2004). Therefore future efforts should aim at finding better ZGA-specific markers.

.

In future experiments, it would be interesting to analyze each of the individual cells in the early embryos and to determine the level of pluripotency and lineage commitment in individual cells during OZT and blastocyst formation. These findings would provide novel insights into understanding, when the pluripotency network is built up during development and also at what point individual cells start to differentiate. It would also provide a toolkit for analysis of ncRNAs, transcription factors and signaling pathways as one could treat embryos by various inhibitors and siRNAs.

To my knowledge, there is one recent publication studying lineage commitment and coherency of pluripotent network in individual cells in morula and blastocyst using Fluidigm™ (Guo et al, 2010). The authors applied Fluidigm for detailed analysis of three distinct population of cells in blastocyst. While, this study brings new light in the field, an important question how and when pluripotency network starts being built up in earlier stages of embryogenesis remains unanswered.

Employing the assay for quality monitoring of embryonic stem cells suggested that the assay reflects differences in transcriptomes of ESCs cultured in different media. Further characterization of our data suggests that NANOG promotes pluripotency state of the ESCs and expression of *Nanog* correlates with the expression of pluripotency genes, such as *Oct4*, *Sox2* and *Esrrb*. As expected, we observed that the absence of LIF in the culturing media initiates differentiation program. This was manifested by increased of expression of differentiation markers *Gata4* and *Tcfap2a*. LIF removal simultaneously inhibited transcription of pluripotency genes, such as *Stat3*, *Esrrb* and *Klf4*. This data are in line with the study, in which authors tested the effect of LIF removal. Their conclusion was that *Stat3* is the main effector of LIF-mediated pathway in murine ESCs (Cartwright et al, 2005). To our knowledge, we were the first group, which developed a qPCR array assay for quality assessment of murine ESCs.

## 8.3 Outlook

Understanding natural formation of ESCs as well as generation of induced pluripotent stem cells are of a high priority. Every novel factor, which improves efficiency of reprogramming and pluripotency maintenance can have a deep impact on future treatment of patients suffering from devastating diseases. Implication of induced pluripotent stem cells holds great promises in next generation of therapy, as one can imagine that it would be possible to obtain pluripotent stem cells from any patient's organ. Finally, these newly reprogrammed cells could be adjusted to patients, where they could replace not functional cells. This approach would revolutionize modern therapy as patient's would not have to wait long period of time for suitable donor of organs. And from point of immunology, patients would not have to use suppressive therapeutics to prevent reaction against host organ.

Conclusions:

- I have identified 143 maternal and 223 ZGA ncRNAs using Affymetrix MOE430 microarray. These ncRNAs might contribute to activation of zygotic program in the mouse.

- I have developed an assay for monitoring of oocyte-to-zygote transition. The assay can be used for phenotyping of individual cells during oocyte-to-zygote transition

- I have developed an assay for monitoring lineage commitment and pluripotency in mouse embryonic stem cells.

# 9 Appendix

| Number | ID | Official name | Reason |
|---|---|---|---|
| 1 | ActB | Actin b | Housekeeping gene |
| 2 | Bmp15 | Bone morphogenetic protein 15 | Maternal |
| 3 | Cbx2 | Chromobox homolog 2 (Drosophila Pc class) | miR-290 target |
| 4 | Ccnb1 | Cyclin B | Dormant maternal |
| 5 | Cdh1 | Cadherin 1 | Pluripotency network |
| 6 | Cdkn1 | Cyclin-dependent kinase inhibitor 1A (P21) | miR-290 target, senescence marker |
| 7 | Dazl | Deleted in azoospermia-like | Maternal |
| 8 | Dcp1a | Dcp1 decapping enzyme homolog A | Dormant maternal |
| 9 | Dnmt3b | DNA methyltransferase 3B | ZGA |
| 10 | E2f2 | E2F transcription factor 2 | Cell cycle regulator |
| 11 | E2f5 | E2F transcription factor 5 | Cell cycle regulator |
| 12 | E2f7 | E2F transcription factor 7 | Cell cycle regulator |
| 13 | Eif1a | Eukaryotic translation initiation factor 1A | ZGA |
| 14 | Eif3l | Eukaryotic translation initiation factor 3, subunit L | Meiotic mRNA degradation |
| 15 | Esrrb | Estrogen related receptor, beta | Pluripotency network |
| 16 | Gapdh | Glyceraldehyde-3-phosphate dehydrogenase | Housekeeping gene |
| 17 | Gdf9 | Growth differentiation factor 9 | Maternal |
| 18 | globin | Globin rabbit | External spike |
| 19 | Hdac1 | Histone deacetylase 1 | ZGA |
| 20 | Hmga2 | High mobility group AT-hook 2 | let-7 target, peaks at 8CC |
| 21 | Hprt1 | Hypoxanthine phosphoribosyltransferase 1 | Housekeeping gene |
| 22 | IAP | Internal A particule | ZGA |
| 23 | Igbp1 | Immunoglobulin (CD79A) binding protein 1 | ZGA 2C |
| 24 | Klf4 | Kruppel-like factor 4 (gut) | Pluripotency network |
| 25 | Lhx8 | LIM homeobox protein | Maternal |
| 26 | Lin28a | Lin-28 homolog A (C. elegans) | let-7 target |
| 27 | Mos | Moloney sarcoma oncogene | Dormant maternal |
| 28 | MT | Mouse transcript | ZGA 2C |
| 29 | Muerv | Murine endogenous retrovirus | ZGA 2C |
| 30 | Nanog | Nanog homeobox | Pluripotency network |
| 31 | Nlrp5 | NLR family, pyrin domain containing 5 | Maternal |
| 32 | Nobox | Oogenesis homeobox | Maternal |
| 33 | Nr5a2 | Nuclear receptor subfamily 5, group A, member 2 | ZGA, peaks at 4 CC |
| 34 | Obox1 | Oocyte specific homeobox 1 | Maternal |
| 35 | Oog1 | Oogenesin 1 | Maternal |
| 36 | Plat | Plasminogen activator, tissue | Dormant maternal |
| 37 | Pou5f1 | POU domain, class 5, transcription factor 1 | Pluripotency network |
| 38 | Ppil3 | Peptidylprolyl isomerase (cyclophilin)-like 3 | Meiotic mRNA degradation |
| 39 | Rb | Retinoblastoma-like 1 (p107) | Cell cycle regulator |
| 40 | Rbl2 | Retinoblastoma-like 2 | miR-290 target, senescence marker |
| 41 | Rpl18a | Ribosomal protein L18A | Meiotic mRNA degradation |
| 42 | Rps17 | Ribosomal protein S17 | Meiotic mRNA degradation |
| 43 | Sox2 | SRY-box containing gene 2 | Pluripotency network |
| 44 | Spin1 | Spindlin 1 | Maternal |
| 45 | Trim71 | Tripartite motif-containing 71 | let-7 target, MGA, peaks at 8CC |
| 46 | YY1 | YY1 transcription factor | ZGA 2C |
| 47 | Zar1 | Zygote arrest 1 | Significant drop 1C-2C |
| 48 | Zp3 | Zona pellucida glycoprotein 3 | Maternal |

| Number | ID | Official name | Reason |
|---|---|---|---|
| 1 | Actb | actin, beta | Housekeeping gene |
| 2 | Bax | BCL2-associated X protein | Apoptotic |
| 3 | Bcl2 | B cell leukemia/lymphoma 2 | Apoptotic |
| 4 | Bim | BCL2-like 11 (apoptosis facilitator) | Apoptotic |
| 5 | Brachyury T | brachyury | Mesoderm |
| 6 | CBX8 | chromobox homolog 8 | Trophoectoderm |
| 7 | Cdkn1a | cyclin-dependent kinase inhibitor 1A (P21) | Senescence |
| 8 | Cdkn2a | cyclin-dependent kinase inhibitor 2A | Senescence |
| 9 | Cdx2 | caudal type homeobox 2 | Trophoectoderm |
| 10 | cMyc | myelocytomatosis oncogene cellular | Pluripotency markers |
| 11 | Crabp2 | cellular retinoic acid binding protein II | Ectoderm marker - neural commitment |
| 12 | Creb3l2 | cAMP responsive element binding protein 3-like 2 | ICM |
| 13 | Dppa3 | developmental pluripotency-associated 3 | Pluripotency marker |
| 14 | Esrrb | estrogen related receptor, beta | Pluripotency marker |
| 15 | Fabp7 | fatty acid binding protein 7, brain | Ectoderm marker - neural commitment |
| 16 | FGF4 | fibroblast growth factor 4 | Epiblast |
| 17 | Fgfr2 | fibroblast growth factor receptor 2 | Primitive endoderm |
| 18 | Gapdh | glyceraldehyde-3-phosphate dehydrogenase | Housekeeping gene |
| 19 | Gata3 | GATA binding protein 3 | Trophoectoderm |
| 20 | Gata4 | GATA binding protein 4 | Primitive endoderm, and also definitive endoderm |
| 21 | Gata6 | GATA binding protein 6 | Primitive endoderm, and also definitive endoderm |
| 22 | Gsc | goosecoid homeobox | Mesendoderm |
| 23 | Hprt1 | hypoxanthine phosphoribosyltransferase 1 | Housekeeping gene |
| 24 | Klf4 | Kruppel-like factor 4 (gut) | Pluripotency marker |
| 25 | Lefty1 | left right determination factor 1 | Mesoderm |
| 26 | Lin28 | lin-28 homolog (C. elegans) | Pluripotency marker |
| 27 | Msc | musculin | Epiblast |
| 28 | Msx1 | homeobox, msh-like 1 | Mesendoderm |
| 29 | Nanog | Nanog homeobox | Pluripotency marker |
| 30 | Nes | nestin | Ectoderm marker |
| 31 | Nr5a2 | nuclear receptor subfamily 5, group A | ES pluripotency |
| 32 | Oct4/Pou5f1 | POU domain, class 5, transcription factor 1 | Pluripotency marker |
| 33 | Pax3 | pawered box gene 3 | Ectoderm marker |
| 34 | Pdgfra | platelet-derived growth factor receptor, alpha polypeptide | Primitive endoderm |
| 35 | Pecam1 | platelet/endothelial cell adhesion molecule 1 | ICM |
| 36 | Rbl2 | retinoblastoma-like 2 | Low in ES cells |
| 37 | Sall4 | sal-like 4 (Drosophila) | ICM |
| 38 | Sox13 | SRY-box containing gene 13 | ICM |
| 39 | Sox17 | SRY-box containing gene 17 | Primitive endoderm, and also definitive endoderm |
| 40 | Sox2 | SRY-box containing gene 2 | Pluripotency marker |
| 41 | Sox3 | SRY-box containing gene 3 | Ectoderm marker - neural commitment |
| 42 | Stat3 | signal transducer and activator of transcription 3 | Pluripotency markers |
| 43 | Tbx6 | T-box 6 | Mesoderm |
| 44 | Tcfap2a | transcription factor AP-2, alpha | Trophoectoderm |
| 45 | Tead4 | TEA domain family member 4 | Trophoectoderm |
| 46 | Trp53 | transformation related protein 53 | Senescence |
| 47 | VNP | transgene | External |
| 48 | Zfp42 | zinc finger protein 42 | Pluripotency marker |

*Blue typing marks genes, which were used in Guo et al., 2010 publication

# 10 List of references


Ahmed FE (2006) Microarray RNA transcriptional profiling: part I. Platforms, experimental design and standardization. *Expert Rev Mol Diagn* **6:** 535-550

Ahmed K, Dehghani H, Rugg-Gunn P, Fussner E, Rossant J, Bazett-Jones DP (2010) Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. *PLoS One* **5:** e10531

Aoki F, Worrad DM, Schultz RM (1997) Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. *Dev Biol* **181:** 296-307

Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R (2003) Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* **17:** 126-140

Azzalin CM, Reichenbach P, Khoriauli L, Giulotto E, Lingner J (2007) Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* **318:** 798-801

Bar T, Kubista M, Tichopad A (2012) Validation of kinetics similarity in qPCR. *Nucleic Acids Res* **40:** 1395-1406

Bergsland M, Ramskold D, Zaouter C, Klum S, Sandberg R, Muhr J (2011) Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev* **25:** 2453-2464

Boyer LA, Mathur D, Jaenisch R (2006) Molecular control of pluripotency. *Curr Opin Genet Dev* **16:** 455-462

Braude PR (1979) Time-dependent effects of alpha-amanitin on blastocyst formation in the mouse. *J Embryol Exp Morphol* **52:** 193-202

Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, Willard HF (1991) A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349:** 38-44

Bultman SJ, Gebuhr TC, Pan H, Svoboda P, Schultz RM, Magnuson T (2006) Maternal BRG1 regulates zygotic genome activation in the mouse. *Genes Dev* **20:** 1744-1754

Cartwright P, McLean C, Sheppard A, Rivett D, Jones K, Dalton S (2005) LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development* **132:** 885-896

Colledge WH, Carlton MB, Udy GB, Evans MJ (1994) Disruption of c-mos causes parthenogenetic development of unfertilized mouse eggs. *Nature* **370:** 65-68

Davis W, Jr., Schultz RM (2000) Developmental change in TATA-box utilization during preimplantation mouse development. *Dev Biol* **218:** 275-283

Edson MA, Nagaraja AK, Matzuk MM (2009) The mammalian ovary from genesis to revelation. *Endocr Rev* **30:** 624-712

Evans MJ, Kaufman MH (1981) Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292:** 154-156

Flach G, Johnson MH, Braude PR, Taylor RA, Bolton VN (1982) The transition from maternal to embryonic control in the 2-cell mouse embryo. *Embo J* **1:** 681-686

Gong C, Maquat LE (2011) lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* **470:** 284-288

Goulding MD, Chalepakis G, Deutsch U, Erselius JR, Gruss P (1991) Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. *Embo J* **10:** 1135-1147

Guo G, Huss M, Tong GQ, Wang C, Li Sun L, Clarke ND, Robson P (2010) Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev Cell* **18:** 675-685

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458:** 223-227

Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, Yang X, Amit I, Meissner A, Regev A, Rinn JL, Root DE, Lander ES (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477:** 295-300

Guttman M, Rinn JL (2012) Modular regulatory principles of large non-coding RNAs. *Nature* **482:** 339-346

Hamatani T, Carter MG, Sharov AA, Ko MS (2004) Dynamics of global gene expression changes during mouse preimplantation development. *Dev Cell* **6:** 117-131

Hanna J, Saha K, Pando B, van Zon J, Lengner CJ, Creyghton MP, van Oudenaarden A, Jaenisch R (2009) Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462:** 595-601

Hoffmann A, Czichos S, Kaps C, Bachner D, Mayer H, Kurkalli BG, Zilberman Y, Turgeman G, Pelled G, Gross G, Gazit D (2002) The T-box transcription factor Brachyury mediates cartilage development in mesenchymal stem cell line C3H10T1/2. *J Cell Sci* **115:** 769-781

Huarte J, Belin D, Vassalli A, Strickland S, Vassalli JD (1987) Meiotic maturation of mouse oocytes triggers the translation and polyadenylation of dormant tissue-type plasminogen activator mRNA. *Genes Dev* **1:** 1201-1211

Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, Khalil AM, Zuk O, Amit I, Rabani M, Attardi LD, Regev A, Lander ES, Jacks T, Rinn JL (2010) A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142:** 409-419

Chen J, Melton C, Suh N, Oh JS, Horner K, Xie F, Sette C, Blelloch R, Conti M (2011) Genome-wide analysis of translation reveals a critical role for deleted in azoospermia-like (Dazl) at the oocyte-to-zygote transition. *Genes Dev* **25:** 755-766

Chen X, Vega VB, Ng HH (2008) Transcriptional regulatory networks in embryonic stem cells. *Cold Spring Harb Symp Quant Biol* **73:** 203-209

Johnson MH, Ziomek CA (1981) The foundation of two distinct cell lineages within the mouse morula. *Cell* **24:** 71-80

Kigami D, Minami N, Takayama H, Imai H (2003) MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol Reprod* **68:** 651-654

Koutsourakis M, Langeveld A, Patient R, Beddington R, Grosveld F (1999) The transcription factor GATA6 is essential for early extraembryonic development. *Development* **126:** 723-732

Kretz M, Webster DE, Flockhart RJ, Lee CS, Zehnder A, Lopez-Pajares V, Qu K, Zheng GX, Chow J, Kim GE, Rinn JL, Chang HY, Siprashvili Z, Khavari PA (2012) Suppression of progenitor differentiation requires the long noncoding RNA ANCR. *Genes Dev* **26:** 338-343

Krishna M, Generoso WM (1977) Timing of sperm penetration, pronuclear formation, pronuclear DNA synthesis, and first cleavage in naturally ovulated mouse eggs. *J Exp Zool* **202:** 245-252

Litscher ES, Williams Z, Wassarman PM (2009) Zona pellucida glycoprotein ZP3 and fertilization in mammals. *Mol Reprod Dev* **76:** 933-941

Loewer S, Cabili MN, Guttman M, Loh YH, Thomas K, Park IH, Garber M, Curran M, Onder T, Agarwal S, Manos PD, Datta S, Lander ES, Schlaeger TM, Daley GQ, Rinn JL (2010) Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet* **42:** 1113-1117

Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J, Wong KY, Sung KW, Lee CW, Zhao XD, Chiu KP, Lipovich L, Kuznetsov VA, Robson P, Stanton LW, Wei CL, Ruan Y, Lim B, Ng HH (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38:** 431-440

Ma J, Flemr M, Stein P, Berninger P, Malik R, Zavolan M, Svoboda P, Schultz RM (2010) MicroRNA activity is suppressed in mouse oocytes. *Curr Biol* **20:** 265-270

Maekawa M, Yamaguchi K, Nakamura T, Shibukawa R, Kodanaka I, Ichisaka T, Kawamura Y, Mochizuki H, Goshima N, Yamanaka S (2011) Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1. *Nature* **474:** 225-229

Mansour SL (1990) Gene targeting in murine embryonic stem cells: introduction of specific alterations into the mammalian genome. *Genet Anal Tech Appl* **7:** 219-227

Mardis ER (2008) Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* **9:** 387-402

Miller G, Fuchs R, Lai E (1997) IMAGE cDNA clones, UniGene clustering, and ACeDB: an integrated resource for expressed sequence information. *Genome Res* **7:** 1027-1032

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

Murai S, Stein P, Buffone MG, Yamashita S, Schultz RM (2010) Recruitment of Orc6l, a dormant maternal mRNA in mouse oocytes, is essential for DNA replication in 1-cell embryos. *Dev Biol* **341:** 205-212

Murchison EP, Stein P, Xuan Z, Pan H, Zhang MQ, Schultz RM, Hannon GJ (2007) Critical roles for Dicer in the female germline. *Genes Dev* **21:** 682-693

Mutter GL, Grills GS, Wolgemuth DJ (1988) Evidence for the involvement of the proto-oncogene c-mos in mammalian meiotic maturation and possibly very early embryogenesis. *Embo J* **7:** 683-689

Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Scholer H, Smith A (1998) Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95:** 379-391

Niinuma T, Suzuki H, Nojima M, Nosho K, Yamamoto H, Takamaru H, Yamamoto E, Maruyama R, Nobuoka T, Miyazaki Y, Nishida T, Bamba T, Kanda T, Ajioka Y, Taguchi T, Okahara S, Takahashi H, Nishida Y, Hosokawa M, Hasegawa T, Tokino T, Hirata K, Imai K, Toyota M, Shinomura Y (2012) Upregulation of miR-196a and HOTAIR Drive Malignant Character in Gastrointestinal Stromal Tumors. *Cancer Res* **72:** 1126-1136

Niwa H, Toyooka Y, Shimosato D, Strumpf D, Takahashi K, Yagi R, Rossant J (2005) Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123:** 917-929

Nolan T, Hands RE, Bustin SA (2006) Quantification of mRNA using real-time RT-PCR. *Nat Protoc* **1:** 1559-1582

Oh B, Hwang SY, Solter D, Knowles BB (1997) Spindlin, a major maternal transcript expressed in the mouse during the transition from oocyte to embryo. *Development* **124:** 493-503

Ohinata Y, Payer B, O'Carroll D, Ancelin K, Ono Y, Sano M, Barton SC, Obukhanych T, Nussenzweig M, Tarakhovsky A, Saitou M, Surani MA (2005) Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* **436:** 207-213

Ohnishi Y, Totoki Y, Toyoda A, Watanabe T, Yamamoto Y, Tokunaga K, Sakaki Y, Sasaki H, Hohjoh H (2010) Small RNA class transition from siRNA/piRNA to miRNA during pre-implantation mouse development. *Nucleic Acids Res* **38:** 5141-5151

Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q, Guigo R, Shiekhattar R (2010) Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143:** 46-58

Pedersen RA, Wu K, Balakier H (1986) Origin of the inner cell mass in mouse embryos: cell lineage analysis by microinjection. *Dev Biol* **117:** 581-595

Piko L, Clegg KB (1982) Quantitative changes in total RNA, total poly(A), and ribosomes in early mouse embryos. *Dev Biol* **89:** 362-378

Qureshi IA, Mattick JS, Mehler MF (2010) Long non-coding RNAs in nervous system function and disease. *Brain Res* **1338:** 20-35

Rodda DJ, Chew JL, Lim LH, Loh YH, Wang B, Ng HH, Robson P (2005) Transcriptional regulation of nanog by OCT4 and SOX2. *J Biol Chem* **280:** 24731-24737

Rogers YH, Venter JC (2005) Genomics: massively parallel sequencing. *Nature* **437:** 326-327

Saitou M, Payer B, Lange UC, Erhardt S, Barton SC, Surani MA (2003) Specification of germ cell fate in mice. *Philos Trans R Soc Lond B Biol Sci* **358:** 1363-1370

Sessa L, Breiling A, Lavorgna G, Silvestri L, Casari G, Orlando V (2007) Noncoding RNA synthesis and loss of Polycomb group repression accompanies the colinear activation of the human HOXA cluster. *Rna* **13:** 223-239

Schuster SC (2008) Next-generation sequencing transforms today's biology. *Nat Methods* **5:** 16-18

Silva J, Barrandon O, Nichols J, Kawaguchi J, Theunissen TW, Smith A (2008) Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol* **6:** e253

Simeone A, Acampora D, Arcioni L, Andrews PW, Boncinelli E, Mavilio F (1990) Sequential activation of HOX2 homeobox genes by retinoic acid in human embryonal carcinoma cells. *Nature* **346:** 763-766

Sorensen RA, Wassarman PM (1976) Relationship between growth and meiotic maturation of the mouse oocyte. *Dev Biol* **50:** 531-536

Soudais C, Bielinska M, Heikinheimo M, MacArthur CA, Narita N, Saffitz JE, Simon MC, Leiden JM, Wilson DB (1995) Targeted mutagenesis of the transcription factor GATA-4 gene in mouse embryonic stem cells disrupts visceral endoderm differentiation in vitro. *Development* **121:** 3877-3888

Spurgeon SL, Jones RC, Ramakrishnan R (2008) High throughput gene expression measurement with real time PCR in a microfluidic dynamic array. *PLoS One* **3:** e1662

Stitzel ML, Seydoux G (2007) Regulation of the oocyte-to-zygote transition. *Science* **316:** 407-408

Strumpf D, Mao CA, Yamanaka Y, Ralston A, Chawengsaksophak K, Beck F, Rossant J (2005) Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development* **132:** 2093-2102

Stutz A, Conne B, Huarte J, Gubler P, Volkel V, Flandin P, Vassalli JD (1998) Masking, unmasking, and regulated polyadenylation cooperate in the translational control of a dormant mRNA in mouse oocytes. *Genes Dev* **12:** 2535-2548

Svoboda P, Stein P, Anger M, Bernstein E, Hannon GJ, Schultz RM (2004) RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev Biol* **269:** 276-285

Tang F, Barbacioru C, Nordman E, Bao S, Lee C, Wang X, Tuch BB, Heard E, Lao K, Surani MA (2011) Deterministic and stochastic allele specific gene expression in single mouse blastomeres. *PLoS One* **6:** e21208

Tong ZB, Gold L, Pfeifer KE, Dorward H, Lee E, Bondy CA, Dean J, Nelson LM (2000) Mater, a maternal effect gene required for early embryonic development in mice. *Nat Genet* **26:** 267-268

Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA, Blencowe BJ, Prasanth SG, Prasanth KV (2010) The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* **39:** 925-938

Valen E, Preker P, Andersen PR, Zhao X, Chen Y, Ender C, Dueck A, Meister G, Sandelin A, Jensen TH (2011) Biogenic mechanisms and utilization of small RNAs derived from human protein-coding genes. *Nat Struct Mol Biol* **18:** 1075-1082

Vassalli JD, Stutz A (1995) Translational control. Awakening dormant mRNAs. *Curr Biol* **5:** 476-479

Velkey JM, O'Shea KS (2003) Oct4 RNA interference induces trophectoderm differentiation in mouse embryonic stem cells. *Genesis* **37:** 18-24

Werner T (2010) Next generation sequencing in functional genomics. *Brief Bioinform* **11:** 499-511

Wu X, Viveiros MM, Eppig JJ, Bai Y, Fitzpatrick SL, Matzuk MM (2003) Zygote arrest 1 (Zar1) is a novel maternal-effect gene critical for the oocyte-to-embryo transition. *Nat Genet* **33:** 187-191

Yamada G, Mansouri A, Torres M, Stuart ET, Blum M, Schultz M, De Robertis EM, Gruss P (1995) Targeted mutation of the murine goosecoid gene results in craniofacial defects and neonatal death. *Development* **121:** 2917-2922

Yang CS, Li Z, Rana TM (2011) microRNAs modulate iPS cell generation. *Rna* **17:** 1451-1460

Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin, II, Thomson JA (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318:** 1917-1920

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

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

Zhang Z, Gao Y, Gordon A, Wang ZZ, Qian Z, Wu WS (2011) Efficient generation of fully reprogrammed human iPS cells via polycistronic retroviral vector and a new cocktail of chemical compounds. *PLoS One* **6:** e26592

Ziomek CA, Johnson MH (1980) Cell surface interaction induces polarization of mouse 8-cell blastomeres at compaction. *Cell* **21:** 935-942