Analysis of genome activation in early bovine embryos by bioinformatic evaluation of RNA-Seq data

Alexander Stefan Graf

Inaugural-Dissertation zur Erlangung der Doktorwürde
der Tierärztlichen Fakultät der Ludwig-Maximilians-Universität München

Analysis of genome activation in early bovine embryos by bioinformatic evaluation of RNA-Seq data

von

Alexander Stefan Graf

aus Nürnberg

München 2015

Aus dem
Department für Veterinärwissenschaften der Tierärztlichen Fakultät
der Ludwig-Maximilians-Universität München
Lehrstuhl für Molekulare Tierzucht und Biotechnologie

Arbeit angefertigt unter der Leitung von:
Univ.-Prof. Dr. Eckhard Wolf

im
Labor für Funktionale Genomanalyse (LAFUGA)
Genzentrum der Ludwig-Maximilians-Universität München
Genomics

Mentor: Dr. Helmut Blum

Gedruckt mit Genehmigung der Tierärztlichen Fakultät der
Ludwig-Maximilians-Universität München

Dekan:	Univ.-Prof. Dr. Joachim Braun
Berichterstatter:	Univ.-Prof. Dr. Eckhard Wolf
Korreferenten:	Priv.-Doz. Dr. Sabine André
	Univ.-Prof. Dr. Joachim Braun
	Univ.-Prof. Dr. Dr. h.c. Erwin P. Märtlbauer
	Univ.-Prof. Dr. Dušan Palić

Tag der Promotion: 31. Januar 2015

Contents

1 Introduction

The journey of life starts with the remarkable transformation of a fertilized oocyte to a totipotent zygote, later differentiating into a complete organism. After fertilization, the early embryo is in a transcriptionally inactive state and thus is dependent on maternal transcripts and proteins stored in the oocyte during oogenesis (reviewed in Tadros et al., 2009). One of the main goals of these transcripts is to promote oocyte maturation and the support of embryonic development. These oocyte stored transcripts must be kept in a stable form and translation must be prevented. This may be achieved by a de-polyadenylation of the 3' end of the mRNA, which hinders translation in association with ribonucleoproteins. Translation initation seems to be dependent on polyA tail length since short tails (approximately 20 As) are known to repress translation. A re-polyadenylation of mRNAs with specific regulatory motifs and proteins is involved to activate translation (reviewed in Sirard, 2012). During embryonic development a switch from maternal to embryonic transcripts occurs, which is termed maternal-to-embryonic transition (MET). This process involves the degradation of maternal transcripts together with a subsequent replacement with embryonic ones and the increase of new embryo derived transcripts (reviewed in Sirard, 2010). As of today, two ways of degradation of maternally stored transcripts are known. First, the degradation is controlled by proteins that induce deadenylation by binding to sequences in the 3' end of the untranslated region of the transcripts (reviewed in Schier, 2007). It has been shown in Drosophila, that the *SMAUG* (SMG) gene is involved in the regulation of maternal transcript destabilization by recruiting deadenylases, which trigger the degradation of the transcripts by nucleases (Tadros et al., 2007). Second, specific microRNAs might be involved in the degradation of maternal transcripts. Studies in zebrafish and rainbow trout revealed that specific microRNAs produced by the embryo are involved in the degradation of maternal transcripts (Giraldez et al., 2006; Ramachandra et al., 2008). In bovine embryos, microRNAs including miR-130a, miR-21 were speculated to be involved in maternal transcript degradation (Mondou et al., 2012). Several different aspects might

be responsible for the repression of embryonic transcription. Maternal transcripts are known to alter chromatin structure, leading to changes in the activity of transcription factors that allow or restrict access to regulatory elements of the genome (Prioleau et al., 1994; Østrup et al., 2013). A further point might be the rapid cleavage in early embryos, in which the DNA replication might interfere with nascent zygotic transcription (reviewed in Schier, 2007).

The ongoing gradual degradation of maternal transcripts is accompanied by the initiation of first embryonic genome activation (EGA). EGA occurs in several waves and is a crucial event during MET. In mammals, the timing of the main EGA is species-specific. It occurs at the two-cell stage in mouse embryos (Wang et al., 2006), at the four- to eight-cell stage in human (Braude et al., 1988) and pig embryos (reviewed in Sirard, 2010), and at the eight- to 16-cell stage in bovine and rabbit embryos (Telford et al., 1990; Sirard, 2012). This major activation is preceded by an initial minor embryonic transcription of a few genes which are known to be involved in cell proliferation, mitotic cell cycle, regulation of transcription as well as DNA and protein metabolism (Kanka et al., 2012). Bovine EGA was first studied with autoradiographic analyses of embryos followed by short term incorporation of [^{3}H]uridine (Camous et al., 1986; Frei et al., 1989). These studies identified major EGA to appear at the eight- to 16-cell stage. This was in accordance with another study using polypeptide profiles of bovine *in vitro* embryos treated with α-amanitin (Svarcova et al., 2007). Long term exposure with [^{3}H]uridine revealed a first minor genome activation at the two- to four-cell stage (Plante et al., 1994; Hyttel et al., 1996; Viuff et al., 1996; Memili et al., 1998). Additional studies suggest that EGA occurs even earlier in one-cell embryos (Memili et al., 1999). First identification of activated genes at major EGA was achieved by comparison of transcription profiles generated by microarrays of untreated bovine eight-cell embryos with those treated with α-amanitin. This method identified 233 genes and could assign them to gene ontology (GO) categories of "Regulation of transcription", "Cell adhesion", "Apoptosis/cell death", "Protein folding", "Transport", "Metabolism", and "Immune response" (Misirlioglu et al., 2006). Another study created transcriptome profiles with Affymetrix microarrays from bovine metaphase II (MII), zygotes, two-cell, four-cell, eight-cell, morula and blastocyst stages (Kues et al., 2008). The authors found the highest number of differentially expressed genes with more than 7,000 genes between the four-cell and eight-cell stage, which properly reflects major EGA at this stage. Additionally, several hundred transcripts were identified to be activated before major EGA at the two- to four-cell stage. They found stage specific gene expression, e.g. genes from the eukaryotic translation initiation factor family (*EIF2*, *EIF3*, *EIF4* and *EIF5*) that are exclusively expressed at the eight-cell

stage of major genome activation (Kues et al., 2008). However, expression profiling with microarray based techniques or RT-PCR analyses of candidate genes enable only the determination of relative transcript abundances. These techniques cannot distinguish between newly synthesized transcripts and maternal derived ones. A discrimination between maternal and embryonic transcripts can only be facilitated for newly synthesized embryonic transcripts not present in oocytes.

RNA sequencing (RNA-Seq) enables the discrimination between maternal and paternal transcripts by the use of single nucleotide polymorphisms (SNPs). Compared to microarrays, RNA-Seq allows transcriptome analyses at single nucleotide resolution and has a higher sensitivity and a higher dynamic range (Wang et al., 2009). First studies used RNA-Seq to compare normal bovine IVF blastocysts with degenerated ones (Huang et al., 2010). Novel, unannotated transcriptional units and alternative splice events were identified, suggesting more genes to be involved in early embryonic development. Up to now, no detailed analyses of specific genes related to early bovine embryonic development exist and the underlying mechanisms of embryonic genome activation are only partially understood.

Therefore, a RNA-Seq study was performed to generate comprehensive transcriptome profiles of bovine germinal vesicle oocytes (GV), metaphase II oocytes and of embryos at the four-cell, eight-cell, 16-cell and blastocyst stage. Embryos were produced by *in vitro* fertilization of German Simmental (*Bos taurus taurus*) oocytes with sperm from a single, genetically distant Brahman (*Bos taurus indicus*) bull. Aim of this thesis was to perform a bioinformatic evaluation of these transcriptome profiles and to develop new tailored procedures as a basis for detailed analysis of early embryonic development. Based on these results, the major genome activation in early bovine embryos should be analyzed and involved genes should be captured and characterized. Using RNA-Seq, this cross-breeding design was used to establish several strategies to provide detailed insights into the timing of bovine EGA.

2 Publications

Proceedings of the National Academy of Science of the United States of America

Fine mapping of genome activation in bovine embryos by RNA sequencing

Alexander Graf[a], Stefan Krebs[a], Valeri Zakhartchenko[b], Björn Schwalb[c], Helmut Blum[a,1,2], Eckhard Wolf[a,b,1,2]

[a]Laboratory for Functional Genome Analysis (LAFUGA), [b]Chair for Molecular Animal Breeding and Biotechnology, and [c]Department of Biochemistry, Center for Integrated Protein Science (CIPSM), Gene Center, Ludwig-Maximilians-Universität München, Munich, Germany

[1]H.B. and E.W. contributed equally to this work
[2]To whom correspondence may be addressed. E-mail: ewolf@lmb.uni-muenchen.de and blum@lmb.uni-muenchen.de

Short title: RNA-Seq of bovine embryos

Author contributions: H.B. and E.W. designed research; A.G. and S.K. performed research; V.Z. contributed new reagents/analytic tools; A.G., S.K. and B.S. analyzed data; and A.G., S.K., H.B. and E.W. wrote the paper.

Reference: Proc Natl Acad Sci U S A. 2014 Mar 18;111(11):4139–4144

Abstract

During maternal-to-embryonic transition control of embryonic development gradually switches from maternal RNAs and proteins stored in the oocyte to gene products generated after embryonic genome activation (EGA). Detailed insight into the onset of embryonic transcription is obscured by the presence of maternal transcripts. Using the bovine model system, we established by RNA sequencing a comprehensive catalogue of transcripts in germinal vesicle and metaphase II oocytes, and in embryos at the four-cell, eight-cell, 16-cell and blastocyst stages. These were produced by in vitro fertilization of *Bos taurus taurus* oocytes with sperm from a *Bos taurus indicus* bull to facilitate parent-specific transcriptome analysis. Transcripts from 12.4 to 13.7 $\times$ 10^3 different genes were detected in the various developmental stages. EGA was analyzed by (*i*) detection of embryonic transcripts which are not present in oocytes; (*ii*) detection of transcripts from the paternal allele; and (*iii*) detection of primary transcripts with intronic sequences. These strategies revealed (*i*) 220, (*ii*) 937, and (*iii*) 6,848 genes to be activated from the four-cell to the blastocyst stage. The largest proportion of gene activation [i.e., (*i*) 59%, (*ii*) 42%, and (*iii*) 58%] was found in eight-cell embryos, indicating major EGA at this stage. Gene ontology analysis of genes activated at the four-cell stage identified categories related to RNA processing, translation, and transport, consistent with preparation for major EGA. Our study provides the largest transcriptome data set of bovine oocyte maturation and early embryonic development and detailed insight into the timing of embryonic activation of specific genes.

Significance

Gene expression profiling is widely used to get insight into mechanisms of early embryonic development and to characterize embryos generated by various techniques or exposed to different culture conditions. Transcripts in early embryos may be of maternal or embryonic origin, which is difficult to distinguish by conventional techniques. RNA sequencing in bovine oocytes and embryos facilitated mapping of the onset of embryonic expression for almost 7,400 genes. The timing of embryonic gen(om)e activation offers an additional level of information for embryo biosystems research and for detecting disturbances of early development due to genetic, epigenetic, and environmental factors.

Introduction

Early embryonic development is governed by maternal transcripts and proteins stored within the oocyte during oogenesis (reviewed in Tadros et al., 2009). As development proceeds, maternally derived transcripts and proteins are degraded, whereas embryonic genome activation (EGA) is initiated. The period when control of development is shifted from maternal gene products to embryonic ones is referred to as the maternal-to-embryonic transition (MET).

EGA occurs in several waves, and the timing of major EGA is species dependent: it occurs at the two-cell stage in mouse embryos (reviewed in Sirard, 2012), at the four- to eight-cell stage in human (Braude et al., 1988) and pig embryos (reviewed in Sirard, 2012), and at the eight- to 16-cell stage in bovine embryos (reviewed in Sirard, 2012). At the time of EGA, both maternal and embryonic transcripts are present in the embryo, thus hampering a precise mapping of the onset of embryonic expression of specific genes. First insight into the timing of global EGA came from incorporation studies of radiolabeled UTP. ^{35}S-UTP incorporation was high at the germinal vesicle (GV) stage of oocytes, decreased to background levels in metaphase II (MII) oocytes, increased again in two-cell embryos, remained at the same level during the four-cell stage, but increased significantly at the eight-cell stage (Memili et al., 1998). The authors concluded that bovine two-cell embryos are already transcriptionally competent and active, but that major EGA occurs at the eight-cell stage. To identify the genes that are activated during major EGA in bovine embryos, subsequent studies used microarrays to screen for transcriptome differences between untreated eight-cell embryos and eight-cell embryos treated with the transcription inhibitor α-amanitin (Misirlioglu et al., 2006; Vigneault et al., 2009). These studies identified several hundred transcripts with increased abundance in transcriptionally active eight-cell embryos. Gene ontology classification of the differentially expressed genes showed that they were involved in chromatin structure, transcription, RNA processing, protein biosynthesis, signal transduction, cell adhesion, and maintenance of pluripotency. Naturally the discovery of genes being activated was limited by the probe sets present on the respective microarrays. Further, to date there is no systematic study addressing the activation of specific genes during several stages of early bovine embryo development. Detailed insight into the time course of embryonic genome activation is important, because embryos are particularly susceptible during the period of EGA (e.g., to changing culture conditions) (Gad et al., 2012). However, the underlying mechanisms are only partially understood, and new molecular readouts, such as effects on the timing of EGA, are required.

We used high-throughput sequencing to generate comprehensive transcriptome profiles of bovine GV and MII oocytes, of four-cell, eight-cell, and 16-cell embryos, and of blastocysts. By combining a dedicated cross breeding design of *Bos taurus taurus* × *Bos taurus indicus* with the sensitivity and single nucleotide resolution of RNA-Seq, we established various strategies for identification of de novo transcribed RNAs, providing detailed insight into the timing of gene activation during early bovine embryo development.

Results

Hybrid embryos as model to study parent-specific gene expression in bovine embryos. Pools of German Simmental oocytes (*Bos t. taurus*) from randomly selected donor animals were in vitro matured and fertilized with semen from a single Brahman bull (*Bos t. indicus*). Because the two *B. taurus* subspecies are relatively distant (Troy et al., 2001), a large number of informative SNPs for unequivocal identification of transcripts from the paternal allele can be expected in hybrid embryos. Presumptive zygotes were cultured according to a standard protocol (Bauersachs et al., 2009). Pools of 10 oocytes (GV oocytes and in vitro matured MII oocytes) or embryos (four-cell, eight-cell, 16-cell and blastocyst stages) were lysed, and cDNA was synthesized using a combination of oligo-dT and random primers to cover the whole transcriptome except for ribosomal RNAs. After single primer isothermic amplification, cDNAs were used for library preparation and sequenced on an Illumina Genome Analyzer IIx. The sequenced reads were filtered and mapped against the bovine reference genome sequence. Three biological replicates of each oocyte and embryo pool were analyzed. The number of reads per biological replicate was 9–58 × 10^6 (Table S1). In general, all parts of the transcripts (5' to 3') were equally well represented in all analyzed developmental stages (Figure S1). In oocytes and four-cell embryos approximately 60% of the reads mapped to coding sequences, whereas less than 10% of the reads mapped to intronic sequences. The proportion of intron-specific reads increased approximately threefold from the four-cell (6.5%) to the eight-cell stage (20%) and further approximately 1.4-fold between the eight-cell and the 16-cell stage (28%), suggesting an increase of primary transcripts (Figure S2).

Global view on the transcriptome of bovine oocytes and preimplantation embryos. In all developmental stages, transcripts from 12.4 to 13.7 $\times$ 10^3 different genes were detected (Figure 1). Comparison of transcript abundances between the various developmental stages showed relatively few differentially abundant transcripts between GV and MII oocytes as well as between GV oocytes and four-cell embryos. The number of differentially abundant transcripts increased 10-fold between the four-cell and eight-cell stages, and even more between subsequent stages. Our study provides the most comprehensive resource of transcriptome data for bovine oocytes and early embryos. The raw FASTQ files and the normalized read counts per gene are publically available at Gene Expression Omnibus (GEO) (accession no. GSE52415). Interestingly, during early development up to the four-cell stage approximately 85% of the differentially abundant transcripts were increased, and only approximately 15% decreased in abundance. The proportion of differentially abundant transcripts with reduced abundance in eight-cell vs. four-cell embryos corresponded to 24% and increased further to 55% in 16-cell vs. eight-cell embryos. In blastocysts the proportions of transcripts with increased and decreased abundance compared with the 16-cell stage were 51% and 49%, respectively. The observed transcriptome changes reflect the phenomenon of MET, in which maternal transcripts synthesized in immature oocytes are gradually degraded, whereas the embryonic genome is activated. We developed and tested different strategies to unravel the temporal pattern of EGA.

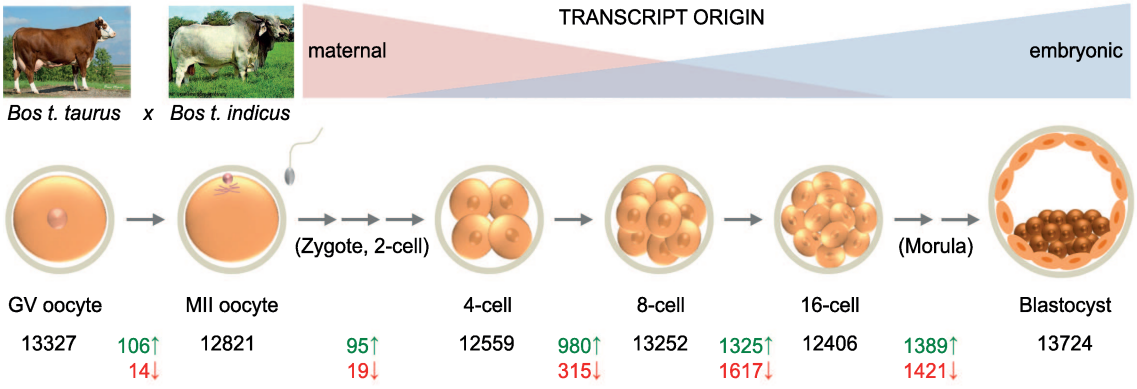

Fig. 1. Experimental design. Oocytes from *Bos t. taurus* cows were fertilized with sperm of a single *Bos t. indicus* bull. Pools of 10 oocytes/embryos were harvested at the GV, MII, four-cell, eight-cell, 16-cell, and blastocyst stage and processed for RNA-Seq. For each stage, the total number of genes with detectable transcripts is indicated in black. The numbers of differentially abundant transcripts between two stages are shown in green (increased abundance) and red (decreased abundance) in the subsequent vs. the previous stage.

Transcripts first detected in embryonic stages. The most obvious strategy to identify genes activated during early development of embryos is to look for transcripts that are not detected in oocytes. Genes were considered as first expressed in embryos when fewer than five reads were found in both oocyte stages (GV and MII) and at least 20 reads in one of the stages after fertilization. In addition, the transcript abundance had to be differentially up-regulated for the analyzed developmental stage to be designated as first expressed (Appendix Dataset 1). One example is the Nanog homeobox (*NANOG*) gene, which was – in accordance with a previous report (Khan et al., 2012) – found to be first expressed at the eight-cell stage (Figure 2A). In total, this approach revealed eight genes to be first expressed at the four-cell stage, 129 genes at the eight-cell stage, 36 genes at the 16-cell stage, and 47 genes at the blastocyst stage (Figure 2B and Appendix Dataset 2).

Embryonic gene activation as detected by transcripts from the paternal allele. In addition to the newly transcribed genes, the cross-breeding design was used to detect the onset of EGA by the appearance of transcripts from the paternal allele as identified by breed-specific SNPs. We identified 61,371 *B. t. indicus* (Brahman)-specific SNPs in exons that were distributed over 4,048 different genes, thus covering ∼20% of the 19,994 coding genes of the bovine reference genome. As shown in Figure 2C, transcripts of the 5-azacytidine-induced protein 2 (*AZI2*) gene were present in GV and MII oocytes and in four-cell embryos as well. However, transcripts from the paternal *AZI2* allele ("C", indicated as blue bar) were first detected in eight-cell embryos, suggesting embryonic activation of the *AZI2* gene at this stage. Using first expression of the paternal allele as a marker, 16 genes were found to be activated at the four-cell stage, 395 genes at the eight-cell stage, 314 genes at the 16-cell stage, and 212 genes at the blastocyst stage (Figure 2D and Appendix Dataset 2).

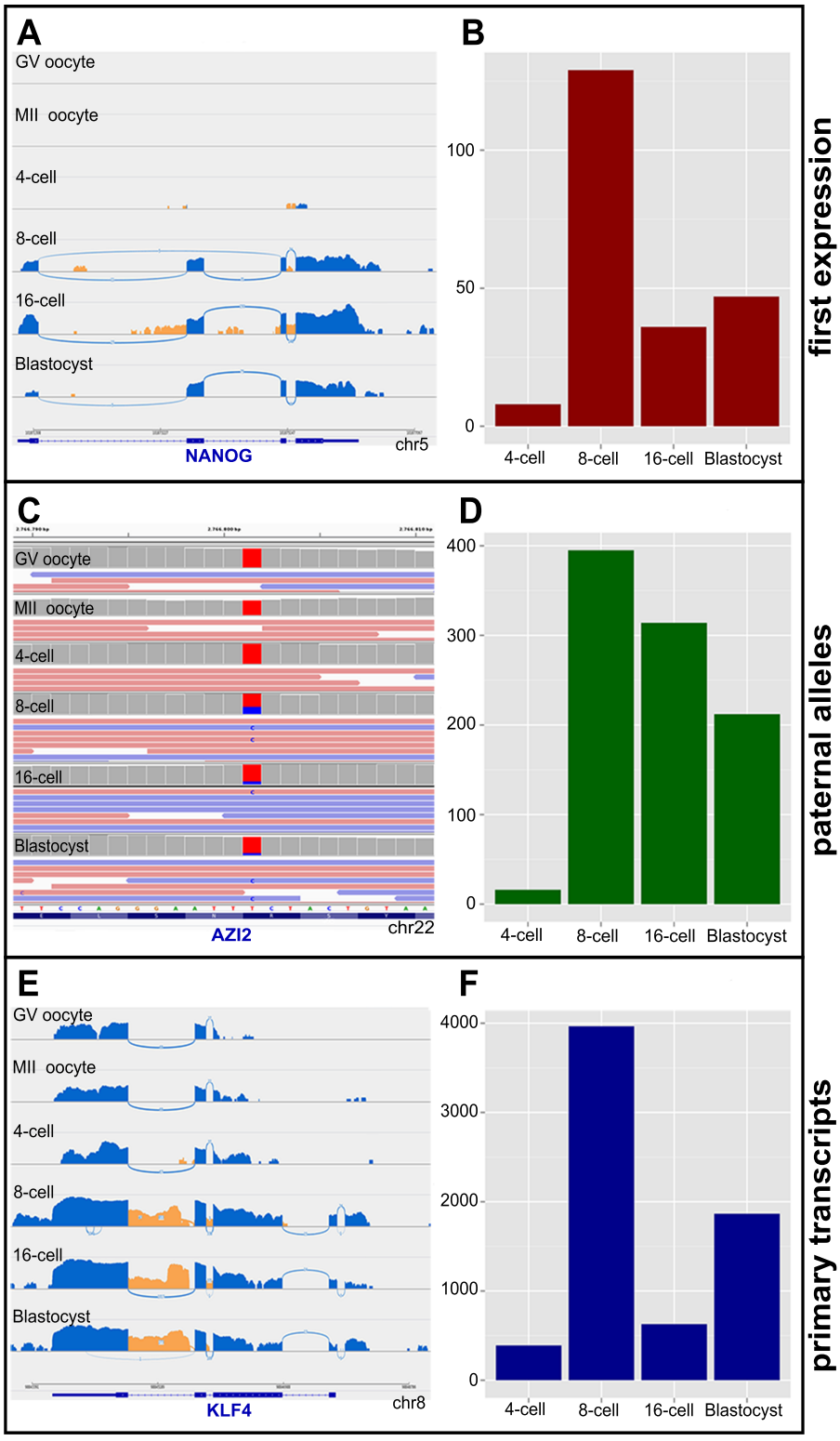

Fig. 2. Different strategies for fine mapping of genome activation in bovine embryos by RNA-Seq. (A and B) Detection of transcripts that are not present in oocytes. (A) Sashimi plot (Katz et al., 2010) of *NANOG* showing the absence of transcripts in oocytes, very limited transcription at the four-cell stage, and clear activation of embryonic transcription at the eight-cell stage. (B) Total numbers of genes activated at the respective embryonic stages as detected by first appearance of specific transcripts. (C and D) Detection of embryonic gene activation by the occurrence of paternal transcripts. (C) IGV (Thorvaldsdóttir et al., 2013) plot for *AZI2* indicating first appearance of transcripts with a paternal specific SNP (blue) at the eight-cell stage identified using three replicates. (D) Total numbers of genes activated at the respective embryonic stages as detected by first appearance of paternal transcripts. (E and F) Detection of embryonic gene activation by the appearance of primary transcripts. (E) Sashimi plot (Katz et al., 2010) for *KLF4* indicating that transcripts are present in oocytes and all embryonic stages, but transcripts with intronic reads (in orange) are first detectable at the eight-cell stage, suggesting embryonic activation of *KLF4* at this stage. (F) Total numbers of genes activated at the respective embryonic stages as detected by first appearance of primary transcripts.

Gene activation as detected by the appearance of incompletely spliced transcripts. Another parameter for detecting de novo synthesized transcripts is the presence of intronic sequences due to incomplete cotranscriptional splicing (Ameur et al., 2011). As shown in Figure 2A, the onset of expression of *NANOG* is accompanied by the presence of reads covering intronic regions. The gross assignment of reads to exonic and intronic sequences already indicated a marked (threefold) increase in the proportion of intronic reads between the four-cell and eight-cell stages (Figure S2). To discriminate intronic reads in primary transcripts from intronic reads resulting, e.g., from repetitive sequences, we defined the parameter RINP as a measure for the coverage of all intronic sequences in a transcript. It indicates the **r**atio of **i**ntronic read counts to **n**ot-covered intronic **p**ositions. A fold-change $\geq$10 in RINP between subsequent replicates of the embryonic stages was considered as indicative of nascent transcription. Background was defined as the 75th percentile of RINP in the oocyte stages (Figure S3). As an example, the activation of the Krüppel-like factor 4 (*KLF4*) gene is shown in Figure 2E. *KLF4* transcripts are present in GV and MII oocytes and are maintained through the four-cell stage; a substantial increase of intronic reads in eight-cell embryos clearly indicates embryonic activation of the *KLF4* gene at this stage. In total, the detection of intronic sequences revealed 390 genes to be activated at the four-cell stage, 3,965 genes at the eight-cell stage, 628 genes at the 16-cell stage, and 1,865 genes at the blastocyst stage (Figure 2F and Appendix Dataset 2).

Proportion of intronic sequences in transcripts in relation to gene length and developmental stage. The length of activated genes was determined and compared for all early embryonic stages (Figure S4). The length of primary transcripts increased significantly (Mann-Whitney test; $P < 0.01$) from the four-cell (median 19 kb) to the eight-cell stage (median 28 kb) and from the 16-cell (median 26 kb) to the blastocyst stage (median 31 kb).

To get a global view of the relationship between the proportions of intronic sequences in transcripts and gene size during early embryonic development, all annotated genes were ranked according to primary transcript length. For each intron of a gene the distance from transcript start to the center of each intron was calculated, and a dot was plotted if its RINP value was above background (75th percentile of RINP values calculated for MII oocytes = 0.0014; Figure S3). We found a nearly random distribution of dots in GV and MII oocytes and in four-cell embryos (Figure 3), indicating the presence of mainly mature transcripts. The density of dots increased markedly at the eight-cell

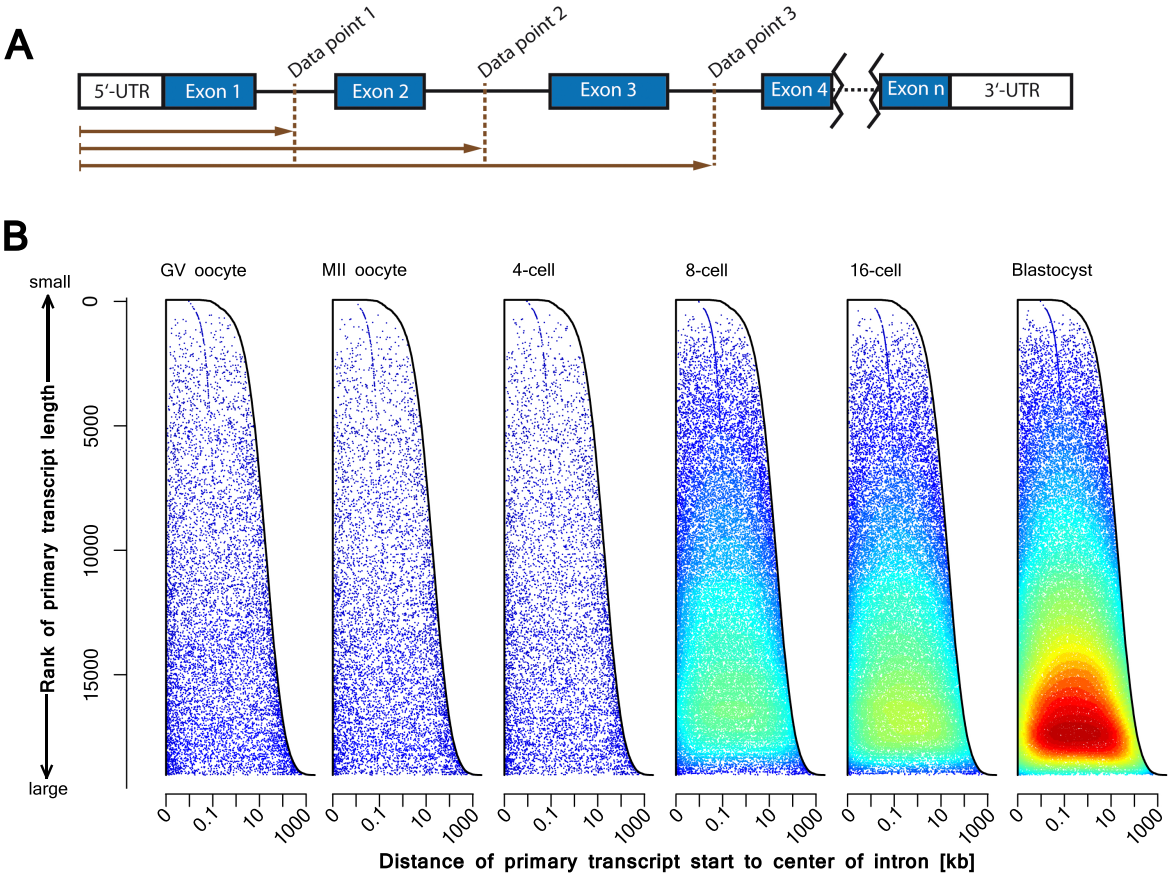

Fig. 3. Global view of intron transcription of all annotated genes. For each intron of all annotated genes a data point was computed by calculating the distance of the transcription start site to the center of the intron in bases (A). Furthermore, an RINP value was calculated by summing up all mapped intronic reads for each intron divided by the positions (bases) where no coverage was observed. On the basis of the RINP values a plot was generated for each oocyte and embryonic stage (B). For each data point, a dot was plotted if its RINP value was above background (Figure S3). The influence of gene length was visualized by ranking genes on the y axis according to length in descending order. The density of dots was visualized by a colored scatterplot generated by the R package *LSD*, with colors ranging from blue (low density) to red (high density). This algorithm causes a minor underestimation of dot density in the peripheral regions.

stage, corresponding with the major wave of embryonic genome activation. For smaller primary transcripts (upper half of the plot), the density of dots remained nearly constant from the eight-cell to the blastocyst stage, whereas for larger ones (lower half of the plot), the density increased during development.

Functional classification of genes activated before and during major EGA. Genes switched on at the four-cell stage or earlier are particularly interesting, because they may be involved in major EGA. We identified eight genes that were transcribed for the first time in four-cell embryos. Among them were "upstream binding transcription factor, RNA polymerase I-like 1" (*UBTFL1*; *LOC100140569*), "heterogeneous nuclear ribonucleoprotein A2/B1" (*HNRNPA2B1*; *LOC516616*), "Krüppel-like factor 17" (*KLF17*), and "Kelch-like family member 28" (*KLHL28*).

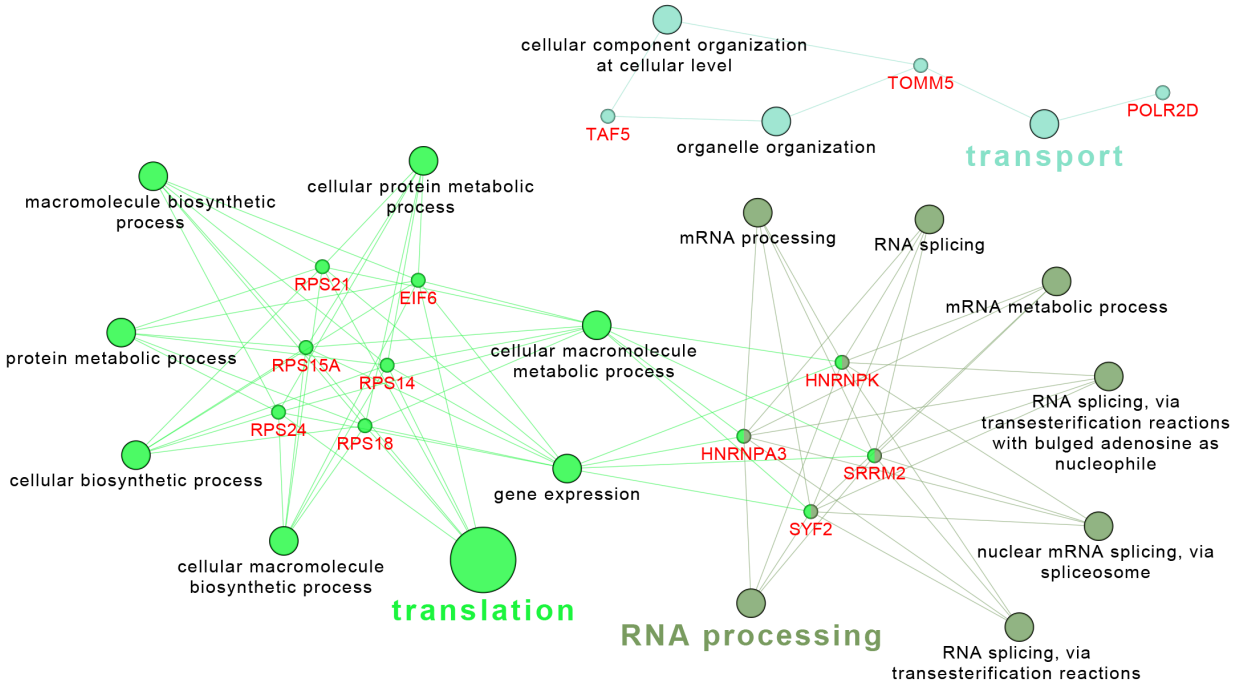

Fig. 4. Functionally grouped GO terms for the genes activated before major genome activation in bovine embryos. Genes activated at the four-cell stage were detected by the presence of de novo transcripts ($n = 8$), transcripts with paternal-specific SNPs ($n = 16$), or primary transcripts with intronic sequences ($n = 390$) and functionally analyzed with the *ClueGO* (Bindea et al., 2009) plugin of *Cytoscape*. The major significant GO terms were "RNA processing", "translation", and "transport". Genes enriched in the GO terms were colored in red. The significance of the GO terms is reflected by the size of the nodes.

The gene ontology (GO) analysis of the 414 genes identified by our three approaches as activated at the four-cell stage classified the GO terms "RNA processing", "translation", and "transport" as significantly overrepresented (Figure 4). The analysis of the 4,255 genes activated at the eight-cell stage revealed the GO term "RNA splicing" as the most prominent and additionally the GO terms "mRNA transcription from RNA polymerase II promoter", "regulation of transcription from RNA polymerase II promoter",

"purine nucleotide biosynthetic process", and "5S class rRNA transcription from RNA polymerase III type 1 promoter" (Figure S5).

Discussion

Expression profiling – either by RT-PCR analyses of candidate genes (reviewed in Wrenzycki et al., 2005) or by holistic approaches using array-based techniques (Kues et al., 2008) – has been widely used to identify molecular characteristics of bovine embryos of different origin or with different developmental potential. These techniques determine relative transcript abundances but fail to differentiate between transcripts of embryonic vs. maternal origin, except for embryonic transcripts that are not present in the oocyte. In general RNA-Seq is considered superior to hybridization-based methods of transcriptome profiling (reviewed in Wang et al., 2009). RNA-Seq directly determines the cDNA sequence; thus the read counts for a particular transcript provide a digital value of its abundance. Moreover, RNA-Seq facilitates parent-specific analyses of gene expression by the detection of parental SNPs. We constructed sequencing libraries without prior polyA+ selection or rRNA depletion. Furthermore, exonic as well as intronic parts of transcripts were detected. This approach enabled us – by the occurrence of intronic sequences in transcripts – to capture de novo transcription of genes in embryos with high sensitivity, even if transcripts of these genes were already present in oocytes.

Recent studies performed RNA-Seq analyses of bovine blastocysts (Chitwood et al., 2013) and of bovine conceptuses (days 10, 13, 16, and 19) (Mamo et al., 2011); however, no comprehensive transcriptome analysis covering the stages from the GV oocyte to the blastocyst stage is available to date.

Although it is technically feasible to perform RNA-Seq on single embryos (Chitwood et al., 2013) or even single embryonic cells (Xue et al., 2013), we decided to analyze three biological replicates of pools of 10 oocytes or embryos per developmental stage. Individual embryos may suffer from a considerable proportion of cytogenetic abnormalities (Demyda-Peyrás et al., 2013), which may affect their gene expression profile. In consequence, RNA-Seq analysis of single embryos or even single blastomeres may reflect the abnormality of a particular embryo or embryonic cell rather than the characteristic transcriptome profile of a specific developmental stage. A limitation of our study is the use of in vitro-produced embryos, which are known to be developmentally less competent than in vivo-derived embryos (reviewed in Lonergan et al., 2008). Future

studies comparing EGA of in vitro vs. in vivo embryos may provide new insights into these developmental differences.

In the various stages of bovine oocytes and embryos analyzed in the present study, transcripts from 12.4 to 13.7 $\times$ 10^3 different genes per developmental stage were identified. This was on the same order of magnitude or even higher than the number of expressed genes detected by single-cell RNA-Seq in human embryos (Xue et al., 2013).

In this study the proportion of uniquely mapped reads decreased from the early oocyte stage (74%) to the blastocyst stage (approximately 50%). Simultaneously, the percentage of intronic mapped reads increased from 7% in oocytes to 30% in blastocysts. Introns are known to contain a higher proportion of repetitive elements than exons, which has been shown to reduce the mappability of intron-derived reads and thus could explain the decreased number of uniquely mapped reads (Zhu et al., 2009). If multiple mapped reads were allowed in the alignments, we observed an increased portion of reads mapping to repetitive sequences after genome activation. In contrast, the restriction of alignment parameters to uniquely mapped reads led to a higher fraction of unmapped reads.

Our RNA-Seq analysis revealed relatively few differentially abundant transcripts between GV and MII oocytes and between GV oocytes and four-cell embryos. A marked increase in differentially abundant transcripts was observed between the four-cell and eight-cell stages, and even more between subsequent stages. Interestingly, the proportion of transcripts with decreased abundance was initially small (17% in four-cell embryos vs. MII oocytes; 24% in eight-cell vs. four-cell embryos), but increased to 55% in 16-cell vs. eight-cell embryos. This observation may, at least in part, be due to the degradation of maternal transcripts (reviewed in Tadros et al., 2009).

To get insight into time course of EGA, we tested three approaches: detection of (*i*) transcripts arising after fertilization, (*ii*) paternal SNPs, and (*iii*) primary transcripts.

We could show that GV and MII oocytes store transcripts of approximately 13,000 genes, whereas only a small number of genes were transcribed for the first time after fertilization (in total 220; 129 (59%) of them at the eight-cell stage).

Our experimental design allowed us to capture the active transcription of genes according to the detection of paternal SNPs, albeit corresponding transcripts being present in oocytes. This was achieved by fertilizing *B. t. taurus* oocytes in vitro with semen from a single bull of the genetically distant subspecies *B. t. indicus*. In total ~61,000 paternal SNPs could be identified, covering ~20% of all known bovine genes. On the basis of

this data set we were able to detect the embryonic activation of 937 genes, 395 (42%) of which were actively transcribed during EGA at the eight-cell stage.

The third approach to determine the onset of embryonic gene expression was the detection of transcripts with intronic sequences. In total 6,848 genes were found to be switched on from the four-cell to the blastocyst stage. The majority of these genes (3,965; 58%) were activated at the eight-cell stage. No spatial clustering of activated genes to certain chromosomal locations was observed.

Notably, the results of the three methods to detect the onset of gene expression were consistent with respect to the timing of minor EGA at the four-cell stage or before and major EGA at the eight-cell stage; however, the absolute numbers of activated genes detected were rather different. This is because (*i*) method 1 covered only genes that are not transcribed in oocytes; and (*ii*) method 2 relied on SNPs distinguishing the parental alleles, which were – in our experiment – found only in approximately 20% of the known bovine genes. Thus, method 3, based on the presence of primary transcripts, identified the largest proportion of activated genes. Nevertheless, the results of method 2 and 3 were remarkably concordant (Figure S6). The limitations of our study in detecting all activated genes could be overcome by labeling and enriching nascent RNA and by increasing the sequencing depth. In comparison with a set of transcripts enriched in normal vs. α-amanitin-treated bovine eight-cell embryos (Vigneault et al., 2009), we found a significant overlap (58%; Fisher's exact test; $P < 0.01$) with our eight-cell activated genes (Figure S7).

Interestingly, the proportion of intronic sequences increased for longer transcripts after the eight-cell stage. This could result from less-efficient splicing of larger transcripts. Alternatively, intron delays (i.e., transcriptional delays implemented by intron length), in combination with the cell cycle constraint imposed by rapid cleavage in early embryos, may lead to early rounds of incomplete transcription of large genes (reviewed in Swinburne et al., 2008). Therefore, processed transcripts of large genes would be expected in more advanced stages, as observed in our study.

Among the genes first expressed at the four-cell stage we found the homologous gene *UBTFL1* (*LOC100140569*), which has been shown in mouse to play an essential role for the earliest stages of pre-implantation embryos (Yamada et al., 2010). Further, we identified *HNRNPA2B1* (*LOC516616*), which interacts with *SOX2* (Fang et al., 2011), a key transcription factor for embryonic stem cell pluripotency (Masui et al., 2007). Another gene activated before major EGA is *KLF17*. Its product, Krüppel-like factor 17, can activate or suppress transcription (Vliet et al., 2006). Array analyses of

polysomal mRNA from mouse one-cell embryos detected a markedly higher expression of *KLF17* compared with MII oocytes [National Center for Biotechnology Information (NCBI) GEO Profile: 3138385], indicating that the onset of expression of this gene during minor EGA is conserved between mouse and bovine.

In summary, our study provides a comprehensive transcriptome data set of bovine oocyte maturation and early embryonic development and detailed insight into the timing of embryonic activation of specific genes. This offers an additional level of information for studies in embryo biosystems research and for detecting disturbances of early development due to genetic, epigenetic, and environmental factors.

Methods

In vitro production of bovine embryos. In vitro production of bovine embryos was essentially done as described previously (Bauersachs et al., 2009). Commercially available semen from a Zebu bull was used for in vitro fertilization. Pools of 10 embryos were picked after visual inspection and snap-frozen in liquid nitrogen after washing in PBS. Stages collected for sequencing were denuded oocytes before and after maturation and embryos at the four-cell, eight-cell, 16-cell, and blastocyst stages.

Library preparation and sequencing. Frozen pools of 10 oocytes or embryos were thawed and lysed in 10 μl of Lysis Buffer (Prelude kit from NuGEN). cDNA was generated and amplified with the Ovation RNAseq v2 kit (NuGEN). In brief, 1 μl of the lysate was used for mixed random-/polyA-primed first-strand cDNA synthesis. After second strand synthesis the double-stranded cDNA was amplified by single primer isothermal amplification, and the amplified cDNA was bead-purified (AmpureXP, Beckman-Coulter) and fragmented by sonication (Bioruptor, Diagenode; 25 cycles 30 s on/30 s off). Five hundred nanograms of fragmented cDNA were used for preparation of Illumina-compatible sequencing libraries using the NuGEN Rapid library kit according to the manufacturer's protocol. Adapter ligation was done with sample-specific barcodes. The resulting library was amplified (KAPA hifi polymerase, eight cycles, 95°C 80 s, 55°C 30 s, 72°C 60 s) and quantified on a Bioanalyzer 2100 (Agilent). Barcoded libraries were pooled at 10-nM concentration for multiplexed sequencing. Three replicates of each stage were sequenced on an Illumina GAIIx to a mean coverage of 20×10^6 reads each. Sequencing runs were done in single-read mode with an 80-base read-length.

Preprocessing. For each replicate the raw reads (80 bases) from the Illumina Genome Analyzer IIx were filtered for adapter sequences. The first five bases were removed from each read because of random priming effects, and the reads were filtered from the 3' and 5' end with a quality cutoff of 20. Reads below a length of 30 were discarded.

Mapping and gene expression analysis. For each developmental stage and replicate the filtered reads were mapped with *Tophat2* (Kim et al., 2013) (v.2.0.3) to the bovine reference genome (UMD 3.1) supplied by annotated gene models in the GTF format from the online available *iGenomes* project of Illumina. Only uniquely mapped reads were used to calculate the number of reads falling into each gene with the *HTSeq-count* script (v.0.5.3) in the union mode and using no strand information from the *HTSeq* package. Differentially expressed genes were calculated with the *DESeq* package (Anders et al., 2010). Genes were regarded as differentially expressed between subsequent developmental stages when the adjusted P value was < 0.05.

Number of detectable genes in RNA-Seq. The mapped reads from each replicate were merged, and the numbers of reads falling into the exonic regions of the annotated genes were counted. A gene was determined as expressed if more than 15 reads could be properly aligned to that gene.

Genome activation by first expression. The number of reads calculated for each gene was used to analyze first expressed genes after fertilization. Genes were assumed to be first expressed in embryos if fewer than five reads were found in both oocyte stages (GV and MII) and at least 20 reads in one of the embryonic stages and if the transcript abundance in a particular embryonic stage was significantly higher (adjusted P value $<$ 0.05 with *DESeq*) than in the previous stage.

Genome activation by breed-specific SNPs. For SNP detection the uniquely mapped reads were used to generate a pileup for each replicate with *SAMtools* (Li et al., 2009) (v.0.1.13). From the resulting pileups, SNPs were called using *Varscan* (Koboldt et al., 2009) (v.2.3) with a minimum coverage of 1 and a minimum variant frequency of 0.01. SNPs occurring outside of the coding sequences of annotated genes were discarded. Furthermore, an SNP was considered only if the coverage was above 40 reads in both the Brahman $\times$ Simmental and the Jersey $\times$ Simmental hybrid embryos in all developmental stages. Last, a SNP was identified as breed specific, if it was absent

in all oocyte stages and in all stages of the Simmental × Jersey embryos. A valid SNP had to be verified by both strands of mapped reads. SNPs occurring in the first base of a read were discarded because this position is more artifact-prone. Genome activation was analyzed using the list of breed-specific SNPs. A breed-specific SNP was used for detection of genome activation if its minor allele frequency reached at least 20% in at least one of the replicates of the embryonic stages. This threshold was chosen to account for an expected frequency of a transcript with a paternal-specific SNP if the bull was heterozygous at this position. The probability of an SNP being called erroneously ($P = 0.00047$) was calculated for nucleotides differing from the expected alleles at all breed-specific positions. The P value was calculated with a binomial distribution B(n,p) as the probability of 1 up to n bases representing a paternal allele being called erroneously. At least two of three replicates at a developmental stage were required to have a P value for a paternal allele below 0.1. If an SNP fulfilled all of the above criteria, first occurrence of the paternal variant was considered as indicative of embryonic activation of the respective gene.

Genome activation by intronic reads. The mapped reads from the three replicates were counted as falling into intronic positions of a gene if at least 15 reads mapped to exons of the oocytes and at least six reads to the intronic part of the respective gene. To assess intronic coverage we counted all reads that completely mapped to introns as well as all positions that were not covered by any read using the *HTSeq-count* module with the intersection-strict parameter. The RINP value was used for detection of unspliced primary transcripts and was calculated by summing up all mapped intronic reads for each gene (or each intron as for Figure 3) divided by the positions (bases) where no coverage was observed. To distinguish between background noise and intronic expression the threshold was set to the 75th percentile of the RINP value obtained for MII oocytes (0.0014; Figure S3). Genes with RINP values below the threshold were discarded. All three replicates of a particular developmental stage were compared with the three replicates of the subsequent stage in all possible permutations, resulting in six sets of unique pairwise comparisons. A gene was considered as activated if the fold-change between subsequent stages was $\geq$10 in at least two out of three pairwise comparisons in all sets of permutations.

Global view on transcription of introns of all annotated genes. For the oocyte and embryonic stages the number of mapped reads for each intron was calculated from the merged three replicates. Only reads aligning exclusively to intronic sequences were counted. For each intron an RINP value was calculated, and data points were created as the distance from the transcription start site to the center of each intron. The annotated genes were ranked by their transcript length in descending order. For each intron with an RINP value above the background (Figure S3) a dot was plotted at the data point. The dot density was visualized with a colored scatter plot from the *LSD* package in R, with colors ranging from blue (low density) to red (high density).

Read distribution. The merged mapped reads of the replicates were used to determine the total number of mapped reads and the percentage of reads that could be assigned to exons, introns, ribosome, or reads containing polyA signals. The number of reads mapped to exons and introns was calculated with the *HTSeq-count* script in the union mode and intersection-strict mode, respectively. The reference for ribosomal sequences was obtained from NCBI and used to identify ribosomal reads by mapping with *Tophat2* (Kim et al., 2013), allowing multiple hits. The percentage of all reads mapping to ribosomal sequences was calculated, and multiple mapped reads were counted only once. Reads with polyA signals had to contain at least six polyA or polyT stretches at their 3' or 5' end and had to be properly aligned to the reference genome after trimming the stretches. The ratio of "unassigned" reads was determined on the basis of reads belonging to no other group.

Functional GO clustering. The *Cytoscape* plugin *ClueGO* (Bindea et al., 2009) was used to functionally group the genes activated in bovine four-cell embryos into GO terms "biological processes" as annotated for their human orthologs. The evidence was set to "Inferred by Curator (IC)", and the statistical test was set to a right-sided hypergeometrical test with a Bonferroni (step down) P value correction and a κ score of 0.3. The GO term restriction levels were set to 3–8, with a minimum of five genes or 1% genes in each GO term. The functional grouping was used with an initial group size of 2 and 50% for a group merge. To achieve a visualization of the eight-cell activated genes, the parameters of restriction levels were adjusted to 7–15 and the function "GO Term fusion" was additionally selected.

Acknowledgments. We thank Tuna Güngör and Sylvia Mallok for their excellent technical assistance, and Phillipp Torkler for his fruitful help to visualize the time course of intron transcription. This study was supported by the European Union Grant Plurisys, HEALTH-F4-2009-223485 FP7 Health 534 project, by the Deutsche Forschungsgemeinschaft (FOR 1041), and by BioSysNet.

Supporting information

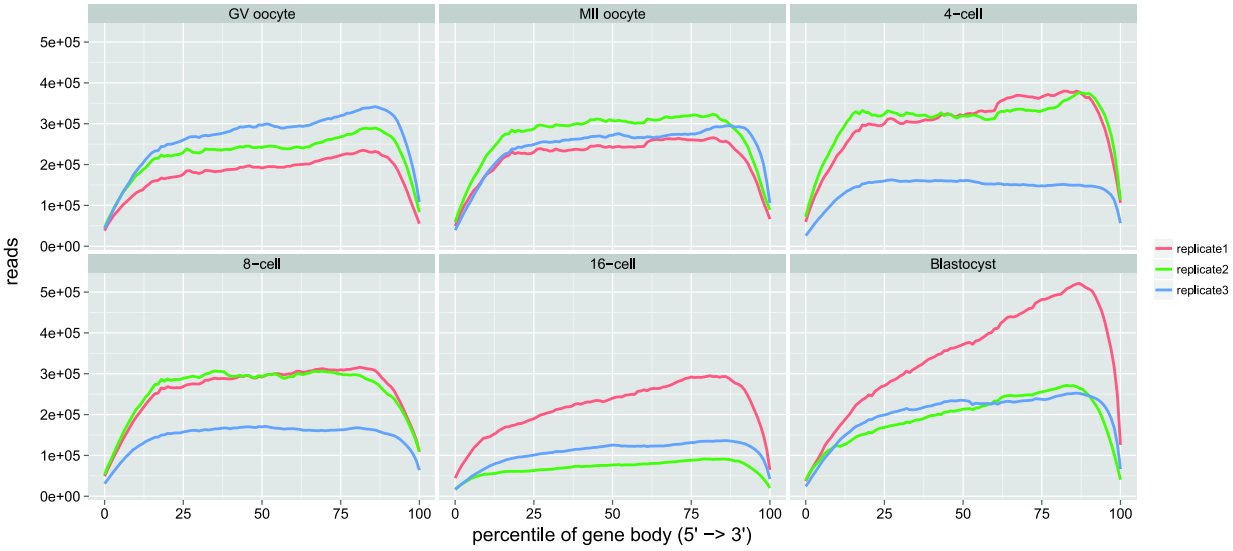

Fig. S1. Read coverage of annotated genes by RNA-Seq. The coverage of reads across each annotated gene was derived with *RSeQC* (Wang et al., 2012), which scales all genes to a uniform length (100 nt) and calculates the number of reads at each nucleotide position from the uniquely mapped reads. For each replicate of the oocyte and embryonic stages the read coverage for the annotated genes was summed an plotted in R.

Table S1. Number of filtered reads for the GV oocyte, MII oocyte, and embryonic stages for the three replicates.

Developmental stage	Replicate 1	Replicate 2	Replicate 3
GV oocyte	12,348,334 (74.6)	17,508,014 (74.2)	17,367,957 (74.0)
MII oocyte	15,439,935 (72.6)	20,664,911 (72.1)	17,194,714 (72.8)
Four-cell	22,381,046 (73.6)	27,684,867 (61.6)	10,368,947 (72.3)
Eight-cell	28,442,940 (72.7)	25,635,286 (61.6)	18,725,594 (70.0)
16-cell	19,049,174 (66.0)	8,866,945 (64.2)	17,973,667 (71.5)
Blastocyst	58,721,268 (41.8)	29,669,929 (49.8)	35,610,695 (54.5)

The percentage of uniquely mapped reads with *Tophat2* is shown in brackets.

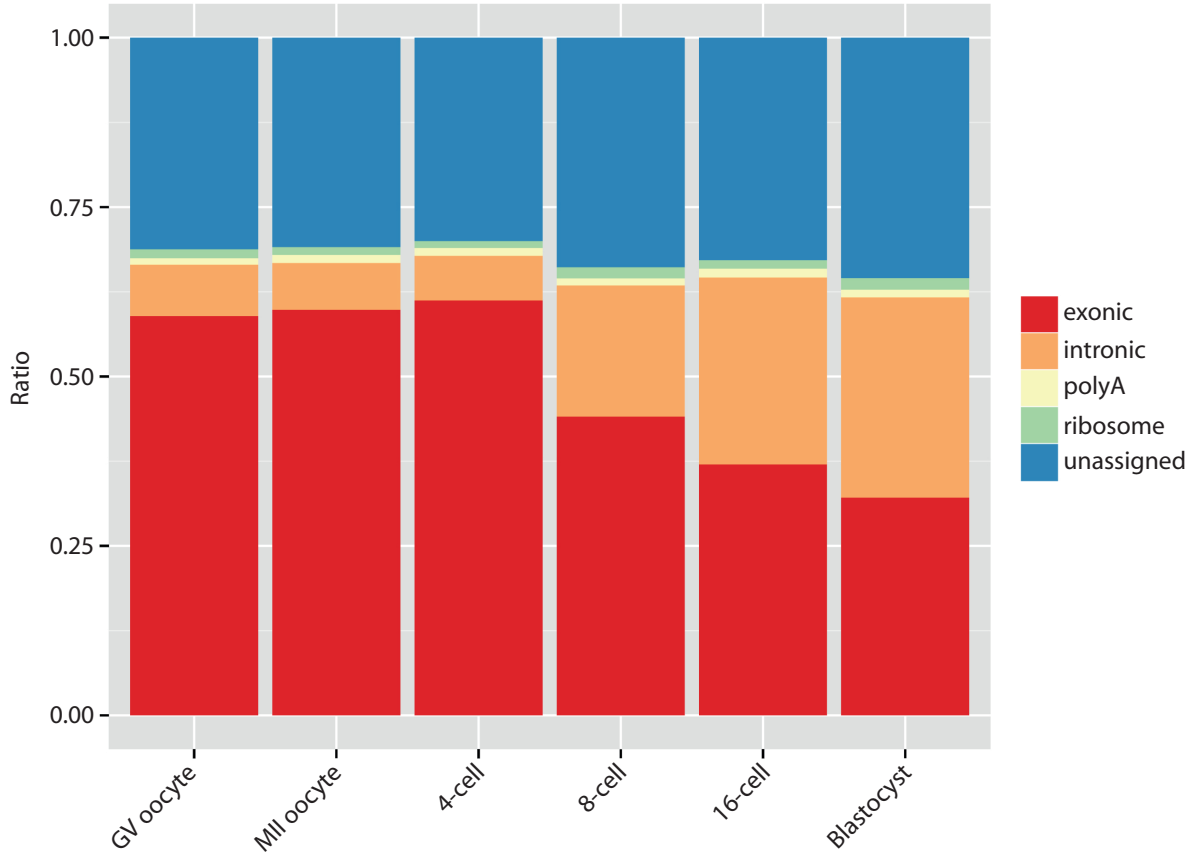

Fig. S2. Distribution of mapped reads. For the oocytes and the embryonic stages, the mapped reads of the three replicates were merged, and the percentages of exonic (red), intronic (orange), ribosomal (green), and unassigned mapped reads (blue) were calculated. The percentage of polyA reads (light yellow) was determined from those reads containing a stretch of at least six "A"s at one of the ends and that were then successfully mapped to the reference genome.

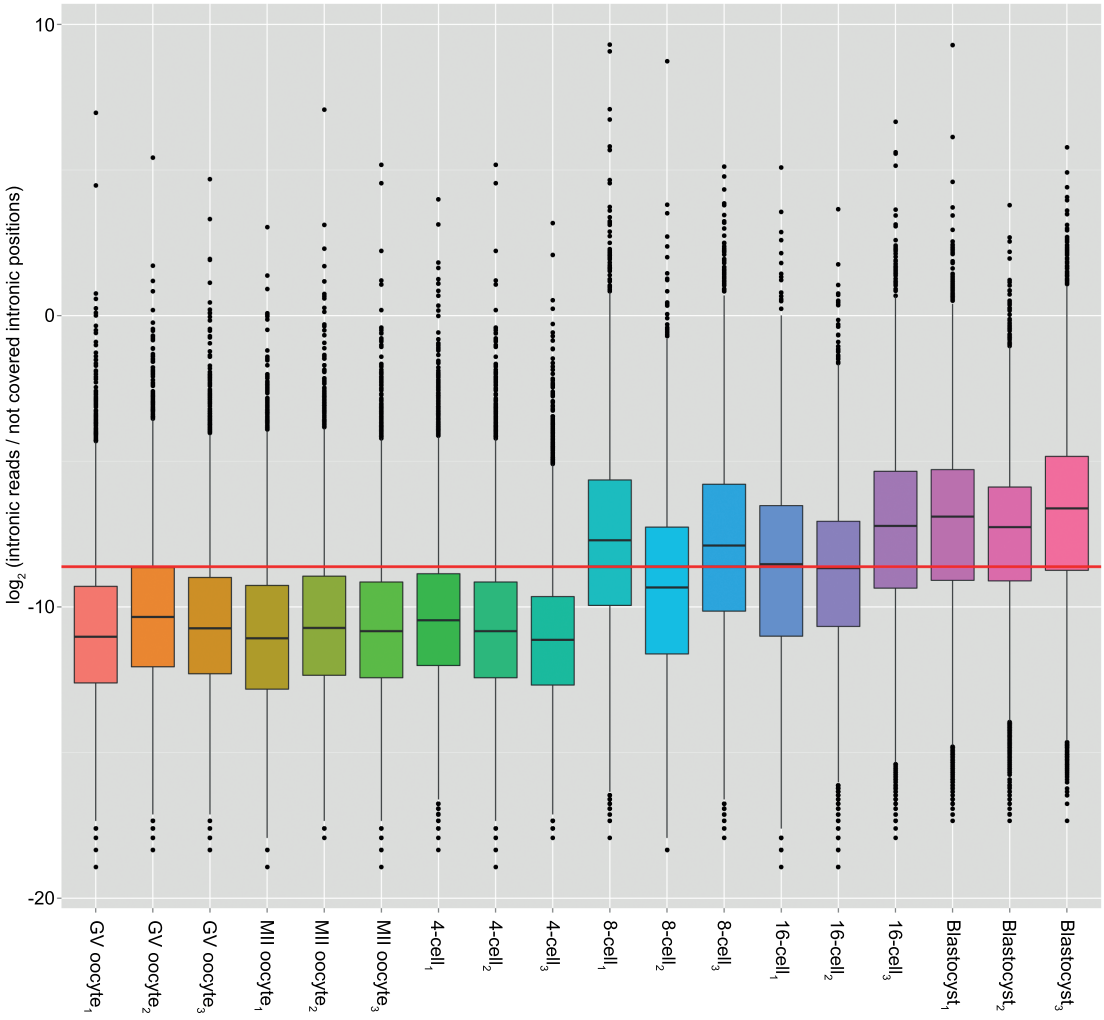

Fig. S3. Boxplots of **r**atios of **i**ntronic read counts to **n**ot-covered intronic **p**ositions (RINP) for all genes. For each replicate of the germinal vesicle (GV) and metaphase II (MII) oocyte, the four-cell, eight-cell, 16-cell, and blastocyst the RINP values were calculated for each gene and visualized in a boxplot. The background was determined as the 75th percentile of the GV oocyte as indicated by a red line. This RINP value corresponds to 0.0014.

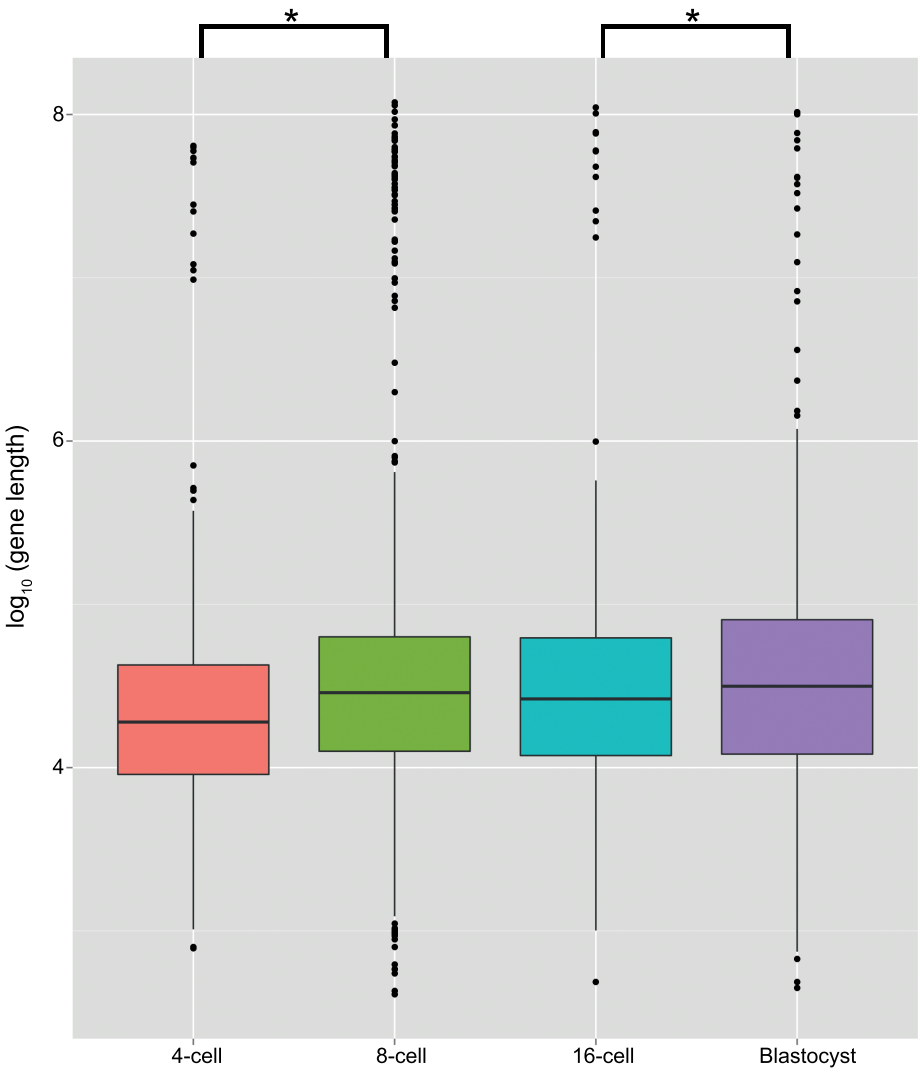

Fig. S4. Length of genes activated during early embryonic development. Boxplots for the gene length of embryonic activated genes. A significant increase of gene length is indicated with an asterisk and was observed between four-cell and eight-cell stages (Mann-Whitney test, $P < 0.01$) and between 16-cell and blastocyst stages (Mann-Whitney test, $P < 0.01$). No significant difference was observed between eight-cell and 16-cell stages ($P = 0.2$).

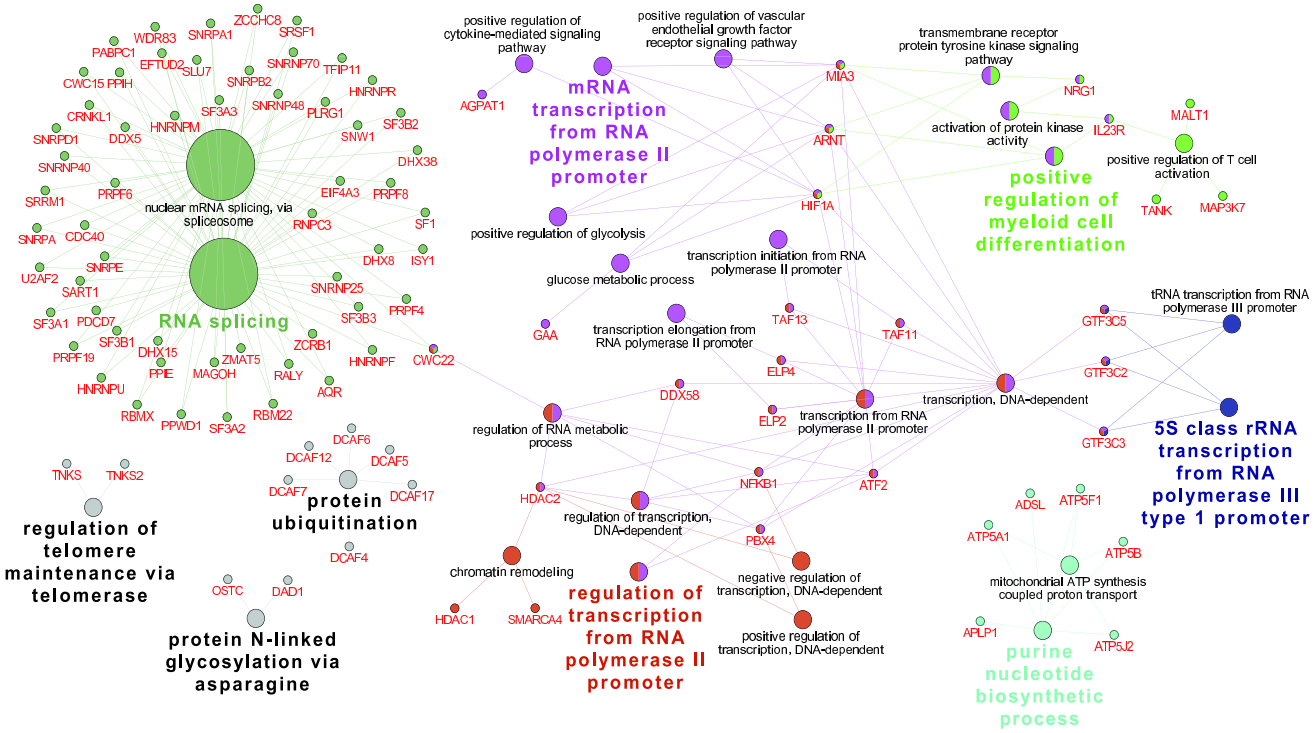

Fig. S5. Functionally grouped Gene Ontology (GO) terms for genes activated at major genome activation in bovine embryos. Genes activated at the eight-cell stage were detected by the presence of de novo transcripts (n = 129), transcripts with paternal-specific SNPs (n = 395), or primary transcripts with intronic sequences (n = 3965). The significantly enriched GO terms were analyzed with the *ClueGO* (Bindea et al., 2009) plugin from *Cytoscape*. Genes enriched in GO terms were colored in red. The significance of GO terms is reflected by the size of the nodes. The main significant GO terms were "RNA splicing", "mRNA transcription from RNA polymerase II promotor", "regulation of transcription from RNA polymerase II promotor", "positive regulation of myeloid cell differentiation", and "protein ubiquitination".

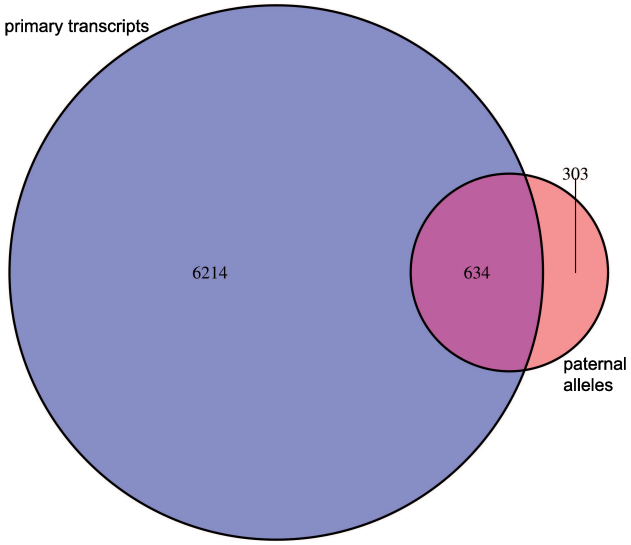

Fig. S6. Venn diagram showing the overlap of activated genes in early bovine embryos as detected by method 2 (appearance of transcripts from the paternal allele) and method 3 (appearance of primary transcripts with intron sequences).

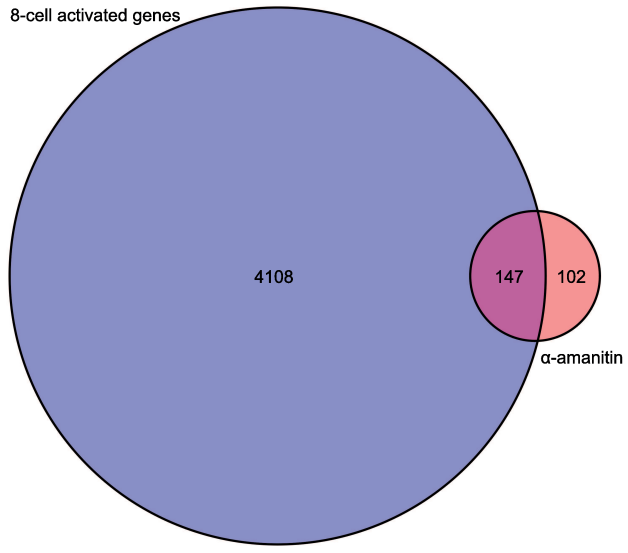

Fig. S7. Venn diagram comparing eight-cell activated genes identified in our study as first transcribed genes, by paternal alleles or primary transcripts with differentially expressed genes captured by a microarray analysis study of untreated bovine eight-cell embryos vs. eight-cell embryos treated with the transcription inhibitor α-amanitin (Vigneault et al., 2009).

References

Ameur, A., Zaghlool, A., Halvardson, J., Wetterbom, A., Gyllensten, U., Cavelier, L., Feuk, L. 2011. Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. Nat. Struct. Mol. Biol. 18 (12), 1435–1440.

Anders, S., Huber, W. 2010. Differential expression analysis for sequence count data. Genome Biol. 11 (10), R106.

Bauersachs, S., Ulbrich, S. E., Zakhartchenko, V., Minten, M., Reichenbach, M., Reichenbach, H. D., Blum, H., Spencer, T. E., Wolf, E. 2009. The endometrium responds differently to cloned versus fertilized embryos. Proc. Natl. Acad. Sci. U.S.A. 106 (14), 5681–5686.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W. H., Pages, F., Trajanoski, Z., Galon, J. 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 25 (8), 1091–1093.

Braude, P., Bolton, V., Moore, S. 1988. Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. Nature 332 (6163), 459–461.

Chitwood, J. L., Rincon, G., Kaiser, G. G., Medrano, J. F., Ross, P. J. 2013. RNA-seq analysis of single bovine blastocysts. BMC Genomics 14, 350.

Demyda-Peyrás, S., Dorado, J., Hidalgo, M., Anter, J., De Luca, L., Genero, E., Moreno-Millan, M. 2013. Effects of oocyte quality, incubation time and maturation environment on the number of chromosomal abnormalities in IVF-derived early bovine embryos. Reprod. Fertil. Dev. 25 (7), 1077–1084.

Fang, X., Yoon, J. G., Li, L., Tsai, Y. S., Zheng, S., Hood, L., Goodlett, D. R., Foltz, G., Lin, B. 2011. Landscape of the SOX2 protein-protein interactome. Proteomics 11 (5), 921–934.

Gad, A., Hoelker, M., Besenfelder, U., Havlicek, V., Cinar, U., Rings, F., Held, E., Dufort, I., Sirard, M. A., Schellander, K., Tesfaye, D. 2012. Molecular mechanisms and pathways involved in bovine embryonic genome activation and their regulation by alternative in vivo and in vitro culture conditions. Biol. Reprod. 87 (4), 100.

Katz, Y., Wang, E. T., Airoldi, E. M., Burge, C. B. 2010. Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat. Methods 7 (12), 1009–1015.

Khan, D. R., Dube, D., Gall, L., Peynot, N., Ruffini, S., Laffont, L., Le Bourhis, D., Degrelle, S., Jouneau, A., Duranthon, V. 2012. Expression of pluripotency master regulators during two key developmental transitions: EGA and early lineage specification in the bovine embryo. PLoS ONE 7 (3), e34110.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S. L. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14 (4), R36.

Koboldt, D. C., Chen, K., Wylie, T., Larson, D. E., McLellan, M. D., Mardis, E. R., Weinstock, G. M., Wilson, R. K., Ding, L. 2009. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics 25 (17), 2283–2285.

Kues, W. A., Sudheer, S., Herrmann, D., Carnwath, J. W., Havlicek, V., Besenfelder, U., Lehrach, H., Adjaye, J., Niemann, H. 2008. Genome-wide expression profiling reveals distinct clusters of transcriptional regulation during bovine preimplantation development in vivo. Proc. Natl. Acad. Sci. U.S.A. 105 (50), 19768–19773.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25 (16), 2078–2079.

Lonergan, P., Fair, T. 2008. In vitro-produced bovine embryos: dealing with the warts. Theriogenology 69 (1), 17–22.

Mamo, S., Mehta, J. P., McGettigan, P., Fair, T., Spencer, T. E., Bazer, F. W., Lonergan, P. 2011. RNA sequencing reveals novel gene clusters in bovine conceptuses associated with maternal recognition of pregnancy and implantation. Biol. Reprod. 85 (6), 1143–1151.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A. A., Ko, M. S., Niwa, H. 2007. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat. Cell Biol. 9 (6), 625–635.

Memili, E., Dominko, T., First, N. L. 1998. Onset of transcription in bovine oocytes and preimplantation embryos. Mol. Reprod. Dev. 51 (1), 36–41.

Misirlioglu, M., Page, G. P., Sagirkaya, H., Kaya, A., Parrish, J. J., First, N. L., Memili, E. 2006. Dynamics of global transcriptome in bovine matured oocytes and preimplantation embryos. Proc. Natl. Acad. Sci. U.S.A. 103 (50), 18905–18910.

Sirard, M. A. 2012. Factors affecting oocyte and embryo transcriptomes. Reprod. Domest. Anim. 47 Suppl 4, 148–155.

Swinburne, I. A., Silver, P. A. 2008. Intron delays and transcriptional timing during development. Dev. Cell 14 (3), 324–330.

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

Thorvaldsdóttir, H., Robinson, J. T., Mesirov, J. P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinformatics 14 (2), 178–192.

Troy, C. S., MacHugh, D. E., Bailey, J. F., Magee, D. A., Loftus, R. T., Cunningham, P., Chamberlain, A. T., Sykes, B. C., Bradley, D. G. 2001. Genetic evidence for Near-Eastern origins of European cattle. Nature 410 (6832), 1088–1091.

Vigneault, C., Gravel, C., Vallee, M., McGraw, S., Sirard, M. A. 2009. Unveiling the bovine embryo transcriptome during the maternal-to-embryonic transition. Reproduction 137 (2), 245–257.

Vliet, J. van, Crofts, L. A., Quinlan, K. G., Czolij, R., Perkins, A. C., Crossley, M. 2006. Human KLF17 is a new member of the Sp/KLF family of transcription factors. Genomics 87 (4), 474–482.

Wang, L., Wang, S., Li, W. 2012. RSeQC: quality control of RNA-seq experiments. Bioinformatics 28 (16), 2184–2185.

Wang, Z., Gerstein, M., Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10 (1), 57–63.

Wrenzycki, C., Herrmann, D., Lucas-Hahn, A., Korsawe, K., Lemme, E., Niemann, H. 2005. Messenger RNA expression patterns in bovine embryos derived from in vitro procedures and their implications for development. Reprod. Fertil. Dev. 17 (1-2), 23–35.

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

Yamada, M., Hamatani, T., Akutsu, H., Chikazawa, N., Kuji, N., Yoshimura, Y., Umezawa, A. 2010. Involvement of a novel preimplantation-specific gene encoding the high mobility group box protein Hmgpi in early embryonic development. Hum. Mol. Genet. 19 (3), 480–493.

Zhu, L., Zhang, Y., Zhang, W., Yang, S., Chen, J. Q., Tian, D. 2009. Patterns of exon-intron architecture variation of genes in eukaryotic genomes. BMC Genomics 10, 47.

Animal Reproduction Science

Genome activation in bovine embryos: Review of the literature and new insights from RNA sequencing experiments

Alexander Graf[a], Stefan Krebs[a], Mari Heininen-Brown[b], Valeri Zakhartchenko[c], Helmut Blum[a,**,1], Eckhard Wolf[a,b,c,*,1]

[a]Laboratory for Functional Genome Analysis (LAFUGA), Gene Center, Ludwig-Maximilians-Universität München, Munich, Germany

[b] Graduate School of Quantitative Bioscience (QBM), Gene Center, Ludwig-Maximilians-Universität München, Munich, Germany

[c]Chair for Molecular Animal Breeding and Biotechnology, Gene Center, Ludwig-Maximilians-Universität München, Munich, Germany

[1]Equal last author contributions

[*]Corresponding author at: Gene Center, LMU Munich, Feodor-Lynen-Str. 25, D-81377 Munich, Germany, Tel: +49 89 2180 76800

[**]Corresponding author at: Gene Center, LMU Munich, Feodor-Lynen-Str. 25, D-81377 Munich, Germany, Tel: +49 89 2180 76700

E-mail adresses: blum@lmb.uni-muenchen.de (H. Blum), ewolf@lmb.uni-muenchen.de (E. Wolf)

Reference: Anim Reprod Sci. 2014 Sep;149(1-2):46-58.

Abstract

Maternal-to-embryonic transition (MET) is the period in early embryonic development when maternal RNAs and proteins stored in the oocyte are gradually degraded and transcription of the embryonic genome is activated. First insights into the timing of embryonic genome activation (EGA) came from autoradiographic analyses of embryos following incorporation of [^{3}H]uridine. These studies identified the eight- to 16-cell stage of bovine embryos as the period of major EGA, but detected first transcriptional activity already in one-cell embryos. Subsequent studies compared the transcriptome profiles of untreated embryos and of embryos incubated with the transcription inhibitor α-amanitin to reveal transcripts of embryonic origin. In addition, candidate gene-based and global gene expression studies over several stages of early development were performed and characteristic profiles were revealed. However, the onset of embryonic transcription was obscured by the presence of maternal transcripts and could only be determined for genes which are not expressed in oocytes. Using RNA sequencing of bovine germinal vesicle and metaphase II oocytes, and of four-cell, eight-cell, 16-cell and blastocyst stage embryos, we established the most comprehensive transcriptome data set of bovine oocyte maturation and early development. EGA was analyzed by (i) detection of embryonic transcripts which are not present in oocytes; (ii) detection of transcripts from the paternal allele; and (iii) detection of primary transcripts with intronic sequences. Using these three approaches we were able to map the onset of embryonic transcription for almost 7,400 genes. Genes activated at the four-cell stage or before were functionally related to RNA processing, translation, and transport, preparing the embryo for major EGA at the eight-cell stage, when genes from a broad range of functional categories were found to be activated. These included transcriptional and translational functions as well as protein ubiquitination. The functions of the genes activated at the 16-cell stage were consistent with ongoing transcription and translation, while the genes activated in blastocysts included regulators of early lineage specification. Fine mapping of EGA provides a new layer of information for detecting disturbances of early development due to genetic, epigenetic, and environmental factors.

Introduction

The fusion of a male and a female gamete gives rise to an embryo. Initiation of development and the early embryonic developmental program is controlled by maternal transcripts and proteins produced and stored during oogenesis (reviewed in Tadros et al., 2009). In the mouse a number of so-called maternal effect genes have been discovered which are transcribed during oogenesis. Functions of their products include processing of the male genome after fertilization, degradation of maternal RNAs and proteins, and the activation of the embryonic genome (reviewed in Li et al., 2010). As development proceeds, control is switched from maternal to embryo-derived transcripts and proteins. This crucial process in development has been termed maternal-to-embryonic transition (MET) and involves the following events: depletion of maternal transcripts by degradation and translation; replacement of maternal transcripts stored in oocytes by embryonic transcripts, e.g. ribosomal RNAs; and the generation of new embryo-specific transcripts (reviewed in Sirard, 2010). In zebrafish and rainbow trout, specific microRNAs (miRNAs) produced by the embryo have been shown to be involved in the destruction of maternal transcripts (Giraldez et al., 2006; Ramachandra et al., 2008). A role of specific miRNAs in MET was also suggested for bovine embryos. Mondou et al. (2012) observed an increase in the abundance of the mature forms of miR-130a and miR-21 and of the precursor form of miR-130a from the one-cell to the eight-cell stage, correlated with MET. Transcriptional inhibition of two-cell embryos by exposure to α-amanitin decreased the abundances of miR-21, pre-miR-21, and miR-130a, suggesting that these miRNAs were – at least in part – of embryonic origin. The authors suggested that miR-21 and miR-130a are involved in gene regulation during MET and may play a role in the degradation of maternal mRNAs. Another miRNA which was found increased in abundance from the two-cell to the eight-cell stage of bovine embryogenesis, is miR-212. It was suggested as a negative regulator of maternal factor in the germ line alpha (*FIGLA*) transcripts during MET in bovine embryos (Tripurani et al., 2013). Other factors involved in the clearance of maternal transcripts during early development of metazoan embryos include RNA-binding proteins acting as specificity factors to direct the maternal degradation machinery to target mRNAs; signaling pathways that trigger production and/or activation of the clearance mechanism in early embryos; and mechanisms for spatial control of transcript clearance (reviewed in Walser et al., 2011).

During MET, nuclear reprogramming is required to activate the transcriptionally inactive embryonic genome, which lasts two hours in Drosophila and takes one, three, and up to 3,000 cell cycles in mouse, bovine and Xenopus embryos, respectively (reviewed in

Sirard, 2010). Oocyte-stored products play an essential role in this process by altering the chromatin structure (Østrup et al., 2013). The chromatin structure of early embryos, which impacts gene expression, can be altered by epigenetic modifications of DNA and histone proteins (Dean et al., 2001; Santos et al., 2003; Lepikhov et al., 2008; Wossidlo et al., 2011). Alterations in chromatin structure modulate the activity of transcription factors by permitting or restricting their access to regulatory elements of the genome, but are solely not sufficient to activate transcription. The oocyte cytoplasm plays also an important role in transcription activation by providing active transcription factors and RNA polymerase II (reviewed in Kanka, 2003).

The initiation of gene expression largely based on the products of an embryo is referred to as embryonic genome activation (EGA) and, as a part of MET, is the most important event in the pre-implantation development of mammals. The mechanisms regulating the onset of EGA are thought to be broadly conserved in mammals, despite species-specific differences in the timing of major EGA which ranges from the two-cell stage in mouse embryos (reviewed in Wang et al., 2006) to the four- to eight-cell stage in human (Braude et al., 1988) and pig embryos (reviewed in Sirard, 2012), and the eight- to 16-cell stage in bovine and rabbit embryos (Telford et al., 1990; Sirard, 2012). Cell cycle chronology, which is species-specific, and a cell cycle-dependent localization of RNA polymerase II in the nuclei are probably related to embryonic transcription (Marcucio et al., 1995) and may, at least partly, account for the differences in the onset of EGA.

EGA appears to start gradually and is preceded by an initial minor embryonic transcription. Gene expression studies during preimplantation mouse embryo development revealed three successive, overlapping waves of gene expression corresponding to minor EGA (one-cell stage), major EGA (two- to four-cell stage), and mid-preimplantation gene activation (MGA; four- to eight-cell stage). Subsequent waves of gene expression were found to be associated with morula compaction and blastocyst cavitation (reviewed in Wang et al., 2006). Genes involved in cell proliferation, mitotic cell cycle, regulation of transcription, DNA and protein metabolism were found to be early expressed (Kanka et al., 2012). Although specific mechanisms of the initiation of EGA remain to be elucidated, the involvement of some factors like maternal cyclin A2 (CCNA2), retinoblastoma protein (RB1), catalytic subunit of SWI/SNF related chromatin remodeling complex (BRG1) and sex determining region Y-box2 (SOX2), were recently suggested in a model of EGA in mouse (Kanka et al., 2012).

In bovine embryos, major EGA has been described to occur at the eight- to 16-cell stage, but the onset of EGA has not been precisely defined and varied dependent on

the respective techniques used for detecting embryonic transcription. Using [^{3}H]uridine incorporation after short incubation as an indicator, EGA in bovine embryos appeared to occur at the eight- to 16-cell stage (Camous et al., 1986; Frei et al., 1989). A further evidence for bovine EGA at the eight-cell stage was given by a study using polypeptide profiles of bovine in vitro embryos treated with α-amanitin (Barnes et al., 1991). The authors showed that these embryos were able to develop only up to the eight-cell stage, indicating the requirement of embryonic transcripts for further development at this stage. The eight-cell stage was also characterized by major changes in the structure of blastomere nucleoli, i.e. nucleolus precursor bodies (NPBs), electron-dense spherical masses of tightly packed fibrils, transformed into a fibrillo-granular nucleolus including formation of primary eccentric and secondary peripheral vacuoles (reviewed in Svarcova et al., 2007). However, after long-term exposure to [^{3}H]uridine, transcriptional activity could be already detected in two- to four-cell (Plante et al., 1994; Hyttel et al., 1996; Viuff et al., 1996; Memili et al., 1998) or even in one-cell bovine embryos (Memili et al., 1999), suggesting minor EGA already at these early stages of development. First ribosomal RNA transcription was visualized in four-cell embryos using a combination of fluorescence in situ hybridization (FISH) and silver staining (Viuff et al., 1998).

Subsequent studies aimed at the characterization of EGA in bovine embryos by using experimental approaches combining incubation with the transcription inhibitor α-amanitin and subsequent expression profiling by reverse transcriptase-polymerase chain reaction (RT-PCR) or array hybridization-based methods and are reviewed below.

Insights into EGA by RT-PCR and microarray studies

To investigate the onset of EGA with a resolution higher than global transcription initiation, bovine embryos from the one-cell to the expanded/hatched blastocyst stage were cultured with and without α-amanitin for four or 12 h, and stage-specific α-amanitin sensitive transcripts were evaluated by differential display (DD)-RT-PCR (Natale et al., 2000). Sensitivity of the DD-RT-PCR band pattern to α-amanitin was first detected at the two- to five-cell stage but became predominant following the six- to eight-cell stage of development. Only a few of the differential bands could be identified based on their similarity to human copy DNA (cDNA) sequences.

A subsequent study asked which genes are activated during early bovine embryogenesis. This question was addressed by comparing the transcriptome profiles of untreated eight-cell embryos with the profiles of embryos that developed to the eight-cell stage

in the presence of α-amanitin, and by comparing the transcriptome profiles of both groups of eight-cell embryos with that of metaphase II (MII) oocytes (Misirlioglu et al., 2006). Affymetrix BovineMidi euk2v3 Genome Arrays were used, which contained 24,072 probe sets representing about 23,000 transcripts, including assemblies from 19,000 UniGene clusters. About 50% and 53% of the probe sets were called as present in MII oocytes and in both groups of eight-cell embryos, respectively, with a P value of 0.04. For a total of 1,490 probe sets the hybridization signals were significantly ($P < 0.01$) stronger (fold-change ≥ 2) in untreated vs. α-amanitin treated eight-cell embryos. Among the significant probe sets were 233 specific for genes, and a gene ontology (GO) classification of those revealed the categories "Regulation of transcription", "Cell adhesion", "Apoptosis/cell death", "Protein folding", "Transport", "Metabolism", and "Immune response" as most prominent. Comparison of either group of eight-cell embryos with MII oocytes resulted in even higher numbers of differentially abundant transcripts, among them transcripts belonging to the GO categories "Regulation of transcription", "Chromatin modification", "Cell adhesion", "Apoptosis/cell death", "Protein folding" and "Signal transduction". The authors concluded that these changes in gene expression may result in a unique chromatin structure capable of maintaining totipotency during early embryogenesis and leading to differentiation during postimplantation development (Misirlioglu et al., 2006).

Another study used suppression subtractive hybridization (SSH; Clontech PCR-Select cDNA subtraction kit) followed by cDNA microarray hybridization to identify transcripts which are overrepresented in untreated as compared to α-amanitin treated bovine eight-cell embryos (Vigneault et al., 2009). An SSH library of more than 1,000 clones overexpressed in control vs. α-amanitin treated eight-cell embryos was spotted onto glass slides, which were hybridized with Alexa-Fluor® labelled amplified antisense RNA (aRNA) probes of either group of eight-cell embryos. The array analysis revealed 310 transcripts with at least twofold increased abundance in untreated vs. α-amanitin treated eight-cell embryos, among them 208 transcripts from genes with established function. GO classification of the differentially expressed genes found the primary molecular functions "RNA processing", "Regulation of transcription", "Protein biosynthesis", "Mitochondrial activity", "Protein modification/transport" and "Protein degradation" to be the most overrepresented (Vigneault et al., 2009).

Interestingly, a previous study comparing untreated with α-amanitin treated bovine embryos revealed that the presence of some key nucleolar proteins (RNA polymerase I, upstream binding factor, topoisomerase I, fibrillarin) was dependent on embryonic

transcription, whereas nucleolin and nucleophosmin were also detected in transcription inhibited embryos, although their localization was negatively affected by α-amanitin treatment (Svarcova et al., 2007). Silencing of nucleophosmin mRNA in in vitro produced bovine embryos using an RNA interference (RNAi) approach resulted in a marked reduction of nucleophosmin, but development to the blastocyst stage was not significantly impaired (Toralová et al., 2012). Localization of residual nucleophosmin in the RNAi treated embryos was comparable to control embryos, with only a slight delay in relocalization from the nucleoplasm to the nucleoli. The authors concluded that the amount of maternal nucleophosmin is sufficient for preimplantation development of bovine embryos.

Other approaches to study EGA in bovine embryos used expression profiling over several developmental stages. In a study by Vallee et al. (2009), transcriptome profiles of GV oocytes, and of in vivo derived embryos at the two-cell, eight-cell and blastocyst stage were analyzed using a custom bovine developmental microarray containing 1,153 cDNAs derived from four different subtracted libraries (GV oocytes – somatic tissues; GV oocytes – blastocysts; blastocysts – GV oocytes; blastocysts – somatic tissues). A pair-wise comparison of gene expression data for all developmental stages revealed the biggest difference between the transcriptome profiles of GV oocytes and blastocysts. The majority of the genes expressed during bovine preimplantation development could be grouped into three main clusters with characteristic stage-dependent changes of their transcript abundance: (i) genes with the highest transcript abundance in GV oocytes and a lower transcript abundance in embryos which was similar in all stages (72% of the genes); (ii) genes with a developmental expression profile as in cluster (i), but a further decreased transcript abundance in blastocysts (17% of the genes); and (iii) genes as in cluster (i), but with increased transcript abundance in blastocysts (10% of the genes). The authors concluded that the first two clusters represented maternal transcripts, whereas cluster (iii) reflected MET and EGA (Vallee et al., 2009). However the array used in this study represented only a small fraction of the genes that are actually expressed during early bovine embryonic development.

To overcome this problem, another study analyzing changes in transcriptome profiles during early bovine development used microarrays (Affymetrix GeneChip Bovine Genome Arrays) covering approximately 23,000 transcripts (Kues et al., 2008). The numbers of individual transcripts detected in MII oocytes, zygotes, two-cell, four-cell, eight-cell, morula and blastocyst stages ranged from 12.0 to 14.4 $\times$ 10^3. A pair-wise comparison of adjacent stages revealed the largest number of differentially expressed genes in

eight-cell vs. four-cell embryos, which is consistent with major EGA in bovine embryos at the eight- to 16-cell stage. However this study could not discriminate between maternal transcripts still present in eight-cell embryos and de novo transcripts synthesized by the embryo. The same is true for numerous studies analyzing developmental expression profiles of candidate genes by semi-quantitative or quantitative RT-PCR analyses (reviewed in Wrenzycki et al., 2005). These studies allowed the detection of embryonic gene activation only if corresponding transcripts were not present in oocytes.

RNA sequencing for the study of early bovine development

RNA sequencing (RNA-Seq) has a number of advantages as compared to hybridization-based techniques, such as microarrays, which compare only relative transcript abundances (reviewed in Wang et al., 2009). RNA-Seq enables the direct determination of the cDNA sequences from millions of short fragments, allowing transcriptome analyses at single nucleotide resolution. As compared to hybridization-based techniques, RNA-Seq has a higher sensitivity, a higher dynamic range, and less background (Wang et al., 2009). Importantly, RNA-Seq facilitates the discrimination of maternal and paternal transcripts by the detection of specific single nucleotide polymorphisms (SNPs) or other genetic markers.

A first RNA-Seq study of bovine embryos compared transcriptome profiles of pools of normal IVF blastocysts with degenerated ones (Huang et al., 2010). The authors were able to identify 4,426 alternative splice events, including exon skipping and alternative 5'- and 3'- splice sites within 2,032 genes, and revealed differential alternative splice events in several genes. Additionally, a large set of novel, not annotated transcriptional units (1,785 TUs) within the intergenic regions could be assembled at the blastocyst stage. Two thirds of these novel TUs could be assigned to expressed sequence tags (EST) and a large portion was supported by polyA containing reads. These findings revealed possibilities for more unannotated genes and alternative splicing events involved during embryonic development than previously thought.

Further studies used RNA-Seq to analyze the transcriptome of individual bovine blastocysts (Chitwood et al., 2013). The authors grouped the genes according to their level of expression and identified overrepresented functional categories of "Cytoskeleton", "Ribosomes" and "Mitochondria" for the highly expressed genes in bovine blastocysts. With this approach they were able to discriminate the gender of the sequenced embryo

and identified allelic imbalances which could be caused by imprinting where one of the parental alleles is silenced.

We performed an RNA-Seq study of bovine GV and MII oocytes, and of in vitro produced four-cell, eight-cell, 16-cell and blastocyst stage embryos (Graf et al., 2014). These were produced by in vitro fertilization of German Simmental (Bos taurus taurus) oocytes with sperm from a single bull of the genetically distant Brahman (Bos taurus indicus) breed to obtain a large number of SNPs for identification of the parental origin of transcripts. Furthermore, RNA-Seq libraries were produced without polyA+ selection enabling the identification of intronic sequences in transcripts, which can be found in de novo synthetised transcripts due to incomplete co-transcriptional splicing (Ameur et al., 2011) and can thus be used to discriminate them from spliced maternal transcripts stored in the oocyte (Graf et al., 2014). In all developmental stages investigated, transcripts from 12.4 to 13.7 $\times$ 10^3 different genes were detected. This is the same order of magnitude as compared to a previous microarray study of gene expression during early bovine development (Kues et al., 2008). Our differential gene expression analyses revealed only few transcripts to be differentially abundant between the GV and MII oocyte and between MII oocyte and four-cell stage. According to the main embryonic genome activation we found a marked increase in the number of differentially expressed genes between four- and eight-cell stage and even more between the later stages. The proportion of transcripts with decreased abundance gradually increased during embryonic development from 17% between four-cell embryo and MII oocyte to about 55% between eight- and 16-cell embryos, which could be due to the degradation of maternal transcripts (Figure 1).

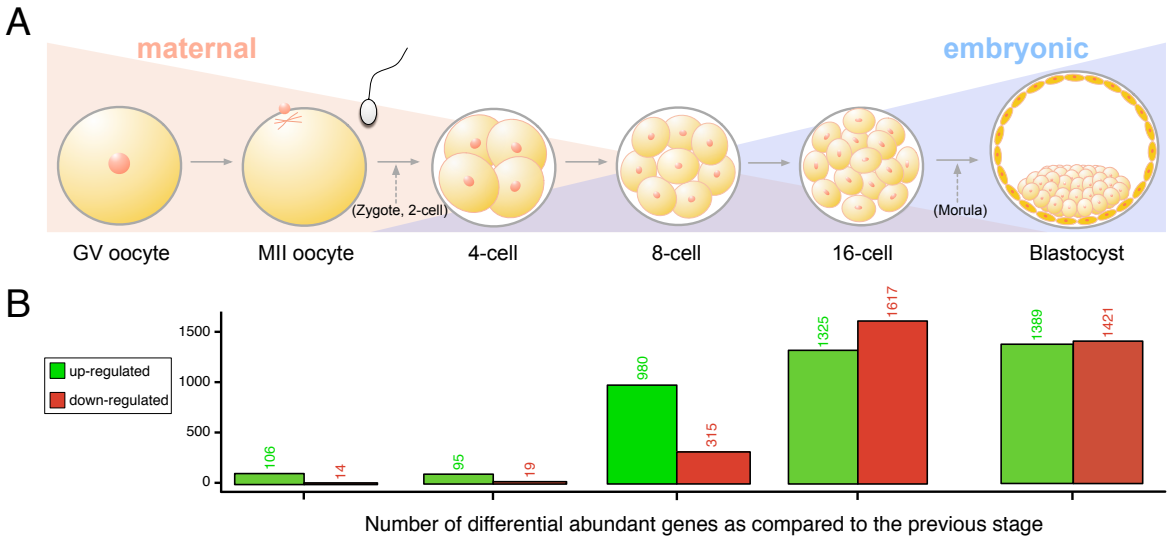

Fig. 1. Differential gene expression during bovine oocyte and early embryonic development as revealed by RNA-Seq. (A) Hybrid embryos were generated by fertilizing oocytes from B. t. taurus donors with sperm of a single B. t. indicus bull. For the GV and MII oocytes and four-cell, eight-cell, 16-cell and blastocyst stage, three pools of 10 per stage were harvested and used for RNA-Seq. (B) The numbers of differentially abundant transcripts between subsequent stages are visualized in the bar diagram. Green bars represent the numbers of transcripts with increased abundance, red bars the numbers of transcripts with decreased abundance as compared to the previous stage.

Strategies to identify EGA by RNA sequencing

In our RNA-Seq study of bovine oocytes (GV and MII) and early embryos (four-cell, eight-cell, 16-cell, and blastocyst) we used three different strategies to fine-map EGA (Graf et al., 2014): (i) detection of embryonic transcripts not present in oocytes; (ii) detection of paternal specific SNPs as a marker for the onset of EGA; and (iii) detection of incompletely spliced transcripts as an indicator of de novo transcription.

The most elementary approach to look for newly expressed genes during EGA is to identify transcripts that are not present in oocytes. In our study a gene was considered not to be expressed in oocytes if less than 5 reads were detected in the GV and MII stages. To be designated as first expressed in an embryonic stage, at least 20 specific reads for a gene had to be detected and the transcript had to be significantly more abundant than in the previous stage. Using this strategy, eight genes were found to be

first expressed at the four-cell stage, 129 genes at the eight-cell stage, 36 genes at the 16-cell stage, and 47 genes at the blastocyst stage.

The cross breeding design (Bos t. taurus × Bos t. indicus) used to produce the model embryos for our study enabled the analysis of EGA by detection of transcripts from the paternal allele of the Brahman bull. Therefore, paternal (Brahman-specific) SNPs were identified genome wide and limited to the coding regions of genes. An SNP was regarded as paternal, if it was neither found in Bos t. taurus oocytes, nor in any sequenced Bos t. taurus × Bos t. taurus cross breed embryos (Jersey × German Simmental). The first occurrence of a breed-specific SNP was indicative of EGA at a particular developmental stage. The paternal allele approach revealed activation of 16 genes at the four-cell stage, 395 genes at the eight-cell stage, 314 genes at the 16-cell stage, and 212 genes at the blastocyst stage.

In our research design a combination of oligo-dT and random primers was used for cDNA synthesis to cover the whole transcriptome except for ribosomal RNAs. With this approach, we analyzed EGA by the detection of intron derived reads indicating nascent transcription at a particular developmental stage. This strategy permitted the detection of 390 genes that were activated at the four-cell stage, 3,965 genes at the eight-cell stage, 628 genes at the 16-cell stage, and 1,865 at the blastocyst stage, making it the method that identified the largest proportion of embryonic activated genes.

A detailed examination of the intronic transcripts of all annotated genes for the oocyte and embryonic stages revealed a marked increase in primary transcripts around the eight-cell stage, concordant with major EGA in bovine embryos. Interestingly, the proportion of intronic sequences increased in longer genes from the eight-cell stage to the blastocyst stage (Graf et al., 2014), which could be explained by intron delays together with rapid cleavage of early embryos, resulting in early rounds of incomplete transcription of large genes (Swinburne et al., 2008). Alternatively, less-efficient splicing could be responsible for the increased proportion of intronic sequences in longer transcripts after the eight-cell stage.

Genes activated at the four-cell stage or before

Transcripts in early embryos can be of maternal or embryonic origin. This makes it challenging to distinguish the true origin by conventional techniques. RNA-Seq combined with specific analysis strategies uncovered new information about genes activated in different cell stages. These novel strategies enabled the identification of genes activated

before major EGA, which might play a role in preparing the blastomeres for major EGA. A functional classification of the 414 genes found to be activated in bovine embryos at the four-cell stage or before revealed the GO terms "RNA processing", "Translation" and "Transport", indicating the first transcriptional activity before major EGA (Graf et al., 2014).

Among the first expressed genes was *HNRNPA2B1* encoding heterogeneous nuclear ribonucleoprotein A2/B1, which is known to interact with the pluripotency transcription factor SOX2 (Fang et al., 2011). In addition, first expression of the gene *KLF17* coding for Krüppel-like factor 17, which can activate transcription (Vliet et al., 2006), was found at the four-cell stage. *Klf17* was first discovered as a germ cell-specific gene encoding zinc finger protein 393 (Zfp393) in mouse (Yan et al., 2002). The pluripotency related gene *KLF17* was found to be downregulated during the transition of epiblast from human blastocysts to primary embryonic stem (ES) cells (Yan et al., 2013). In zebrafish embryos, the expression of *Klf17* is regulated by POU5F1/OCT4 and is involved in the specification of the extraembryonic enveloping layer (Kotkamp et al., 2014).

Although corresponding transcripts were already detected in oocytes, the activation of embryonic transcription of the H2A histone family, member Z (*H2AFZ*) gene, the epithelial cell adhesion molecule (*EPCAM*) gene, the histone-lysine N-methyltransferase (*MLL2*) gene, and the Yamaguchi sarcoma viral oncogene homolog 1 (*YES1*) gene was found at the four-cell stage by the detection of primary (intron-containing) transcripts. The histone variant H2AFZ affects chromatin structure, promotes the expression of pluripotency genes, and functions in general as a regulator of gene expression, chromosome segregation and X-chromosome inactivation. Homozygous disruption of *H2afz* in mouse embryos resulted in an abnormal inner cell mass and early embryonic lethality (Faast et al., 2001). Knockdown experiments of *H2afz* in mouse ES cells demonstrated that H2AFZ is required for both self-renewal and differentiation (Hu et al., 2013). Inactivation of the *Epcam* gene in mice led to a delay of embryonic development, prominent placental abnormalities, and death in utero by E12.5 (Nagao et al., 2009), arguing against an essential role of this gene during preimplantation development. However, knockdown experiments in mouse ES cells demonstrated that EPCAM is essential for maintaining their self-renewal (Gonzalez et al., 2009). The histone-lysine N-methyltransferase MLL2 regulates transcription by histone modifications and is critical for early development. *Mll2*$^{-/-}$ mouse embryos showed retarded growth and development from E6.5 and died before E11.5 (Glaser et al., 2006). In contrast, conditional *Mll2* knockout experiments demonstrated that MLL2 is dispensable after E11.5 of mouse development (Glaser et al.,

2009). While disruption of the *Yes1* gene in mice did not result in an overt phenotype (Stein et al., 1994), YES1 was shown to play a role in self-renewal and maintenance of ES cells and in the regulation of the pluripotency genes *Pou5f1/Oct4* and *Nanog* (Tamm et al., 2011).

Genes activated at the eight-cell stage

While it was known from previous studies that major EGA in bovine embryos occurs at the eight- to 16-cell stage, our RNA-Seq study was able to detect the onset of embryonic expression for 4,489 specific genes at the eight-cell stage. GO classification of the eight-cell activated genes revealed the functional categories "Transcription, DNA-dependent", "Purine nucleotide metabolic process", "Protein ubiquitination", "Translational initiation" and most prominently "RNA metabolic process" to be significantly overrepresented (Figure 2). These GO terms correlate well with the major EGA events such as the initiation of transcription and translation and with the continuous degradation of maternally stored RNAs and proteins at this developmental stage.

Within the large number of genes which were activated at the eight-cell stage, we found genes relevant for the maintenance of pluripotent cells in early embryos or of ES cells. Among them was sal-like 4 (Drosophila) (SALL4) which was shown to be important for cell fate decisions in early mouse embryos (Zhang et al., 2006). *SALL4* is an essential regulator of *Pou5f1/Oct4* expression and is required to maintain pluripotency of the inner cell mass of blastocysts. The proto-oncogene *MYC* was also found to be activated in bovine eight-cell embryos. A recent study on mouse embryos indicated that MYC levels are heterogeneous among epiblast cells, and that competition refines the epiblast cell population through the elimination of cells with low MYC levels, eventually contributing to the selection of the epiblast cell pool in the early mammalian embryo (Claveria et al., 2013). Another eight-cell activated gene in bovine embryos was *TBX3* (T-box 3). In the mouse, low levels of *Tbx3* transcripts were found in MII oocytes as well as in two- and four-cell embryos, higher levels in eight-cell embryos and morulae, and highest levels in blastocysts (Guo et al., 2010). The pluripotency-related transcription factor TBX3 was shown to be essential for the maintenance of self-renewal of mouse ES cells and for their differentiation into extraembryonic endoderm (Lu et al., 2011). Further, transcription of the Krüppel-like factor 4 (*KLF4*) gene in bovine embryos was first detected at the eight-cell stage. KLF4 was shown to prevent differentiation of mouse ES cells and to regulate the expression of *Nanog* (Zhang et al., 2010). However KLF4

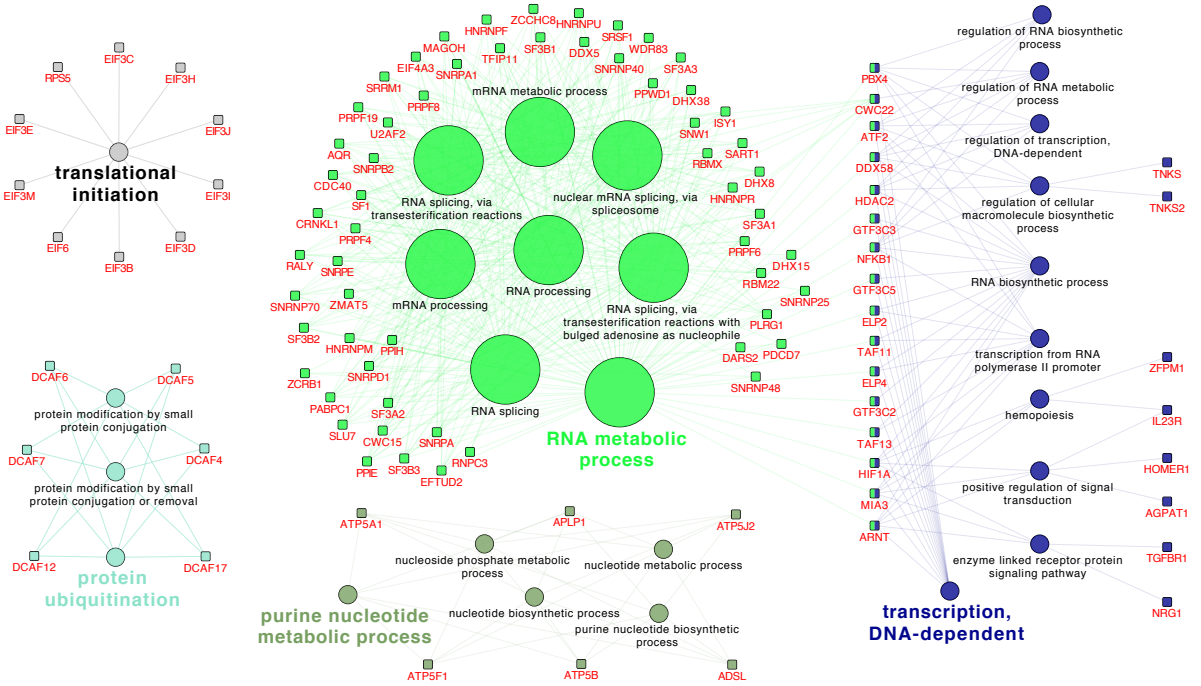

Fig. 2. Functionally grouped Gene Ontology (GO) terms for genes activated at the eight-cell stage. The Cytoscape (Kohl et al., 2011) plugin ClueGO (Bindea et al., 2009) was used to group the genes activated at the eight-cell stages into functional GO terms of 'biological processes'. A right-sided hypergeometrical test with a Bonferroni correction was used with a kappa score of 0.3. The genes were annotated for their human orthologs and the evidence was set to 'Inferred by Curator (IC)' with the GO term restriction levels of 6–15 with a minimum of 5 or 1% genes in each GO term. The initial group size for functional grouping was 2 and 50% for a group merge. The analysis of the 4,489 genes activated at this developmental stages revealed the GO term "RNA metabolic process" as the most prominent and additionally the GO terms "Translation initiation", "Protein ubiquitination", "Purine nucleotide metabolic process", and "Transcription, DNA-dependent". The significance of the GO terms is reflected by the size of the nodes and assigned genes are represented in squares and colored according to their respective GO terms. Shared genes between GO terms are represented in multiple colors.

does not seem to be essential for early development, since *Klf4*$^{-/-}$ mouse embryos develop to term, but die shortly after birth due to a skin defect that results in loss of fluids (Katz et al., 2002). While *SALL4*, *MYC*, *TBX3*, and *KLF4* transcripts were already present in bovine oocytes, *NANOG* (Nanog homeobox) was found to be first expressed in bovine eight-cell embryos, which is in line with a previous report (Khan et al., 2012). NANOG is essential for early development, as mouse embryos with a homozygous disruption of the *Nanog* gene die between E3.5 and E5.5 and are characterized by abnormal embryonic and extraembryonic tissue development (Mitsui et al., 2003). A recent study involving depletion experiments of POU5F1/OCT4, SALL4 and NANOG in early mouse embryos

revealed that an integrated *Pou5f1/Oct4-Sall4-Nanog* regulatory network of protein-coding genes and microRNAs is required to govern progression in pre-implantation development (Tan et al., 2013).

In addition to genes relevant for embryonic cell fate decisions, we found embryonic transcripts of a number of genes coding for subunits of the eukaryotic translation initiation factor 3 (EIF3) complex in bovine eight-cell embryos (Figure 2). The EIF3 complex is composed of 13 subunits, which are organized in the linked modules A (EIF3A, EIF3B, EIF3G, EIF3I), B (EIF3F, EIF3H, EIF3M), and C (EIF3C, EIF3D, EIF3E, EIF3K, EIF3L). This complex is required for several steps in the initiation of protein synthesis. It associates with the 40S ribosome and facilitates the recruitment of EIF1, EIF1A, EIF2:GTP:methionyl-tRNAi and EIF5 to form the 43S preinitiation complex (43S PIC). The EIF3 complex stimulates mRNA recruitment to the 43S PIC and scanning of the mRNA for AUG recognition. The EIF3 complex is also required for disassembly and recycling of post-termination ribosomal complexes and subsequently prevents premature joining of the 40S and 60S ribosomal subunits prior to initiation (reviewed in Hinnebusch, 2006). Thus expression of multiple EIF3 subunit genes in bovine eight-cell embryos prepares them for efficient translation initiation at the time of major EGA.

Another group of genes activated in bovine eight-cell embryos belongs to the family of DDB1 and CUL4-associated factors (DCAFs) (Figure 2). The CUL4-DDB1 ubiquitin ligase regulates cell proliferation, survival, DNA repair, and genomic integrity through targeted ubiquitination of key regulators, and DCAFs have been recently identified as substrate receptors which dictate the specificity of this ubiquitination machinery (reviewed in Lee et al., 2007).

Further, genes encoding subunits of the mitochondrial ATP synthase (ATP5A1, ATP5B, ATP5F1, ATP5J2) were found to be switched on in bovine eight-cell embryos (Figure 2). Mitochondrial membrane ATP synthase produces ATP from ADP in the presence of a proton gradient across the membrane which is generated by electron transport complexes of the respiratory chain (reviewed in Jonckheere et al., 2012). In addition to many other cellular functions, ATP is essential for nucleotide biosynthesis and metabolism (Pelley et al., 2011).

Genes of the DEAD- and DEAH-box families (*DDX5*, *DHX8*, *DHX15*, *DHX38*) were also activated at the eight-cell stage of bovine embryos (Figure 2). The putative RNA helicases of the DEAD/DEAH-box family were shown to be involved in RNA metabolism, such as transcription initiation, ribosome biogenensis and pre-mRNA splicing, and are

believed to have a function in differentiation processes (Abdelhaleem et al., 2003), consistent with major EGA at the eight-cell stage.

Another group of genes activated in bovine eight-cell embryos is functionally related to DNA methylation and chromatin structure. Examples are *DNMT3B* coding DNA (cytosine-5-)-methyltransferase 3 beta and *MTA2* coding metastasis tumor antigen 2. DNMT3B has been shown to be essential for development because *Dnmt3b*$^{-/-}$ mouse embryos had multiple developmental defects including growth impairment and rostral neural tube defects (Okano et al., 1999). MTA2 is a component of the nucleosome remodeling and histone deacetylation (NuRD) complex and essential for the maintenance of monoallelic expression of some imprinted genes in mouse blastocysts (Ma et al., 2010). Additionally, the SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4 (*SMARCA4*) gene was found to be activated in bovine eight-cell embryos. The chromatin remodeling factor SMARCA4 has been shown to be involved in trophoblast stem cell maintenance, pluripotency and self-renewal (Kidder et al., 2010).

Genes activated at the 16-cell stage

The GO classification of genes activated in later developmental stages revealed new roles and functions as compared to the earlier stages. As certain general processes have already been initiated at the four- and eight-cell stages, the functions of the genes activated at the 16-cell stage become diversified and targeted. The 978 genes could be classified into the GO categories "RNA splicing", "Regulation of glycolysis", "ATP biosynthetic process", "Negative regulation of transcription", "5S class rRNA transcription from RNA polymerase III type1 promoter", "Carbon catabolite regulation of transcription from RNA polymerase II promoter", "Cytosolic calcium ion homeostasis", and "Transcription initiation/elongation from RNA polymerase II promoter", implicating the maintenance of transcription and translation and the initiation of metabolic processes (Figure 3).

First activation of the elongator acetyltransferase complex subunit 2 (*ELP2*) gene was found at the 16-cell stage. Studies in embryonal carcinoma P19 cells revealed that ELP2 and ELP3 bind to SF1 which activates the *Pou5f1/Oct4* promoter (Barnea et al., 2000). Tumor susceptibility gene 101 (*TSG101*) was also found to be activated at the 16-cell stage. It is required for cell proliferation before gastrulation and is involved in the regulation of the p53 pathway. *Tsg101*$^{-/-}$ mouse embryos showed a decreased cell proliferation in the ICM and died around E6.5 (Ruland et al., 2001). Furthermore, first expression

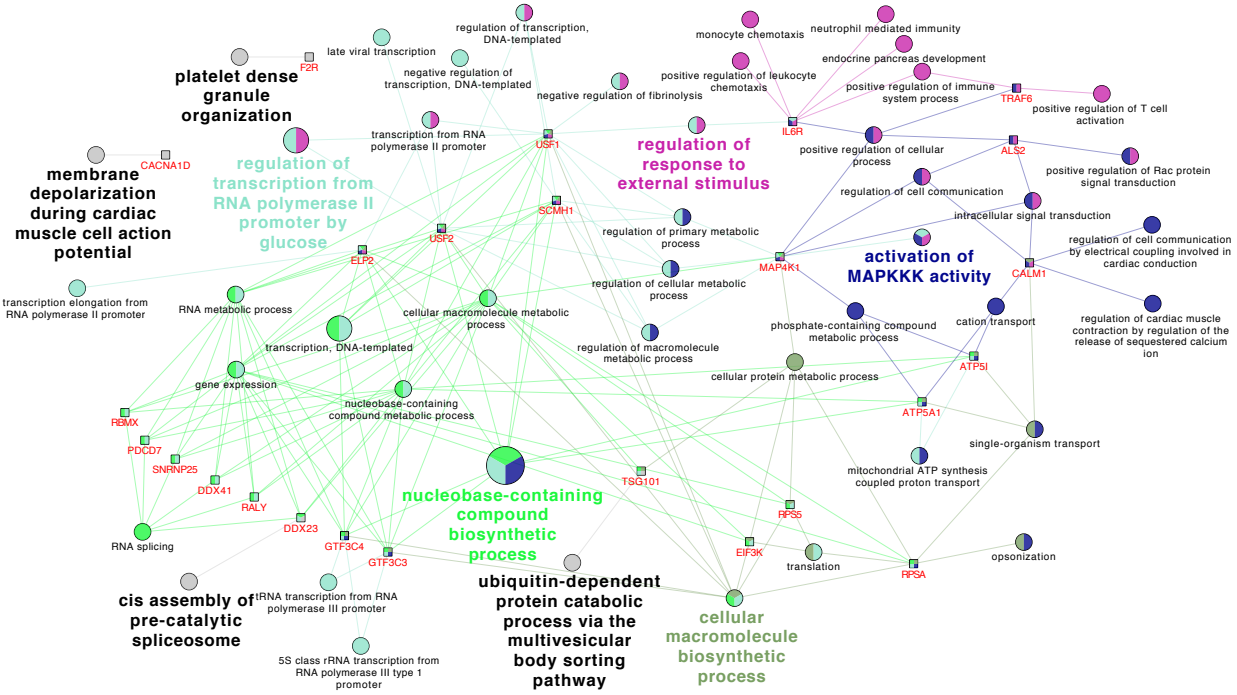

Fig. 3. Functionally grouped Gene Ontology (GO) terms for genes activated at the 16-cell stage. The analysis of the 978 genes activated at the 16-cell stage revealed the GO terms "RNA splicing", "Regulation of glycolysis", "ATP biosynthetic process", "Negative regulation of transcription", "5S class rRNA transcription from RNA polymerase III type1 promoter", "Carbon catabolite regulation of transcription from RNA polymerase II promoter", "Cytosolic calcium ion homeostasis", and "Transcription initiation/elongation from RNA polymerase II promoter". The analysis was done as described in Figure 2. The GO term restriction level was set to 10–15 and the minimum of genes in each GO term was set to 2 or 10%. The significance of the GO terms is reflected by the size of the nodes and assigned genes are represented in squares and colored according to their respective GO terms. Shared genes between GO terms are represented in different colors.

of *TEAD4* coding for TEA domain family member 4 was found at the 16-cell stage, which is required for the development and specification of the trophectoderm. *Tead4*$^{-/-}$ mouse embryos were not able to form clear blastocoels and died at pre-implantation stages (Nishioka et al., 2008). The authors suggested different regulatory functions for TEAD4, such as up-regulation of *Cdx2* in the outer embryonic cells and subsequent down-regulation of *Pou5f1/Oct4*, leading to trophectoderm development.

Additionally, we found activation of the genes chromodomain helicase DNA binding protein 2 (*CHD2*) and bystin-like (*BYSL*) at the 16-cell stage. CHD2 is known to function in chromatin remodeling and it was shown that CHD2 deficient mice embryos displayed growth delays which lead to lethality around birth (Marfella et al., 2006), whereas BYSL is an important factor for the 40S ribosome subunit biogenesis during

the development of preimplantation embryos and is required for blastocyst formation. When mouse embryos at the 16-cell stage were treated with *Bysl* siRNA, differentiation of the trophectoderm was impaired (Adachi et al., 2007).

Further, the gene fuzzy planar cell polarity protein (*FUZ*) was activated at the 16-cell stage. Studies in mice show that FUZ is important for membrane trafficking and is involved in the planar cell polarity signaling pathway which is essential for embryonic development (Gray et al., 2009).

Another 16-cell activated gene was *FOSL1* encoding the Fos-like antigen 1, which acts downstream of the PI3K/AKT signaling pathway, is required for development of the trophoblast lineage, and is a key regulator of trophoblast placentation (Kent et al., 2011).

In addition, we found the genes programmed cell death protein 2 (*PDCD2*) and sal-like 1 (*SALL1*) among the 16-cell activated genes. *Pdcd2*$^{-/-}$ mouse blastocysts showed an interruption of ICM growth, and it is suggested that PDCD2 is involved in self-renewal of pluripotent embryonic cells (Mu et al., 2010). Besides PDCD2, SALL1 was shown to interact with NANOG and SOX2 and to play a role in stem cell pluripotency (Karantzali et al., 2011). The authors showed that SALL1 and NANOG bind together to a large number of self-renewal and differentiation related common target genes of mouse ES cells and that silencing or overexpression of *Sall1* during differentiation had effects on the expression patterns of certain differentiation markers.

Genes activated after the 16-cell up to the blastocyst stage

During embryonic development the morula differentiates into the blastocyst which consists of two morphologically different cell types: the inner cell mass (ICM), which later forms the embryo, and the outer layer forming trophectoderm, which contributes to the placenta. Our functional analysis of the 2,214 genes activated after the 16-cell up to the blastocyst stage revealed GO terms like "Intracellular transport", "Regulation of multicellular organismal process", "Nucleobase-containing compound metabolic process", and "Negative regulation of response to stimulus" as overrepresented (Figure 4). The terms clearly indicate first differentiation processes and the ongoing transcription and translation required for cell specification. Interestingly, the gene homeobox protein CDX-2 (*CDX2*) was among the activated genes. Studies of mouse ES cells showed that CDX2 expression represses *Pou5f1/Oct4* to induce differentiation into the trophectoderm lineage (Niwa et al., 2005). In bovine embryos, *CDX2* was shown to be

important for trophectoderm lineage maintenance as well, but it did not influence *OCT4* levels, suggesting a different role of *CDX2* in embryonic development between mice and cattle (Berg et al., 2011). Besides *CDX2*, the *GATA3* gene was activated during these stages. The transcription factor GATA3 induces trophoblast differentiation and regulates *Cdx2* transcription (Ralston et al., 2010), confirmed by knock-down of *Gata3* in mouse pre-implantation embryos which were stalled during the transition from morula to blastocyst (Home et al., 2009). Additionally, the gene SNW domain-containing protein 1 (*SNW1*) was activated in blastocysts. SNW1 is known to regulate gene expression, and functional screens in Xenopus and zebrafish identified SNW1 as key regulator of bone morphogenetic protein (BMP) activity at the end of gastrulation which is involved in neural crest specification in vertebrates (Wu et al., 2011).

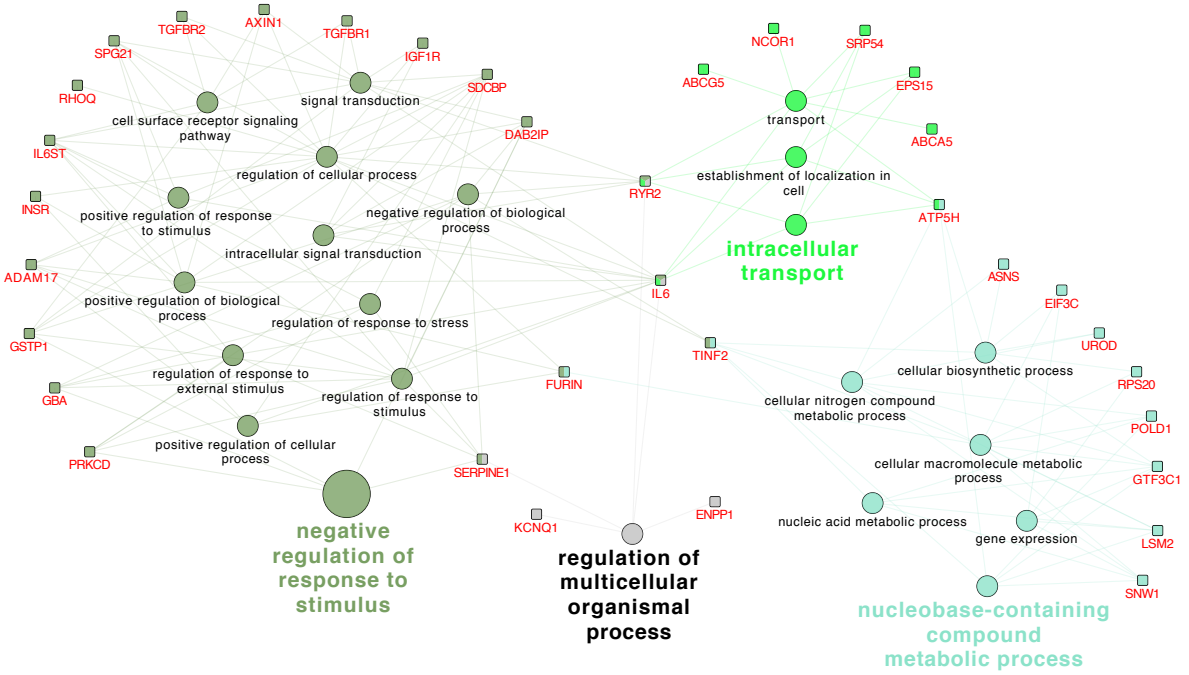

Fig. 4. Functionally grouped Gene Ontology (GO) terms for genes activated after the 16-cell up to the blastocyst stage. The analysis of the 2,124 genes activated between the 16-cell and blastocyst stages was achieved as described in Figure 2 with GO term restriction levels set to 4–8. The functional grouping revealed the GO terms "Negative regulation of response to stimulus" as the most prominent and additionally the GO terms "Intracellular transport", "Regulation of multicellular organismal process" and "Nucleobase-containing compound metabolic process". The significance of the GO terms is reflected by the size of the nodes and assigned genes are represented in squares and colored according to their respective GO terms. Shared genes between GO terms are represented in different colors.

Another activated gene was *FURIN* (furin (paired basic amino acid cleavage enzyme)), a member of the proprotein convertase family, which converts precursor proteins into

their active product. It has been shown in human and mice, that FURIN is important for embryonic development, as it plays a role in placental development and trophoblast fusion by processing insulin-like growth factor 1 receptor (IGF1R) and the vascular endothelial growth factor (VEGF) (Zhou et al., 2013). Deletion of FURIN in mouse embryos leads to lethality on approximately day 11 of development (Roebroek et al., 1998). In addition, we identified the activation of the gene insulin-like growth factor 1 receptor (*IGF1R*) at the blastocyst stage. The IGF1R signaling pathway regulates proliferation, differentiation, and survival of many cell types. Disturbances in the IGF1R signaling pathway induced by specific inhibitors in mouse embryos impaired trophectoderm formation (Bedzhov et al., 2012). The authors showed that the activation of IGF1R requires E-cadherin (CDH1) for proper maintenance of the trophoblast lineage, which is in concordance with our findings of *CDH1* being activated in parallel with *IGF1R* at the blastocyst stage. Gene *CCL17* (chemokine ligand 17), which is speculated to play a role in trophoblast invasion and migration in humans (Li et al., 2014), was found by using the primary transcripts as marker for embryonic expression. Within the genes first expressed after the 16-cell stage we identified *TFCP2L1* (CP2 family transcription factor). In mouse ES cells it was shown that TFCP2L1 is involved in their maintenance and self-renewal in a similar way as NANOG (Ye et al., 2013).

Relevance and outlook

In our study (Graf et al., 2014), we performed RNA-Seq analyses of pools of in vitro produced bovine embryos. These are known to be developmentally less competent than their in vivo derived counterparts (reviewed in Lonergan et al., 2008). Therefore it would be most interesting to repeat the RNA-Seq experiments with in vivo derived embryos, which might provide new molecular insights into the developmental differences of in vitro vs. in vivo derived embryos. Another question is how different culture conditions for embryos affect the timing of EGA and if there is an effect on developmental outcomes. An elegant study switching the environment of bovine embryos at defined stages from in vivo to in vitro and vice versa demonstrated that embryos are particularly sensitive to changes in culture condition around the time of EGA (Gad et al., 2012). Our approaches for detecting the activation of individual genes based on RNA-Seq data could clarify the question if EGA is in general affected by a change of the culture environment or if specific genes or groups of genes are particularly affected.

In our study, we analyzed three replicates of pools of 10 oocytes or embryos per developmental stage. This was mainly done since early embryos may suffer from a considerable proportion of cytogenetic abnormalities (Kawarsky et al., 1996; Demyda-Peyras et al., 2013) which may affect the gene expression profile. Nevertheless, RNA-Seq analysis of individual embryos (Chitwood et al., 2013) or even single embryonic cells (Tang et al., 2009; Xue et al., 2013) is technically feasible and could provide new insights into the inter-embryo variability of gene expression profiles and into the transcriptome changes guiding the first differentiation events in early embryos. RNA-Seq has already been applied for expression profiling of single cells derived from human and mouse embryos, revealing transcriptional modules that are activated during key steps of development (Xue et al., 2013; Yan et al., 2013). It would be interesting to see how conserved these are in bovine embryo and embryos from other livestock species.

Another important aspect is the study of allelic gene expression in early embryos. Although it is widely believed that transcription of autosomal genes occurs from both parental alleles, specific classes of genes have been shown to express only one, randomly selected allele (allelic exclusion) (reviewed in Deng et al., 2014). Furthermore, parental-specific (imprinted) expression has been demonstrated for about 1% of autosomal genes (reviewed in Reik et al., 2001). A recent RNA-Seq analysis of allelic expression in individual cells of mouse preimplantation embryos (CAST/EiJ × C57BL/6J) revealed 12–24% monoallelic expression of autosomal genes. The monoallelic expression appeared random and dynamic because there was considerable variation among closely related embryonic cells. This study provided evidence for a de novo inactivation of the paternal X chromosome starting at the four-cell stage (Deng et al., 2014). It will be interesting to see the dynamics of monoallelic gene expression in bovine embryos. In the Bos t. taurus × Bos t. indicus design of our study (Graf et al., 2014), informative parental SNPs were found in the coding regions of about 20% of the known bovine genes, supporting this type of analysis.

In summary, RNA-Seq analyses provided new insights into embryonic genome activation, allelic expression of genes, and transcriptional networks governing early development of embryos. RNA-Seq data provide a fundamental basis for embryo biosystems research, including genetic, epigenetic and environmental disturbances of early development.

Acknowledgments

Our studies of bovine embryonic genome activation were supported by the EU grants Plurisys (HEALTH-F4-2009-223485 FP7 Health 534 project) and Fecund (FP7-KBBE-2012-6 project 312097), by the Deutsche Forschungsgemeinschaft (FOR 1041), and by BioSysNet. MHB is supported by a DFG fellowship through QBM.

References

Abdelhaleem, M., Maltais, L., Wain, H. 2003. The human DDX and DHX gene families of putative RNA helicases. Genomics 81 (6), 618–622.

Adachi, K., Soeta-Saneyoshi, C., Sagara, H., Iwakura, Y. 2007. Crucial role of Bysl in mammalian preimplantation development as an integral factor for 40S ribosome biogenesis. Mol. Cell. Biol. 27 (6), 2202–2214.

Ameur, A., Zaghlool, A., Halvardson, J., Wetterbom, A., Gyllensten, U., Cavelier, L., Feuk, L. 2011. Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. Nat. Struct. Mol. Biol. 18 (12), 1435–1440.

Barnea, E., Bergman, Y. 2000. Synergy of SF1 and RAR in activation of Oct-3/4 promoter. J. Biol. Chem. 275 (9), 6608–6619.

Barnes, F. L., First, N. L. 1991. Embryonic transcription in in vitro cultured bovine embryos. Mol. Reprod. Dev. 29 (2), 117–123.

Bedzhov, I., Liszewska, E., Kanzler, B., Stemmler, M. P. 2012. Igf1r signaling is indispensable for preimplantation development and is activated via a novel function of E-cadherin. PLoS Genet. 8 (3), e1002609.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W. H., Pages, F., Trajanoski, Z., Galon, J. 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 25 (8), 1091–1093.

Braude, P., Bolton, V., Moore, S. 1988. Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. Nature 332 (6163), 459–461.

Camous, S., Kopecny, V., Flechon, J. E. 1986. Autoradiographic detection of the earliest stage of [3H]-uridine incorporation into the cow embryo. Biol. Cell 58 (3), 195–200.

Chitwood, J. L., Rincon, G., Kaiser, G. G., Medrano, J. F., Ross, P. J. 2013. RNA-seq analysis of single bovine blastocysts. BMC Genomics 14, 350.

Claveria, C., Giovinazzo, G., Sierra, R., Torres, M. 2013. Myc-driven endogenous cell competition in the early mammalian embryo. Nature 500 (7460), 39–44.

Dean, W. L., Kelsey, G., Reik, W. 2001. Generation of monoparental embryos for investigation into genomic imprinting. Methods Mol. Biol. 181, 1–19.

Demyda-Peyras, S., Dorado, J., Hidalgo, M., Anter, J., De Luca, L., Genero, E., Moreno-Millan, M. 2013. Effects of oocyte quality, incubation time and maturation environment on the number of chromosomal abnormalities in IVF-derived early bovine embryos. Reprod. Fertil. Dev. 25 (7), 1077–1084.

Deng, Q., Ramskold, D., Reinius, B., Sandberg, R. 2014. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science 343 (6167), 193–196.

Faast, R., Thonglairoam, V., Schulz, T. C., Beall, J., Wells, J. R., Taylor, H., Matthaei, K., Rathjen, P. D., Tremethick, D. J., Lyons, I. 2001. Histone variant H2A.Z is required for early mammalian development. Curr. Biol. 11 (15), 1183–1187.

Fang, X., Yoon, J. G., Li, L., Tsai, Y. S., Zheng, S., Hood, L., Goodlett, D. R., Foltz, G., Lin, B. 2011. Landscape of the SOX2 protein-protein interactome. Proteomics 11 (5), 921–934.

Frei, R. E., Schultz, G. A., Church, R. B. 1989. Qualitative and quantitative changes in protein synthesis occur at the 8-16-cell stage of embryogenesis in the cow. J. Reprod. Fertil. 86 (2), 637–641.

Gad, A., Hoelker, M., Besenfelder, U., Havlicek, V., Cinar, U., Rings, F., Held, E., Dufort, I., Sirard, M. A., Schellander, K., Tesfaye, D. 2012. Molecular mechanisms and pathways involved in bovine embryonic genome activation and their regulation by alternative in vivo and in vitro culture conditions. Biol. Reprod. 87 (4), 100.

Giraldez, A. J., Mishima, Y., Rihel, J., Grocock, R. J., Van Dongen, S., Inoue, K., Enright, A. J., Schier, A. F. 2006. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science 312 (5770), 75–79.

Glaser, S., Lubitz, S., Loveland, K. L., Ohbo, K., Robb, L., Schwenk, F., Seibler, J., Roellig, D., Kranz, A., Anastassiadis, K., Stewart, A. F. 2009. The histone 3 lysine 4 methyltransferase, Mll2, is only required briefly in development and spermatogenesis. Epigenetics Chromatin 2 (1), 5.

Glaser, S., Schaft, J., Lubitz, S., Vintersten, K., Hoeven, F. van der, Tufteland, K. R., Aasland, R., Anastassiadis, K., Ang, S. L., Stewart, A. F. 2006. Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. Development 133 (8), 1423–1432.

Gonzalez, B., Denzel, S., Mack, B., Conrad, M., Gires, O. 2009. EpCAM is involved in maintenance of the murine embryonic stem cell phenotype. Stem Cells 27 (8), 1782–1791.

Graf, A., Krebs, S., Zakhartchenko, V., Schwalb, B., Blum, H., Wolf, E. 2014. Fine mapping of genome activation in bovine embryos by RNA sequencing. Proc. Natl. Acad. Sci. U.S.A. 111 (11), 4139–4144.

Gray, R. S., Abitua, P. B., Wlodarczyk, B. J., Szabo-Rogers, H. L., Blanchard, O., Lee, I., Weiss, G. S., Liu, K. J., Marcotte, E. M., Wallingford, J. B., Finnell, R. H. 2009. The planar cell polarity effector Fuz is essential for targeted membrane trafficking, ciliogenesis and mouse embryonic development. Nat. Cell Biol. 11 (10), 1225–1232.

Guo, G., Huss, M., Tong, G. Q., Wang, C., Li Sun, L., Clarke, N. D., Robson, P. 2010. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. Dev. Cell 18 (4), 675–685.

Hinnebusch, A. G. 2006. eIF3: a versatile scaffold for translation initiation complexes. Trends Biochem. Sci. 31 (10), 553–562.

Home, P., Ray, S., Dutta, D., Bronshteyn, I., Larson, M., Paul, S. 2009. GATA3 is selectively expressed in the trophectoderm of peri-implantation embryo and directly regulates Cdx2 gene expression. J. Biol. Chem. 284 (42), 28729–28737.

Hu, G., Cui, K., Northrup, D., Liu, C., Wang, C., Tang, Q., Ge, K., Levens, D., Crane-Robinson, C., Zhao, K. 2013. H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. Cell Stem Cell 12 (2), 180–192.

Huang, W., Khatib, H. 2010. Comparison of transcriptomic landscapes of bovine embryos using RNA-Seq. BMC Genomics 11, 711.

Hyttel, P., Viuff, D., Avery, B., Laurincik, J., Greve, T. 1996. Transcription and cell cycle-dependent development of intranuclear bodies and granules in two-cell bovine embryos. J. Reprod. Fertil. 108 (2), 263–270.

Jonckheere, A. I., Smeitink, J. A., Rodenburg, R. J. 2012. Mitochondrial ATP synthase: architecture, function and pathology. J. Inherit. Metab. Dis. 35 (2), 211–225.

Kanka, J. 2003. Gene expression and chromatin structure in the pre-implantation embryo. Theriogenology 59 (1), 3–19.

Kanka, J., Nemcova, L., Toralova, T., Vodickova-Kepkova, K., Vodicka, P., Jeseta, M., Machatkova, M. 2012. Association of the transcription profile of bovine oocytes and embryos with developmental potential. Anim. Reprod. Sci. 134 (1-2), 29–35.

Karantzali, E., Lekakis, V., Ioannou, M., Hadjimichael, C., Papamatheakis, J., Kretsovali, A. 2011. Sall1 regulates embryonic stem cell differentiation in association with nanog. J. Biol. Chem. 286 (2), 1037–1045.

Katz, J. P., Perreault, N., Goldstein, B. G., Lee, C. S., Labosky, P. A., Yang, V. W., Kaestner, K. H. 2002. The zinc-finger transcription factor Klf4 is required for terminal differentiation of goblet cells in the colon. Development 129 (11), 2619–2628.

Kawarsky, S. J., Basrur, P. K., Stubbings, R. B., Hansen, P. J., King, W. A. 1996. Chromosomal abnormalities in bovine embryos and their influence on development. Biol. Reprod. 54 (1), 53–59.

Kent, L. N., Rumi, M. A., Kubota, K., Lee, D. S., Soares, M. J. 2011. FOSL1 is integral to establishing the maternal-fetal interface. Mol. Cell. Biol. 31 (23), 4801–4813.

Khan, D. R., Dube, D., Gall, L., Peynot, N., Ruffini, S., Laffont, L., Le Bourhis, D., Degrelle, S., Jouneau, A., Duranthon, V. 2012. Expression of pluripotency master regulators during two key developmental transitions: EGA and early lineage specification in the bovine embryo. PLoS ONE 7 (3), e34110.

Kidder, B. L., Palmer, S. 2010. Examination of transcriptional networks reveals an important role for TCFAP2C, SMARCA4, and EOMES in trophoblast stem cell maintenance. Genome Res. 20 (4), 458–472.

Kohl, M., Wiese, S., Warscheid, B. 2011. Cytoscape: software for visualization and analysis of biological networks. Methods Mol. Biol. 696, 291–303.

Kotkamp, K., Mossner, R., Allen, A., Onichtchouk, D., Driever, W. 2014. A Pou5f1/Oct4 dependent Klf2a, Klf2b, and Klf17 regulatory sub-network contributes to EVL and ectoderm development during zebrafish embryogenesis. Dev. Biol. 385 (2), 433–447.

Kues, W. A., Sudheer, S., Herrmann, D., Carnwath, J. W., Havlicek, V., Besenfelder, U., Lehrach, H., Adjaye, J., Niemann, H. 2008. Genome-wide expression profiling reveals distinct clusters of transcriptional regulation during bovine preimplantation development in vivo. Proc. Natl. Acad. Sci. U.S.A. 105 (50), 19768–19773.

Lee, J., Zhou, P. 2007. DCAFs, the missing link of the CUL4-DDB1 ubiquitin ligase. Mol. Cell 26 (6), 775–780.

Lepikhov, K., Zakhartchenko, V., Hao, R., Yang, F., Wrenzycki, C., Niemann, H., Wolf, E., Walter, J. 2008. Evidence for conserved DNA and histone H3 methylation reprogramming in mouse, bovine and rabbit zygotes. Epigenetics Chromatin 1 (1), 8.

Li, C. M., Hou, L., Zhang, H., Zhang, W. Y. 2014. CCL17 Induces Trophoblast Migration and Invasion by Regulating Matrix Metalloproteinase and Integrin Expression in Human First-Trimester Placenta. Reprod Sci.

Li, L., Zheng, P., Dean, J. 2010. Maternal control of early mouse development. Development 137 (6), 859–870.

Lonergan, P., Fair, T. 2008. In vitro-produced bovine embryos: dealing with the warts. Theriogenology 69 (1), 17–22.

Lu, R., Yang, A., Jin, Y. 2011. Dual functions of T-box 3 (Tbx3) in the control of self-renewal and extraembryonic endoderm differentiation in mouse embryonic stem cells. J. Biol. Chem. 286 (10), 8425–8436.

Ma, P., Lin, S., Bartolomei, M. S., Schultz, R. M. 2010. Metastasis tumor antigen 2 (MTA2) is involved in proper imprinted expression of H19 and Peg3 during mouse preimplantation development. Biol. Reprod. 83 (6), 1027–1035.

Marcucio, R. S., Hopwood, R. M., Ignotz, G. G., Currie, W. B. 1995. Translation of zygotically-derived mRNA in a cell cycle specific manner in 2-cell cattle embryos. J. Reprod. Fert. Abstr. Ser. 15, (abstract 46).

Marfella, C. G., Ohkawa, Y., Coles, A. H., Garlick, D. S., Jones, S. N., Imbalzano, A. N. 2006. Mutation of the SNF2 family member Chd2 affects mouse development and survival. J. Cell. Physiol. 209 (1), 162–171.

Memili, E., Dominko, T., First, N. L. 1998. Onset of transcription in bovine oocytes and preimplantation embryos. Mol. Reprod. Dev. 51 (1), 36–41.

Memili, E., First, N. L. 1999. Control of gene expression at the onset of bovine embryonic development. Biol. Reprod. 61 (5), 1198–1207.

Misirlioglu, M., Page, G. P., Sagirkaya, H., Kaya, A., Parrish, J. J., First, N. L., Memili, E. 2006. Dynamics of global transcriptome in bovine matured oocytes and preimplantation embryos. Proc. Natl. Acad. Sci. U.S.A. 103 (50), 18905–18910.

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 (5), 631–642.

Mondou, E., Dufort, I., Gohin, M., Fournier, E., Sirard, M. A. 2012. Analysis of microRNAs and their precursors in bovine early embryonic development. Mol. Hum. Reprod. 18 (9), 425–434.

Nagao, K., Zhu, J., Heneghan, M. B., Hanson, J. C., Morasso, M. I., Tessarollo, L., Mackem, S., Udey, M. C. 2009. Abnormal placental development and early embryonic lethality in EpCAM-null mice. PLoS ONE 4 (12), e8543.

Natale, D. R., Kidder, G. M., Westhusin, M. E., Watson, A. J. 2000. Assessment by differential display-RT-PCR of mRNA transcript transitions and alpha-amanitin sensitivity during bovine preattachment development. Mol. Reprod. Dev. 55 (2), 152–163.

Nishioka, N., Yamamoto, S., Kiyonari, H., Sato, H., Sawada, A., Ota, M., Nakao, K., Sasaki, H. 2008. Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. Mech. Dev. 125 (3-4), 270–283.

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 (5), 917–929.

Okano, M., Bell, D. W., Haber, D. A., Li, E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99 (3), 247–257.

Østrup, O., Andersen, I. S., Collas, P. 2013. Chromatin-linked determinants of zygotic genome activation. Cell. Mol. Life Sci. 70 (8), 1425–1437.

Pelley, J. W., Goljan, E. F. 2011. Nucleotide synthesis and metabolism. Rapid Review Biochemistry Elsevier.

Plante, L., Plante, C., Shepherd, D. L., King, W. A. 1994. Cleavage and 3H-uridine incorporation in bovine embryos of high in vitro developmental potential. Mol. Reprod. Dev. 39 (4), 375–383.

Ralston, A., Cox, B. J., Nishioka, N., Sasaki, H., Chea, E., Rugg-Gunn, P., Guo, G., Robson, P., Draper, J. S., Rossant, J. 2010. Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. Development 137 (3), 395–403.

Ramachandra, R. K., Salem, M., Gahr, S., Rexroad, C. E., Yao, J. 2008. Cloning and characterization of microRNAs from rainbow trout (Oncorhynchus mykiss): their expression during early embryonic development. BMC Dev. Biol. 8, 41.

Reik, W., Walter, J. 2001. Genomic imprinting: parental influence on the genome. Nat. Rev. Genet. 2 (1), 21–32.

Roebroek, A. J., Umans, L., Pauli, I. G., Robertson, E. J., Leuven, F. van, Van de Ven, W. J., Constam, D. B. 1998. Failure of ventral closure and axial rotation in embryos lacking the proprotein convertase Furin. Development 125 (24), 4863–4876.

Ruland, J., Sirard, C., Elia, A., MacPherson, D., Wakeham, A., Li, L., Pompa, J. L. de la, Cohen, S. N., Mak, T. W. 2001. p53 accumulation, defective cell proliferation, and early embryonic lethality in mice lacking tsg101. Proc. Natl. Acad. Sci. U.S.A. 98 (4), 1859–1864.

Santos, F., Zakhartchenko, V., Stojkovic, M., Peters, A., Jenuwein, T., Wolf, E., Reik, W., Dean, W. 2003. Epigenetic marking correlates with developmental potential in cloned bovine preimplantation embryos. Curr. Biol. 13 (13), 1116–1121.

Sirard, M. A. 2010. Activation of the embryonic genome. Soc Reprod Fertil Suppl 67, 145–158.

Sirard, M. A. 2012. Factors affecting oocyte and embryo transcriptomes. Reprod. Domest. Anim. 47 Suppl 4, 148–155.

Stein, P. L., Vogel, H., Soriano, P. 1994. Combined deficiencies of Src, Fyn, and Yes tyrosine kinases in mutant mice. Genes Dev. 8 (17), 1999–2007.

Svarcova, O., Laurincik, J., Avery, B., Mlyncek, M., Niemann, H., Maddox-Hyttel, P. 2007. Nucleolar development and allocation of key nucleolar proteins require de novo transcription in bovine embryos. Mol. Reprod. Dev. 74 (11), 1428–1435.

Swinburne, I. A., Silver, P. A. 2008. Intron delays and transcriptional timing during development. Dev. Cell 14 (3), 324–330.

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

Tamm, C., Bower, N., Anneren, C. 2011. Regulation of mouse embryonic stem cell self-renewal by a Yes-YAP-TEAD2 signaling pathway downstream of LIF. J. Cell. Sci. 124 (Pt 7), 1136–1144.

Tan, M. H., Au, K. F., Leong, D. E., Foygel, K., Wong, W. H., Yao, M. W. 2013. An Oct4-Sall4-Nanog network controls developmental progression in the pre-implantation mouse embryo. Mol. Syst. Biol. 9, 632.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K., Surani, M. A. 2009. mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods 6 (5), 377–382.

Telford, N. A., Watson, A. J., Schultz, G. A. 1990. Transition from maternal to embryonic control in early mammalian development: a comparison of several species. Mol. Reprod. Dev. 26 (1), 90–100.

Toralová, T., Benešová, V., Kepková, K. V., Vodička, P., Šušor, A., Kaňka, J. 2012. Bovine preimplantation embryos with silenced nucleophosmin mRNA are able to develop until the blastocyst stage. Reproduction 144 (3), 349–359.

Tripurani, S. K., Wee, G., Lee, K. B., Smith, G. W., Wang, L., Jianboyao 2013. MicroRNA-212 post-transcriptionally regulates oocyte-specific basic-helix-loop-helix transcription factor, factor in the germline alpha (FIGLA), during bovine early embryogenesis. PLOS ONE 8, e761148.

Vallee, M., Dufort, I., Desrosiers, S., Labbe, A., Gravel, C., Gilbert, I., Robert, C., Sirard, M. A. 2009. Revealing the bovine embryo transcript profiles during early in vivo embryonic development. Reproduction 138 (1), 95–105.

Vigneault, C., Gravel, C., Vallee, M., McGraw, S., Sirard, M. A. 2009. Unveiling the bovine embryo transcriptome during the maternal-to-embryonic transition. Reproduction 137 (2), 245–257.

Viuff, D., Avery, B., Greve, T., King, W. A., Hyttel, P. 1996. Transcriptional activity in in vitro produced bovine two- and four-cell embryos. Mol. Reprod. Dev. 43 (2), 171–179.

Viuff, D., Hyttel, P., Avery, B., Vajta, G., Greve, T., Callesen, H., Thomsen, P. D. 1998. Ribosomal ribonucleic acid is transcribed at the 4-cell stage in in vitro-produced bovine embryos. Biol. Reprod. 59 (3), 626–631.

Vliet, J. van, Crofts, L. A., Quinlan, K. G., Czolij, R., Perkins, A. C., Crossley, M. 2006. Human KLF17 is a new member of the Sp/KLF family of transcription factors. Genomics 87 (4), 474–482.

Walser, C. B., Lipshitz, H. D. 2011. Transcript clearance during the maternal-to-zygotic transition. Curr. Opin. Genet. Dev. 21 (4), 431–443.

Wang, H., Dey, S. K. 2006. Roadmap to embryo implantation: clues from mouse models. Nat. Rev. Genet. 7 (3), 185–199.

Wang, Z., Gerstein, M., Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10 (1), 57–63.

Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C. J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., Walter, J. 2011. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun 2, 241.

Wrenzycki, C., Herrmann, D., Lucas-Hahn, A., Korsawe, K., Lemme, E., Niemann, H. 2005. Messenger RNA expression patterns in bovine embryos derived from in vitro procedures and their implications for development. Reprod. Fertil. Dev. 17 (1-2), 23–35.

Wu, M. Y., Ramel, M. C., Howell, M., Hill, C. S. 2011. SNW1 is a critical regulator of spatial BMP activity, neural plate border formation, and neural crest specification in vertebrate embryos. PLoS Biol. 9 (2), e1000593.

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

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., Huang, J., Li, M., Wu, X., Wen, L., Lao, K., Li, R., Qiao, J., Tang, F. 2013. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat. Struct. Mol. Biol. 20 (9), 1131–1139.

Yan, W., Burns, K. H., Ma, L., Matzuk, M. M. 2002. Identification of Zfp393, a germ cell-specific gene encoding a novel zinc finger protein. Mech. Dev. 118 (1-2), 233–239.

Ye, S., Li, P., Tong, C., Ying, Q. L. 2013. Embryonic stem cell self-renewal pathways converge on the transcription factor Tfcp2l1. EMBO J. 32 (19), 2548–2560.

Zhang, J., Tam, W. L., Tong, G. Q., Wu, Q., Chan, H. Y., Soh, B. S., Lou, Y., Yang, J., Ma, Y., Chai, L., Ng, H. H., Lufkin, T., Robson, P., Lim, B. 2006. Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. Nat. Cell Biol. 8 (10), 1114–1123.

Zhang, P., Andrianakos, R., Yang, Y., Liu, C., Lu, W. 2010. Kruppel-like factor 4 (Klf4) prevents embryonic stem (ES) cell differentiation by regulating Nanog gene expression. J. Biol. Chem. 285 (12), 9180–9189.

Zhou, Z., Zhang, Q., Lu, X., Wang, R., Wang, H., Wang, Y. L., Zhu, C., Lin, H. Y., Wang, H. 2013. The proprotein convertase furin is required for trophoblast syncytialization. Cell Death Dis 4, e593.

3 Discussion

Aim of this work was to study genome activation in early bovine embryos by bioinformatic evaluation of transcriptome profiles. These profiles had been generated by whole transcriptome shotgun sequencing (RNA-Seq) of oocytes at the GV and MII stage and of embryos at the four-cell, eight-cell, 16-cell and blastocyst stage. *Bos t. taurus* oocytes fertilized with sperms from a single *Bos t. indicus* sire had been used to construct sequencing libraries without polyA+ selection or rRNA depletion. These transcriptome profiles were evaluated by tailored bioinformatic approaches in order to gain new insights in genome activation in early bovine embryos.

Alignment of RNA-Seq data

The basic step in evaluation of transcriptome profiles is to align the sequence reads against the reference genome. A high quality mapping is a precondition for the significance of subsequent bioinformatic evaluation. For this thesis, a number of powerful software packages were available to handle RNA-Seq data, like *GNSAP* (Wu et al., 2010), *STAR* (Dobin et al., 2013), *MapSplice* (Wang et al., 2010), *PALMapper* (Jean et al., 2010), *ReadsMap* (Seledtsov et al., 2012), *JAGuaR* (Butterfield et al., 2014), *TopHat* (Trapnell et al., 2009) and *TopHat2* (Kim et al., 2013). A comparison of these programs with simulated and real RNA-Seq data sets revealed that *GSNAP*, *MapSplice* and *STAR* outperform the other alignment programs in terms of base wise accuracy (Engström et al., 2013). But in detail, *GSNAP* and *STAR* showed a significant weakness because they spliced reads over several false exon junctions. *TopHat2* performed well compared to other aligners only when a gene annotation was provided because it maps reads directly to the annotated transcripts. This makes it only suitable for well annotated genomes (Engström et al., 2013).

A further aspect is the tolerance for mismatches, required for the detection of parternal SNPs that were used as marker for the detection of EGA. Programs like *GSNAP*, *STAR* and *MapSplice* reported less mismatches compared to *TopHat2*. These tools truncate reads at the end, where multiple SNPs would be introduced in the alignment and thus they were not able to map the entire read (Engström et al., 2013). Therefore, those alignment programs seem not to be suitable for the detection of parent-specific variants in RNA-Seq data. In this thesis *Tophat2* was used because it maps reads spanning exon-exon boundaries correctly while it tolerates mismatches. Thus, *Tophat2* allowed not only to study differential abundance of transcripts but also to detect paternal-specific SNPs in transcriptome profiles. In general, the correct mapping is impaired by reads mapping to repetitive regions and therefore a common practice of alignment algorithms is to discard multi-mapped reads or to place them heuristically to one location (Li et al., 2010). Using *Tophat2*, the number of uniquely mapped reads decreased from 74% of sequence reads derived from GV oocytes to 50% in those from blastocysts. Simultaneously the percentage of reads mapped to intronic regions increased from below 10% in the oocyte stages to about 30% in the blastocysts. These results indicated extensive changes in the transcriptome during early embryonic development. The almost completely spliced RNA of GV oocytes was replaced by *de novo* transcribed RNA species, which in their nascent form are still incompletely spliced. For the first time, the capture of intronic sequences paved the way for a sensitive analysis of *de novo* transcribed genes. In addition, these sequences allow the discrimination between maternally delivered spliced RNA species and nascent ones, generated by the activation of the embryonic genome.

The percentage of uniquely mapped reads and the fraction of intronic mapped sequences (e.g. 30% for blastocysts) were in the same range as found by others in bovine embryos. Transcriptome profiles generated by RNA-Seq from pools of normal and degenerated bovine IVF blastocysts resulted in about 50% uniquely aligned reads. Furthermore, 17% of the reads were mapped to intronic sequences in normal blastocysts and 34% in the degenerated ones (Huang et al., 2010). RNA-Seq data of single bovine blastocysts contained also about 20% of reads mapped to intronic regions (Chitwood et al., 2013).

Comparative analysis of transcriptome profiles

A commonly used strategy for the evaluation of RNA-Seq data sets is the comparative analysis of transcriptome profiles. This approach captures all transcribed genes and detects differentially expressed genes (DEG). Based on this approach, transcripts of about 13,000 genes were detected in GV and MII oocytes and in all investigated embryonic stages. The number of DEG increased during embryonic development from 120 between GV and MII oocytes to approximately 3,000 between 16-cell embryos and blastocysts. The number of detected genes showed that the transcriptome of oocytes contained a large number of RNA species, which was stepwise altered during embryonic development. The increasing number of DEG during early embryonic development correlated inversely with the percentage of uniquely mapped reads, together indicating ongoing alterations of the transcriptome. A RNA-Seq study compared pools of normal bovine IVF blastocysts with degenerated ones (Huang et al., 2010). In their study, they could identify approximately 16,000 genes as transcribed. They defined, that a gene was considered to be expressed if a single read was mapped to the coding region. Without using any biological replicates, they reported about 20% more transcribed genes in the study as presented in this thesis, but it is statistically questionable to take a single read in a single replicate as evidence for the presence of a transcribed gene.

The ability of RNA-Seq to identify transcribed genes and to measure the abundance of certain transcripts relies on the number of reads that were sequenced and mapped to the correct position. Therefore sequencing depth is a limiting factor for the detection of transcribed genes and the identification of DEGs. In order to obtain an accurate quantification of expressed transcripts in the size of a mammalian genome, it was shown that about 700 million reads would be required to cover about 95% of expressed transcripts (Blencowe et al., 2009). The comparison of three human RNA-Seq data sets could show that a saturation of identified transcripts according to read depth was not reached in any of these experiments (Tarazona et al., 2011). Furthermore, a recent comprehensive study in yeast (Schulz et al., 2013) showed that together with transcription of open reading frames a broad diversity of short intermediate and antisense transcripts are initiated, so that a true saturation of transcriptome analysis may not be reachable at all. In this thesis – due to the high costs of RNA-Seq – about 9–58 million reads were produced for each biological replicate. This was sufficient to detect approximately 13,000 transcripts in each investigated sample. Under consideration of the decreasing percentage of uniquely mapped genes during the course of early embryonic

development more sequencing reads would enhance sensitivity toward the capture of low expressed genes and could result in a higher number of identified transcribed genes.

Several other comparable studies of gene expression and genome activation in bovine embryos were performed with microarrays. Misirlioglu et al. (2006) compared transcriptome profiles of eight-cell embryos treated with or without α-amanitin. They identified about 12,000 transcribed genes in untreated eight-cell embryos and 233 DEG in response to α-amanitin treatment. The number of DEG was in the same range as found in a later study by Vigneault et al. (2009). Their array analysis revealed 310 DEG in untreated eight-cell compared to eight-cell embryos treated with α-amanitin. Both microarray based analyses resulted in significantly lower numbers of DEG than found by the RNA-Seq approach of this thesis. However, microarrays may provide comprehensive transcriptome profiles and enable holistic comparative analysis of transcript abundances (Marioni et al., 2008). In contrast to RNA-Seq, the sensitivity for the detection of lowly abundant transcripts is limited by unspecific hybridization (Marcelino et al., 2006) and causing a high noise level. Thus, the lower number of DEG found by Misirlioglu et al. (2006) might be caused by the intrinsic properties of microarrays.

Detection of actively transcribed genes in early bovine embryos

The most elementary approach to detect *de novo* transcription was to look for transcripts that were completely absent in MII and GV oocytes. In total 220 genes could be identified: eight genes in four-cell embryos, 129 genes in eight-cell embryos, 36 genes in 16-cell embryos and 47 genes in blastocysts. Compared to the 13,000 RNA species found to be present in MII and GV oocytes, this number was relatively low. Moreover the increasing numbers of DEG found by comparative analysis of the transcriptome profiles indicated that at least a part of the 13,000 RNA species was actively transcribed. A first hint how to detect *de novo* transcription among this RNA pool was provided by the finding, that during the course of embryonic development the number of uniquely mapped reads decreased with a simultaneously increase of intronic mapped reads. A closer look at the mapped sequencing data revealed that the RNA species found in MII and GV oocytes were completely spliced, whereas in all embryonic stages immature unspliced RNA species were present. This observation was enabled by the fact, that all sequencing libraries were constructed with total RNA without prior enrichment of matured polyadenylated RNA species. Therefore nascent primary transcripts could be

captured as well as incompletely spliced and mature mRNA. In order to distinguish between actively transcribed genes and transcriptional inactive ones, a tailored bioinformatic approach was developed. This approach calculated the **r**atio of **i**ntronic read counts to **n**ot-covered intronic **p**ositions (RINP) for each gene. The method was able to identify actively transcribed introns even if they contain spikes, small regions with unusually high read coverage. These spikes were found both within introns that were uniformly covered, as well as in introns that showed no read coverage outside the spikes. They might be caused by antisense transcripts or residual genomic DNA. In total, the onset of transcription of 6,848 genes could be detected by the presence of transcribed introns from the four-cell to the blastocyst stage. Nearly 60% of these genes were actively transcribed for the first time in the eight-cell stage. The capture of incompletely spliced transcripts enabled us to determine the onset of transcription of genes even if the RNA species were maternally delivered and present in high abundance. A global view on the onset of transcription of all annotated genes showed that activation of genes during the early embryonic development was size dependent. The smaller genes were preferentially activated earlier than the larger ones. This might be caused by less efficient splicing of larger transcripts, leading to accumulation of incompletely spliced RNA species. Alternatively, short cell cycles in the earlier stages might interfere with complete transcription of the longer genes, so that their activation is detected only in the later stages, where the cell cycle is getting long enough.

Parent-specific transcription

Another unique feature of high throughput sequencing is that it provides the correct nucleotide sequence of each transcribed gene. The sequencing libraries were constructed starting with *Bos t. taurus* × *Bos t. indicus* crossbreed embryos fertilized with sperms from a single *Bos t. indicus* sire. Due to the single nucleotide resolution of RNA-Seq, breed-specific single nucleotide polymorphisms (SNPs) could be used to distinguish whether the maternal *Bos t. taurus* allele or the paternal *Bos t. indicus* allele was transcribed. In this thesis, 61,000 *Bos t. indicus* specific SNPs were identified, covering approximately 20% of all annotated genes. Based on these SNPs, the first active transcription of paternal alleles could be determined in 937 genes. Nearly 400 genes were found for the first time in the eight-cell stage. The identification of SNPs from mapped reads was impaired by several aspects. Therefore, tailored optimized software would be needed. Often the alignment algorithms incorrectly align the reads and introduce false

positive SNPs. Misalignments were discovered in highly repetitive regions. Another source for false positive SNPs is found by a wrong alignment of reads spanning the splice site or by sequencing errors generated by the sequencer (Piskol et al., 2013). The per-base quality scores might vary due to the technical properties of the sequencer and the sequence context might lead to incorrect quality scores and thus might impair identification of SNPs. Furthermore, current alignment tools map each single sequence read to the reference genome independently and therefore especially sequence reads spanning indels were mostly misaligned (DePristo et al., 2011). All these aspects might result in the identification of false positive SNPs and lower the significance of results. An improvement was recently suggested by the *Genome Analysis Toolkit (GATK)* (DePristo et al., 2011). *GATK* locally corrects the alignments around the detected indels and the base quality scores were recalibrated to obtain variant calls with high quality. Besides the suggested improvements of the software, the evaluation of the sequencing data of the *Bos t. taurus* × *Bos t. indicus* crossbreed could be improved by using the data generated by whole genome sequencing of 43 Simmental animals. More than 91,000 SNPs were detected in approximately 18,444 genes (Jansen et al., 2013). Additionally, the currently 28 million variants from 234 different cattle breeds identified in the 1000 bull genome project (Daetwyler et al., 2014) could provide informative nucleotide variants for additional detailed studies.

Besides the identification of the onset of transcription, the detection of parental-specific transcripts might be used for detailed studies of allele specific transcription. It has recently been published, that allele-specific silencing is a prerequisite for proper development of early embryos (Miyanari et al., 2012). Thus, the analysis of allele specific transcription might pave the way for detailed insights into the processes regulating the development of early bovine embryos.

Study of genome activation in early bovine embryos

An important phase in the development of early embryos is the activation of the embryonic genome. During this phase the maternally provided RNA species are gradually replaced by *de novo* transcribed embryonic ones. Transcripts of approximately 13,000 genes could be detected, indicating the experimental setup was able to capture a high percentage of the transcriptome. Probably the percentage could be enhanced by an increase of sequencing depth, especially for very low abundant transcripts. Differentially abundant transcripts can only reflect net changes of the transcriptome during embryonic

development. The pure number of DEG significantly increased at the eight-cell stage, indicating the onset of major genome activation at that stage. Nevertheless, this kind of evaluation delivered only limited information because the huge amount of maternally delivered transcripts masks the small changes in transcript abundance caused by EGA.

This limitation could be resolved by analysis of transcription of intronic regions, enabling high sensitive detection of *de novo* transcribed RNA species. Summed up over all embryonic stages de novo transcription was found in nearly 8,000 genes or 60% of all captured genes. The majority of *de novo* transcribed genes was found at the eight-cell stage pointing to the onset of major genome activation at that stage. Analysis of the functional GO terms of these genes resulted, exemplarily in "RNA splicing", "mRNA transcription from RNA polymerase II promotor" and "regulation of transcription from RNA polymerase II promotor". This demonstrated the ability of RNA-Seq to provide transcriptome-wide insight in biological processes occurring during major genome activation. The onset of major genome activation was confirmed by a third approach that analyzed the occurrence of paternal-specific SNPs in actively transcribed genes. Oocytes contain a huge amount of maternally provided RNA, but the spermatozoid carries only neglectable amounts of mRNA. Therefore, the presence of RNA species with paternal SNPs can be associated with *de novo* synthesis from the embryonic genome. This approach identified 40% of all paternal SNP containing transcripts at the eight-cell stage. This result demonstrates, that RNA-Seq of crossbreed embryos enables to unveil fine-tuning of regulation of transcription by detection of allele specific gene expression during major genome activation. First insights in bovine EGA were previously obtained by the evaluation of datasets generated with microarrays (Misirlioglu et al., 2006; Kues et al., 2008; Vigneault et al., 2009), which enable holistic analysis of the transcriptome. These studies were able to identify the major onset of bovine EGA at the eight-cell stage and identified some hundreds of involved genes. In contrast to the published findings, in this thesis a much higher number of genes involved in major genome activation could be identified, demonstrating the superiority of RNA-Seq based analyses.

In summary the bioinformatic analyses performed in this thesis confirmed the onset of major genome activation at the eight-cell stage of bovine embryos and proved, that the majority of the transcriptome is affected. Furthermore, a basis for studying major genome activation with an unprecedented resolution was generated.

Conclusion

The comprehensive bioinformatic analysis of RNA-Seq data derived from bovine GV and MII oocytes and of embryos at the four-cell, eight-cell and 16-cell and blastocysts stage could provide detailed insights into transcriptome changes during early embryonic development. Differentially abundant transcripts were captured by comparative analysis of transcriptome profiles and actively transcribed genes were identified based on transcribed intronic regions. Furthermore, expression of paternal alleles was detected with the occurrence of *Bos t. indicus* specific SNPs. This was accomplished using crossbreed *Bos t. taurus* × *Bos t. indicus* embryos and by developing tailored bioinformatic methods. All results proved that the major genome activation occurs at the eight-cell stage and affects approximately 4,000 genes. According to functional GO analyses, these genes are involved in transcriptional and translational processes and their regulation. This study opens new possibilities for embryo biosystems research and for detecting genetic, epigenetic and environmental disturbances in early embryo development.

4 Summary

In this thesis, comparative analyses of RNA-seq data set were performed. The sequencing libraries had been constructed starting with germinal vesicle (GV) and metaphase II (MII) oocytes and embryos at the four-cell, eight-cell, 16-cell and blastocyst stage. The embryos had been generated *in vitro* by fertilization of *Bos taurus taurus* oocytes with sperm of a *Bos taurus indicus* sire. In total, approximately 13,000 RNA species could be identified in oocytes and each embryonic stages. The number of identified differential abundant transcripts increased in the course of development from roughly 100 to several thousands, with a sharp rise at the eight-cell stage.

A bioinformatic approach could be developed to capture maternally delivered and *de novo* synthesized RNA species separately. It sensitively identified actively transcribed genes despite the fact that comparative analyses failed due to presence of the huge amount of RNA provided by the oocyte. Actively transcribed RNA species could be identified for approximately 8,000 genes, the majority of them at the eight-cell stage. This finding indicated, that the majority of all RNA species provided by oocytes was *de novo* transcribed during early embryonic development. Furthermore, it could be shown that the *de novo* transcription of larger genes was initiated later in embryonic development than smaller ones.

A procedure was established to identify *Bos t. indicus* specific SNPs in RNA-Seq datasets. It identified more than 60,000 SNPs occurring in 20% of all annotated genes. A major part of these SNPs could be detected at the eight-cell stage. This procedure enables a way to capture and study allele-specific transcription during early embryonic development.

The described bioinformatic approaches were used to study major genome activation, an important step in the maternal-to-embryonic transition. More than 4,000 genes were *de novo* transcribed during major genome activation, which was found to occur at the

eight-cell stage. These genes were functionally related to transcription, translation and their regulation.

In summary, this thesis created and applied a powerful tool set for bioinformatic dissection of processes occurring during development of early bovine embryos and provided unprecedented insights in major genome activation.

5 Zusammenfassung

In der vorliegenden Doktorarbeit wurden vergleichende Analysen an RNA-Seq Datensätzen durchgeführt. Die zur Sequenzierung verwendeten Genbanken sind ausgehend von unreifen Oozyten und MII Oozyten, sowie von Embryonen im Vierzell-, Achtzell-, 16-Zell- und Blastozysten-Stadium erstellt worden. Die Embryonen waren *in vitro* durch Befruchtung von *Bos taurus taurus* Oozyten mit Sperma eines *Bos taurus indicus* Bullen erzeugt worden. In jedem Ansatz konnten RNA-Spezies von etwa 13.000 Genen nachgewiesen werden. Die Anzahl der differentiell abundanten Transkripte stieg ausgehend von einigen wenigen in den Oozyten im Laufe der Entwicklung generell an und wies im Achtzell-Stadium einen sprunghaften Anstieg aus.

Es gelang, ein bioinformatisches Verfahren zu entwickeln, das die maternalen RNA-Spezies der Oozyten und die *de novo* transkribierten RNA-Spezies getrennt erfassen konnte. Dieses Verfahren identifizierte aktiv transkribierte Gene mit hoher Nachweisempfindlichkeit, obwohl diese wegen der großen Menge an vorhandener maternaler RNA nicht durch vergleichende Analysen der Genexpression erfasst werden konnten. Insgesamt konnten 8.000 *de novo* transkribierte RNA-Spezies nachgewiesen werden, wobei der größte Anteil im Achtzell-Stadium erfasst wurde. Dieses Ergebnis zeigte, dass der überwiegende Anteil aller von Oozyten stammenden RNA-Spezies im Laufe der Embryonalentwicklung neu transkribiert wird. Außerdem konnte gezeigt werden, dass die *de novo* Transkription von längeren Genen in späteren Entwicklungsstadien angestoßen wird als die von kürzeren Genen.

Ein Verfahren zum Nachweis von *Bos t. indicus* spezifischen SNPs in RNA-Seq Datensätzen konnte entwickelt werden. Dieses identifizierte etwa 60.000 *Bos t. indicus* spezifische SNPs in allen RNA-Seq Datensätzen, die in etwa 20% aller annotierten Gene vorkommen. Ein großer Teil der SNPs konnte im Achtzell-Stadium nachgewiesen werden. Dieses Verfahren ermöglicht somit die Erfassung und Untersuchung allel-spezifischer Transkription während der frühen Embryonalentwicklung.

Die bislang beschriebenen bioinformatischen Verfahren wurden zur Untersuchung der großen Genomaktivierung eingesetzt, einem wichtigen Vorgang während der maternal-embryonalen Übergangsphase. Insgesamt wurden etwa 4.000 Gene während der großen Genomaktivierung *de novo* transkribiert, die im Achtzell-Stadium stattfindet. Diese Gene sind funktionell den Ontologien Transkription und Translation sowie der Regulation zugeordnet.

Zusammenfassend wurden in der vorliegenden Arbeit leistungsfähige bioinformatische Verfahren entwickelt um die Vorgänge während der frühen Entwicklung von Rinderembyonen mit hoher Auflösung zu untersuchen. Dies ermöglichte neue Einblicke in die große Genomaktivierung.

Appendix

Dataset 1

Differentially expressed genes between subsequent developmental stages of GV oocytes, MII oocytes, four-cell, eight-cell, 16-cell and blastocysts were identified with *DESeq* and filtered with an adjusted P value < 0.05. The *baseMean* corresponds to the averaged normalized read count over all compared samples. The *group baseMean* represents the mean normalized read counts of the replicates in each group. The P values were adjusted with the Benjamini-Hochberg procedure.

File: ***Dataset S01***[1] in supporting information of Graf et al. (2014)

Dataset 2

List of first activated genes in bovine embryos identified with method (i) the detection of transcripts which are not present in oocytes (first expression), method (ii) detection of transcripts with the paternal allele as marker for the onset of EGA (paternal SNP) and method (iii) the detection of primary transcripts with intronic sequences (primary transcripts).

File: ***Dataset S02***[2] in supporting information of Graf et al. (2014)

[1] http://www.pnas.org/content/suppl/2014/03/03/1321569111.DCSupplemental/sd01.xls
[2] http://www.pnas.org/content/suppl/2014/03/03/1321569111.DCSupplemental/sd02.xls

References

Blencowe, B. J., Ahmad, S., Lee, L. J. 2009. Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes. Genes Dev. 23 (12), 1379–1386.

Braude, P., Bolton, V., Moore, S. 1988. Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. Nature 332 (6163), 459–461.

Butterfield, Y. S., Kreitzman, M., Thiessen, N., Corbett, R. D., Li, Y., Pang, J., Ma, Y. P., Jones, S. J., Birol, I. 2014. JAGuaR: Junction Alignments to Genome for RNA-Seq Reads. PLoS ONE 9 (7), e102398.

Camous, S., Kopecny, V., Flechon, J. E. 1986. Autoradiographic detection of the earliest stage of [3H]-uridine incorporation into the cow embryo. Biol. Cell 58 (3), 195–200.

Chitwood, J. L., Rincon, G., Kaiser, G. G., Medrano, J. F., Ross, P. J. 2013. RNA-seq analysis of single bovine blastocysts. BMC Genomics 14, 350.

Daetwyler, H. D., Capitan, A., Pausch, H., Stothard, P., Binsbergen, R. van, Brøndum, R. F., Liao, X., Djari, A., Rodriguez, S. C., Grohs, C., Esquerré, D., Bouchez, O., Rossignol, M. N., Klopp, C., Rocha, D., Fritz, S., Eggen, A., Bowman, P. J., Coote, D., Chamberlain, A. J., Anderson, C., VanTassell, C. P., Hulsegge, I., Goddard, M. E., Guldbrandtsen, B., Lund, M. S., Veerkamp, R. F., Boichard, D. A., Fries, R., Hayes, B. J. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat. Genet. 46 (8), 858–865.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., Angel, G. del, Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., Daly, M. J. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43 (5), 491–498.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29 (1), 15–21.

Engström, P. G., Steijger, T., Sipos, B., Grant, G. R., Kahles, A., Rätsch, G., Goldman, N., Hubbard, T. J., Harrow, J., Guigo, R., Bertone, P., Alioto, T., Behr, J., Bertone, P., Bohnert, R., Campagna, D., Davis, C. A., Dobin, A., Engström, P. G., Gingeras, T. R., Goldman, N., Grant, G. R., Guigó, R., Harrow, J., Hubbard, T. J., Jean, G., Kahles, A., Kosarev, P., Li, S., Liu, J., Mason, C. E., Molodtsov, V., Ning, Z., Ponsting, H., Prins, J. F., Rätsch, G., Ribeca, P., Seledtsov, I., Sipos, B., Solovyev, V., Steijger, T., Valle, G., Vitulo, N., Wang, K., Wu, T. D., Zeller, G. 2013. Systematic evaluation of spliced alignment programs for RNA-seq data. Nat. Methods 10 (12), 1185–1191.

Frei, R. E., Schultz, G. A., Church, R. B. 1989. Qualitative and quantitative changes in protein synthesis occur at the 8-16-cell stage of embryogenesis in the cow. J. Reprod. Fertil. 86 (2), 637–641.

Giraldez, A. J., Mishima, Y., Rihel, J., Grocock, R. J., Van Dongen, S., Inoue, K., Enright, A. J., Schier, A. F. 2006. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science 312 (5770), 75–79.

Graf, A., Krebs, S., Zakhartchenko, V., Schwalb, B., Blum, H., Wolf, E. 2014. Fine mapping of genome activation in bovine embryos by RNA sequencing. Proc. Natl. Acad. Sci. U.S.A. 111 (11), 4139–4144.

Huang, W., Khatib, H. 2010. Comparison of transcriptomic landscapes of bovine embryos using RNA-Seq. BMC Genomics 11, 711.

Hyttel, P., Viuff, D., Avery, B., Laurincik, J., Greve, T. 1996. Transcription and cell cycle-dependent development of intranuclear bodies and granules in two-cell bovine embryos. J. Reprod. Fertil. 108 (2), 263–270.

Jansen, S., Aigner, B., Pausch, H., Wysocki, M., Eck, S., Benet-Pages, A., Graf, E., Wieland, T., Strom, T. M., Meitinger, T., Fries, R. 2013. Assessment of the genomic variation in a cattle population by re-sequencing of key animals at low to medium coverage. BMC Genomics 14, 446.

Jean, G., Kahles, A., Sreedharan, V. T., De Bona, F., Rätsch, G. 2010. RNA-Seq read alignments with PALMapper. Curr Protoc Bioinformatics Chapter 11, Unit 11.6.

Kanka, J., Nemcova, L., Toralova, T., Vodickova-Kepkova, K., Vodicka, P., Jeseta, M., Machatkova, M. 2012. Association of the transcription profile of bovine oocytes and embryos with developmental potential. Anim. Reprod. Sci. 134 (1-2), 29–35.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S. L. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14 (4), R36.

Kues, W. A., Sudheer, S., Herrmann, D., Carnwath, J. W., Havlicek, V., Besenfelder, U., Lehrach, H., Adjaye, J., Niemann, H. 2008. Genome-wide expression profiling reveals distinct clusters of transcriptional regulation during bovine preimplantation development in vivo. Proc. Natl. Acad. Sci. U.S.A. 105 (50), 19768–19773.

Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A., Dewey, C. N. 2010. RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics 26 (4), 493–500.

Marcelino, L. A., Backman, V., Donaldson, A., Steadman, C., Thompson, J. R., Preheim, S. P., Lien, C., Lim, E., Veneziano, D., Polz, M. F. 2006. Accurately quantifying low-abundant targets amid similar sequences by revealing hidden correlations in oligonucleotide microarray data. Proc. Natl. Acad. Sci. U.S.A. 103 (37), 13629–13634.

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., Gilad, Y. 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 18 (9), 1509–1517.

Memili, E., Dominko, T., First, N. L. 1998. Onset of transcription in bovine oocytes and preimplantation embryos. Mol. Reprod. Dev. 51 (1), 36–41.

Memili, E., First, N. L. 1999. Control of gene expression at the onset of bovine embryonic development. Biol. Reprod. 61 (5), 1198–1207.

Misirlioglu, M., Page, G. P., Sagirkaya, H., Kaya, A., Parrish, J. J., First, N. L., Memili, E. 2006. Dynamics of global transcriptome in bovine matured oocytes and preimplantation embryos. Proc. Natl. Acad. Sci. U.S.A. 103 (50), 18905–18910.

Miyanari, Y., Torres-Padilla, M. E. 2012. Control of ground-state pluripotency by allelic regulation of Nanog. Nature 483 (7390), 470–473.

Mondou, E., Dufort, I., Gohin, M., Fournier, E., Sirard, M. A. 2012. Analysis of microRNAs and their precursors in bovine early embryonic development. Mol. Hum. Reprod. 18 (9), 425–434.

Østrup, O., Andersen, I. S., Collas, P. 2013. Chromatin-linked determinants of zygotic genome activation. Cell. Mol. Life Sci. 70 (8), 1425–1437.

Piskol, R., Ramaswami, G., Li, J. B. 2013. Reliable identification of genomic variants from RNA-seq data. Am. J. Hum. Genet. 93 (4), 641–651.

Plante, L., Plante, C., Shepherd, D. L., King, W. A. 1994. Cleavage and 3H-uridine incorporation in bovine embryos of high in vitro developmental potential. Mol. Reprod. Dev. 39 (4), 375–383.

Prioleau, M. N., Huet, J., Sentenac, A., Mechali, M. 1994. Competition between chromatin and transcription complex assembly regulates gene expression during early development. Cell 77 (3), 439–449.

Ramachandra, R. K., Salem, M., Gahr, S., Rexroad, C. E., Yao, J. 2008. Cloning and characterization of microRNAs from rainbow trout (Oncorhynchus mykiss): their expression during early embryonic development. BMC Dev. Biol. 8, 41.

Schier, A. F. 2007. The maternal-zygotic transition: death and birth of RNAs. Science 316 (5823), 406–407.

Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J., Cramer, P. 2013. Transcriptome surveillance by selective termination of noncoding RNA synthesis. Cell 155 (5), 1075–1087.

Seledtsov, I., Molodtsov, V., Solovyev, V. 2012. ReadsMap: Aligning next generation sequences to the genome sequence. URL: http://linux5.softberry.com/cgi-bin/berry/programs/ReadsMap (visited on 08/10/2014).

Sirard, M. A. 2010. Activation of the embryonic genome. Soc Reprod Fertil Suppl 67, 145–158.

Sirard, M. A. 2012. Factors affecting oocyte and embryo transcriptomes. Reprod. Domest. Anim. 47 Suppl 4, 148–155.

Svarcova, O., Laurincik, J., Avery, B., Mlyncek, M., Niemann, H., Maddox-Hyttel, P. 2007. Nucleolar development and allocation of key nucleolar proteins require de novo transcription in bovine embryos. Mol. Reprod. Dev. 74 (11), 1428–1435.

Tadros, W., Goldman, A. L., Babak, T., Menzies, F., Vardy, L., Orr-Weaver, T., Hughes, T. R., Westwood, J. T., Smibert, C. A., Lipshitz, H. D. 2007. SMAUG is a major regulator of maternal mRNA destabilization in Drosophila and its translation is activated by the PAN GU kinase. Dev. Cell 12 (1), 143–155.

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

Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A., Conesa, A. 2011. Differential expression in RNA-seq: a matter of depth. Genome Res. 21 (12), 2213–2223.

Telford, N. A., Watson, A. J., Schultz, G. A. 1990. Transition from maternal to embryonic control in early mammalian development: a comparison of several species. Mol. Reprod. Dev. 26 (1), 90–100.

Trapnell, C., Pachter, L., Salzberg, S. L. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25 (9), 1105–1111.

Vigneault, C., Gravel, C., Vallee, M., McGraw, S., Sirard, M. A. 2009. Unveiling the bovine embryo transcriptome during the maternal-to-embryonic transition. Reproduction 137 (2), 245–257.

Viuff, D., Avery, B., Greve, T., King, W. A., Hyttel, P. 1996. Transcriptional activity in in vitro produced bovine two- and four-cell embryos. Mol. Reprod. Dev. 43 (2), 171–179.

Wang, H., Dey, S. K. 2006. Roadmap to embryo implantation: clues from mouse models. Nat. Rev. Genet. 7 (3), 185–199.

Wang, K., Singh, D., Zeng, Z., Coleman, S. J., Huang, Y., Savich, G. L., He, X., Mieczkowski, P., Grimm, S. A., Perou, C. M., MacLeod, J. N., Chiang, D. Y., Prins, J. F., Liu, J. 2010. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res. 38 (18), e178.

Wang, Z., Gerstein, M., Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10 (1), 57–63.

Wu, T. D., Nacu, S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26 (7), 873–881.