# 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], and Eckhard Wolf[a,b,1,2]

[a]Laboratory for Functional Genome Analysis, [b]Chair for Molecular Animal Breeding and Biotechnology, and [c]Department of Biochemistry, Center for Integrated Protein Science, Gene Center, Ludwig-Maximilians-Universität München, 81377 Munich, Germany

Edited* by George E. Seidel, Colorado State University, Fort Collins, CO, and approved February 7, 2014 (received for review November 18, 2013)

**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 × $10^3$ 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.**

Early embryonic development is governed by maternal transcripts and proteins stored within the oocyte during oogenesis (reviewed in ref. 1). 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 ref. 2), at the four- to eight-cell stage in human (3) and pig embryos (reviewed in ref. 2), and at the eight- to 16-cell stage in bovine embryos (reviewed in ref. 2). 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 insights 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 (4). 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 (5, 6). 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) (7). 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

### 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.

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.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE52415).

[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.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1321569111/-/DCSupplemental.

matured and fertilized with semen from a single Brahman bull (*Bos t. indicus*). Because the two *B. taurus* subspecies are relatively distant (8), 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 (9). 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 (Fig. 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 (Fig. S2).

**Global View on the Transcriptome of Bovine Oocytes and Preimplantation Embryos.** In all developmental stages, transcripts from 12.4 to 13.7 × $10^3$ different genes were detected (Fig. 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.

**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 upregulated for the analyzed developmental stage to be designated as first expressed (Dataset S1). One example is the Nanog homeobox (*NANOG*) gene, which was—in accordance with a previous report (10)—found to be first expressed at the eight-cell stage (Fig. 2*A*). 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 (Fig. 2*B* and Dataset S2).

**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 Fig. 2*C*, 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 (Fig. 2*D* and Dataset S2).

**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 (11). As shown in Fig. 2*A*, 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 (Fig. S2). To discriminate intronic reads in primary transcripts from intronic reads resulting, for example, 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 ≥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 (Fig. S3). As an example, the activation of the Krüppel-like factor 4 (*KLF4*) gene is shown in Fig. 2*E*. *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

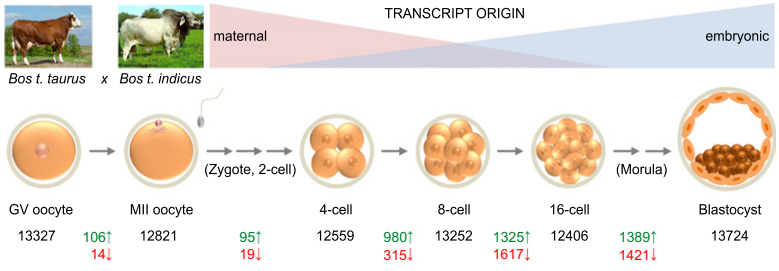



**Fig. 1.** Experimental design. Oocytes from *B. t. taurus* cows were fertilized with sperm of a single *B. t. indicus* bull. Pools of 10 oocytes/embryos were harvested at the GV, MII, four-cell, eight-cell, 16-cell, and blastocyst stage and processed for RNA-Seq. For each stage the total number of 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).

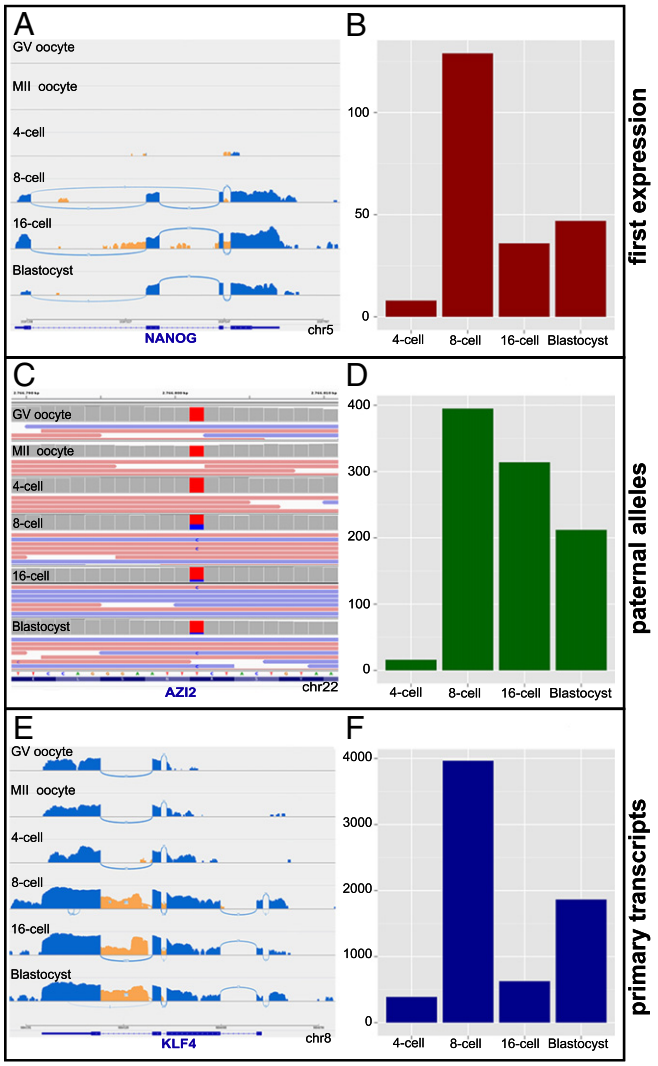



**Fig. 2.** Different strategies for fine mapping of genome activation in bovine embryos by RNA-Seq. (*A* and *B*) Detection of transcripts that are not present in oocytes. (*A*) Sashimi plot (30) of *NANOG* showing the absence of transcripts in oocytes, very limited transcription at the four-cell stage, 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 (31) 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 (30) of *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.

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 (Fig. 2*F* and Dataset S2).

**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 (Fig. 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 the 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; Fig. S3). We found a nearly random distribution of dots in GV and MII oocytes and in four-cell embryos (Fig. 3), indicating the presence of mainly mature transcripts. The density of dots increased markedly at the eight-cell 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*). 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 (Fig. 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

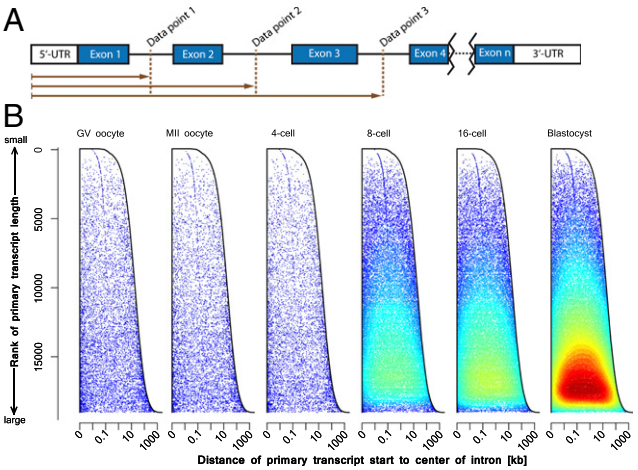



**Fig. 3.** Global view of intron transcription of all annotated genes. For each intron of all annotated genes a data point was computed by calculating the distance of the transcription start site to the center of the intron in bases (*A*). Furthermore, an RINP value was calculated by summing up all 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 (Fig. 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.

DEVELOPMENTAL BIOLOGY

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

## Discussion

Expression profiling—either by RT-PCR analyses of candidate genes (reviewed in ref. 12) or by holistic approaches using array-based techniques (13)—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 ref. 14). 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 (15) and of bovine conceptuses (days 10, 13, 16, and 19) (16); 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 (15) or even single embryonic cells (17), 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 (18), 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 ref. 19). 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 (17).

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 (20). 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 ref. 1).

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

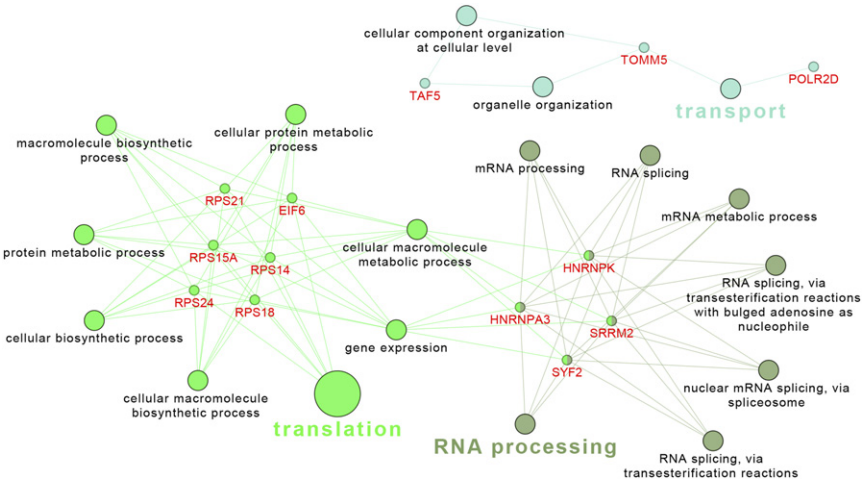



**Fig. 4.** Functionally grouped GO terms for the genes activated before major genome activation in bovine embryos. Genes activated at the four-cell stage were detected by the presence of de novo transcripts ($n$ = 8), transcripts with paternal-specific SNPs ($n$ = 16), or primary transcripts with intronic sequences ($n$ = 390) and functionally analyzed with the *ClueGO* (29) 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.

*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 (Fig. 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 (6), we found a significant overlap (58%; Fisher's exact test, $P < 0.01$) with our eight-cell activated genes (Fig. 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 ref. 21). 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 preimplantation embryos (22). Further, we identified *HNRNPA2B1* (*LOC516616*), which interacts with *SOX2* (23), a key transcription factor for embryonic stem cell pluripotency (24). Another gene activated before major EGA is *KLF17*. Its product, Krüppel-like factor 17, can activate or suppress transcription (25). 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 (9). 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 $20 \times 10^6$ 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* (18) (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 (26). 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* (27) (v.0.1.13). From the resulting pileups, SNPs were called using *Varscan* (28) (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 × Simmental and the Jersey × Simmental hybrid embryos in all developmental stages. Last, an 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 Fig. 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; Fig. 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 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 (Fig. 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* (18), 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* (29) 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.

1. Tadros W, Lipshitz HD (2009) The maternal-to-zygotic transition: a play in two acts. *Development* 136(18):3033–3042.
2. Sirard MA (2012) Factors affecting oocyte and embryo transcriptomes. *Reprod Domest Anim* 47(Suppl 4):148–155.
3. 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.
4. Memili E, Dominko T, First NL (1998) Onset of transcription in bovine oocytes and preimplantation embryos. *Mol Reprod Dev* 51(1):36–41.
5. Misirlioglu M, et al. (2006) Dynamics of global transcriptome in bovine matured oocytes and preimplantation embryos. *Proc Natl Acad Sci USA* 103(50):18905–18910.
6. Vigneault C, Gravel C, Vallée M, McGraw S, Sirard MA (2009) Unveiling the bovine embryo transcriptome during the maternal-to-embryonic transition. *Reproduction* 137(2):245–257.
7. Gad A, et al. (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.
8. Troy CS, et al. (2001) Genetic evidence for Near-Eastern origins of European cattle. *Nature* 410(6832):1088–1091.
9. Bauersachs S, et al. (2009) The endometrium responds differently to cloned versus fertilized embryos. *Proc Natl Acad Sci USA* 106(14):5681–5686.
10. Khan DR, et al. (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.
11. Ameur A, et al. (2011) Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat Struct Mol Biol* 18(12): 1435–1440.
12. Wrenzycki C, et al. (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.
13. Kues WA, et al. (2008) Genome-wide expression profiling reveals distinct clusters of transcriptional regulation during bovine preimplantation development in vivo. *Proc Natl Acad Sci USA* 105(50):19768–19773.
14. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: A revolutionary tool for transcriptomics. *Nat Rev Genet* 10(1):57–63.
15. Chitwood JL, Rincon G, Kaiser GG, Medrano JF, Ross PJ (2013) RNA-seq analysis of single bovine blastocysts. *BMC Genomics* 14:350.
16. Mamo S, et al. (2011) RNA sequencing reveals novel gene clusters in bovine conceptuses associated with maternal recognition of pregnancy and implantation. *Biol Reprod* 85(6):1143–1151.
17. Xue Z, et al. (2013) Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* 500(7464):593–597.
18. Demyda-Peyrás S, et al. (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.
19. Lonergan P, Fair T (2008) In vitro-produced bovine embryos: Dealing with the warts. *Theriogenology* 69(1):17–22.
20. Zhu L, et al. (2009) Patterns of exon-intron architecture variation of genes in eukaryotic genomes. *BMC Genomics* 10:47.
21. Swinburne IA, Silver PA (2008) Intron delays and transcriptional timing during development. *Dev Cell* 14(3):324–330.
22. Yamada M, et al. (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.
23. Fang X, et al. (2011) Landscape of the SOX2 protein-protein interactome. *Proteomics* 11(5):921–934.
24. Masui S, et al. (2007) Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* 9(6):625–635.
25. van Vliet J, et al. (2006) Human KLF17 is a new member of the Sp/KLF family of transcription factors. *Genomics* 87(4):474–482.
26. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11(10):R106.
27. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
28. Koboldt DC, et al. (2009) VarScan: Variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* 25(17):2283–2285.
29. Bindea G, et al. (2009) ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25(8):1091–1093.
30. Katz Y, Wang ET, Airoldi EM, Burge CB (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* 7(12):1009–1015.
31. Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform* 14(2):178–192.