Title

Transcriptome profiling of individual rhesus macaque oocytes and preimplantation embryos

Journal

Biology of Reproduction, 97(3)

ISSN

0006-3363

Authors

Chitwood, James L
Burruel, Victoria R
Halstead, Michelle M
et al.

Publication Date

2017-09-01

DOI

10.1093/biolre/iox114

Peer reviewed

Biology of Reproduction, 2017, 97(3), 353–364
doi:10.1093/biolre/iox114
Research Article
Advance Access Publication Date: 4 September 2017

Research Article

Transcriptome profiling of individual *rhesus macaque* oocytes and preimplantation embryos†

James L. Chitwood[1], Victoria R. Burruel[2], Michelle M. Halstead[1], Stuart A. Meyers[2] and Pablo J. Ross[1,*]

[1]Department of Animal Science, University of California, Davis, California, USA and [2]Department of Anatomy, Physiology, and Cell Biology, School of Veterinary Medicine, University of California, Davis, California, USA

*__Correspondence:__ Department of Animal Science, University of California Davis, One Shields Avenue, Davis, CA 9516, USA. E-mail: pross@ucdavis.edu

†__Grant Support:__ The funding from the National Center for Research Resources (NCRR) (R01RR016581) to SM and NIH/NICHD (R01HD070044) to PR are gratefully acknowledged. MMH was supported by a USDA NIFA National Needs Fellowship.

Received 21 June 2017; Revised 30 August 2017; Accepted 1 September 2017

Abstract

Early mammalian embryonic transcriptomes are dynamic throughout the process of preimplantation development. Cataloging of primate transcriptomics during early development has been accomplished in humans, but global characterization of transcripts is lacking in the rhesus macaque: a key model for human reproductive processes. We report here the systematic classification of individual macaque transcriptomes using RNA-Seq technology from the germinal vesicle stage oocyte through the blastocyst stage embryo. Major differences in gene expression were found between sequential stages, with the 4- to 8-cell stages showing the highest level of differential gene expression. Analysis of putative transcription factor binding sites also revealed a striking increase in key regulatory factors in 8-cell embryos, indicating a strong likelihood of embryonic genome activation occurring at this stage. Furthermore, clustering analyses of gene co-expression throughout this period resulted in distinct groups of transcripts significantly associated to the different embryo stages assayed. The sequence data provided here along with characterizations of major regulatory transcript groups present a comprehensive atlas of polyadenylated transcripts that serves as a useful resource for comparative studies of preimplantation development in humans and other species.

Summary Sentence

The embryonic genome is activated at the 8-cell stage in the rhesus macaque.

Key words: Rhesus, embryo, transcriptome, EGA, oocytes.

Introduction

The transcriptomic profile of the mammalian oocyte and developing embryo is subject to complex changes throughout the entire period of preimplantation development. Molecular shifts are the signatures of crucial biological milestones (e.g. oocyte maturation [1], maternal to zygotic transition of transcription [2], embryonic genome activation [3], and differentiation of cellular lineages [4–6]) that must be achieved to ensure the developmental competence of the oocyte and subsequent embryo. At worst, failure to coordinate these events can lead to embryonic mortality [7,8], compromise the viability of an embryo for transfer or cryopreservation, and affect various aspects of the resultant offspring's health well throughout the duration of its life [9,10].

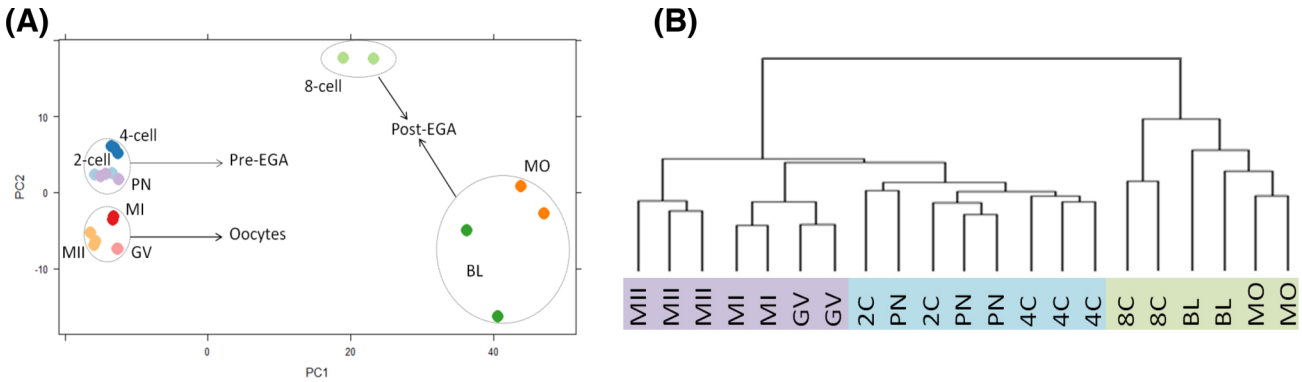

Figure 1. Global transcriptome analysis segregates individual embryos to their respective stages. (**A**) Clustering of samples by global transcript levels following PCA. Circles and arrows indicate clustering of samples within broad stages of preimplantation development: oocyte, pre-EGA embryo, and post-EGA embryo. The post-EGA embryos divided into two groups: the 8-cell samples and the morula/blastocyst samples. (**B**) Hierarchical clustering of sample relatedness based on global transcript levels. Distances of clustering reflect those seen in the PCA graph, with the oocyte stages indicated in purple, the pre-EGA embryo stages in blue, and the post-EGA stages in green.

The onset of significant embryonic de novo gene transcription, known as embryonic genome activation (EGA), is of particular interest during early development. Embryonic survival involves usage of maternally-derived proteins and transcripts [11], while the genome of the embryo itself remains transcriptionally quiescent until a specific developmental time point that varies between species. In mice, EGA occurs at the 2-cell (2C) stage [12]: earlier within the preimplantation period when compared to bovine (8–16 cell stage [13]), humans (4–8 cell stage [14]), and macaques (6–8 cell stage [3]). The initiation of EGA is most obviously manifested through large-scale increased abundance of transcripts; however, degradation of maternal transcripts and protein as well as remodeling of chromatin conformation also occur, underscoring changes in the epigenetic status of the embryonic genome at this time point [15]. Errors in the regulation of these biological phenomena can result in embryonic loss. Therefore, understanding the biology regulating the onset of an organism's transcriptional program is of enormous importance; yet, the mechanisms that promulgate and sustain EGA in mammals have yet to be completely deciphered.

Towards the goal of improving our understanding of early mammalian development, our objective was to characterize the transcriptomes of individual rhesus macaque embryos by sequencing global transcripts (RNA-Seq) obtained from sequential stages of preimplantation development. The macaque is a long-standing model for human reproductive biology [16], and there is a current dearth of global-scale characterization of early development in this species. Furthermore, the expense and difficulty inherent to working with non-human primates makes the data described in this paper a valuable addition to the understanding of early development in primates, and presents a useful supplement to studies conducted with human samples.

We report here an analysis of comprehensive transcriptome data for rhesus macaque embryos for each distinct stage of preimplantation development spanning from the germinal vesicle (GV) to the blastocyst (BL). Our data indicate a large increase in gene transcript levels beginning at the 8-cell (8C) stage, indicating major EGA at this time point. Increases in alignment of reads to intronic sequence at this time point further support this conclusion. Additionally, putative key transcription factors also increased in abundance at the 8C stage, indicating a global shift in the regulatory landscape. Finally, clustering analyses of the data identified distinct groups of co-expressing genes with key significance to early developmental processes.

Materials and methods

Reagents/chemicals

All chemicals were obtained from Sigma-Aldrich Company (St. Louis, MO) unless otherwise stated.

Superovulation, oocyte collection, and fertilization

Animals were housed at the California National Primate Research Center and maintained according to Institutional Animal Care and Use Committee protocols at the University of California. Methods for harvesting of oocytes from superovulated donors, semen collection, and subsequent intracytoplasmic sperm injection (ICSI) have been detailed previously [17–19]. Briefly, rhesus females were superovulated with recombinant human follicle-stimulating hormone (FSH) (30 IU i.m. twice daily, Follistim, Merck) and Acycline (60 μg/kg/day s.c.) for 8 days starting on days 1–4 of menses. Recombinant human Luteinizing hormone (LH) (30 IU s.c. twice daily) was injected during the last 3 days of the 8-day interval. Oocytes were collected by follicular aspiration of ovaries recovered following euthanasia, 35 h after human chorionic gonadotropin (hCG) treatment (1300 IU i.m., Ovidrel, EMD Serono). Sperm used for ICSI were collected from semen by electroejaculation that was washed and suspended in HEPES-Biggers, Whitten, and Whittingham media. Embryos were produced by ICSI of meiosis II (MII) oocytes [20–22] with visibly motile sperm and cultured in 25 μL drops of HECM-9 [23] under oil and cultured at 37°C in 6% CO_2, 5% O_2, and 89% N_2.

Oocyte and embryo processing for complementary DNA synthesis

Individual oocytes and embryos were collected at different developmental stages and frozen until use. GV, MI, and MII oocytes were classified according to the presence or absence of a polar body. ICSI-derived embryos were collected at different developmental stages (2-cell, 4-cell, 8-cell, morula, and blastocyst) based on morphological appearance and frozen in a minimal volume of HECM-9 at –80°C until further processing. The Clontech SMARTer Ultra Low Input RNA Kit for Illumina Sequencing (Clontech, Mountain View, CA) was used for complementary DNA (cDNA) synthesis (oligo dT priming). Fifteen cycles of PCR amplification were done for most samples, except for blastocysts (12 and 13 cycles) and samples requiring

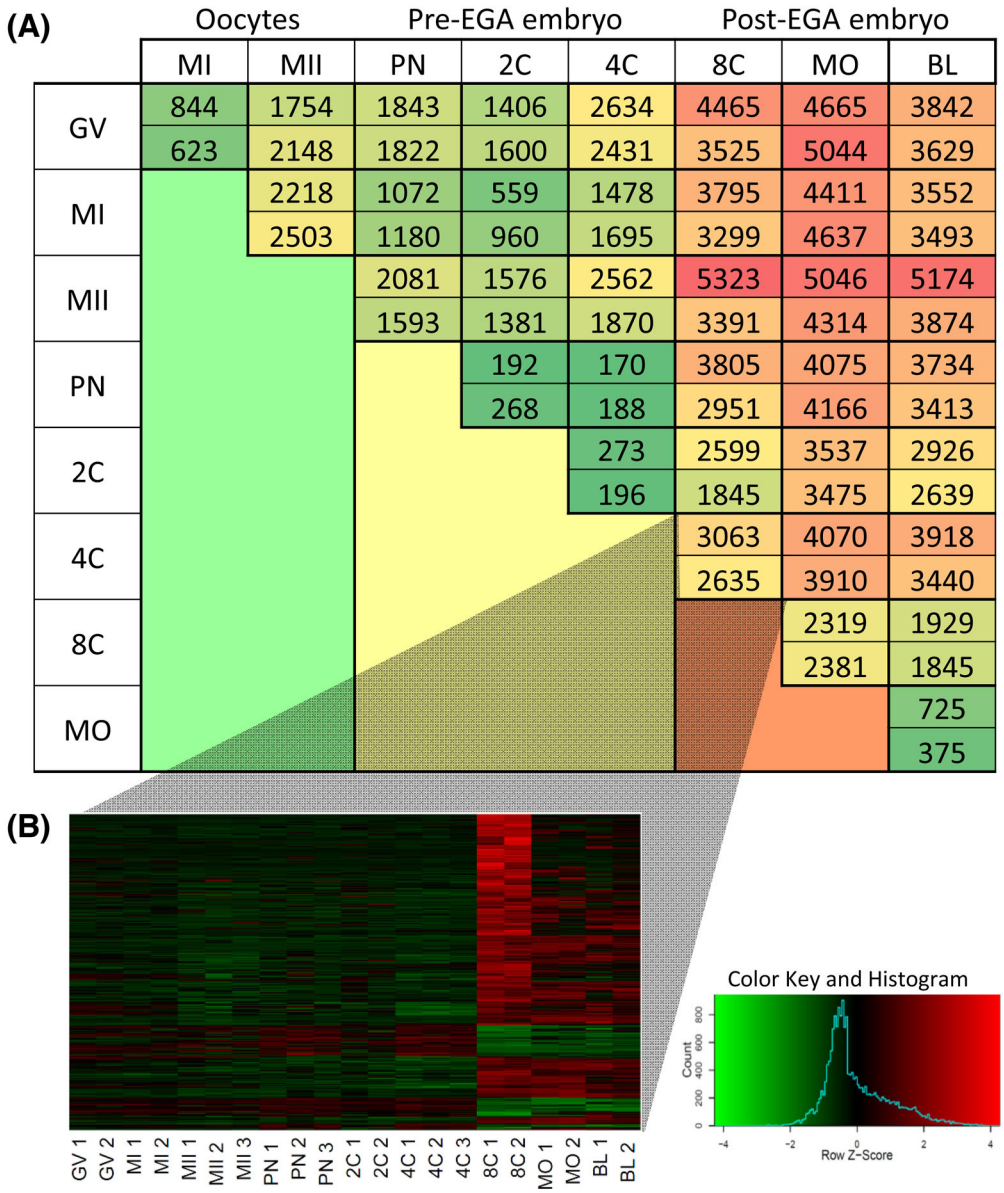

Figure 2. DEGs between sequential developmental stages. (**A**) The number of DEGs between each developmental stage (adjusted *P*-value < 0.05) was found using DESeq2. The top and bottom cell from each comparison indicate number of up- and downregulated genes, respectively (fold change >2 or <0.5). Color gradation indicates lower (green) or higher (red) number of genes. (**B**) Transcript abundance in each sample of genes differentially expressed between the 4-cell and 8-cell stages (adjusted *P*-value < 0.05 and fold change >8 or <0.125).

re-amplification to obtain sufficient quantities of cDNA pronuclear (PN 2, 2C's 1 and 3, 8C's 1 and 3, and MO 2). Re-amplification cycle number varied between 4 and 9 cycles. One microliter of cDNA (of 12 μL total following purification) was used to verify appropriate size, quantity, and lack of overamplification (based on Clontech's protocol) of the product with a High Sensitivity DNA Bioanalyzer chip (Agilent Technologies, Santa Clara, CA).

RNA sequencing library preparation from Clontech complementary DNA

Amplified cDNA was diluted by adding 49 μL of Illumina TruSeq DNA kit Resuspension Buffer to 11 μL of cDNA. Samples were sonicated with an S2 Sonicator (Covaris Inc., Woburn, MA) using parameters provided in the Clontech protocol. Sonicated cDNA was used as input for an Illumina TruSeq DNA library preparation kit (Illumina, San Diego, CA) following the gel-free protocol. The low amount of input to the TruSeq kit necessitated dilution of adapters depending on the amount of cDNA input used (between 1:4 and 1:50). Adapter-ligated products were amplified for 15 PCR cycles. Library size and concentration were assessed using a 1K DNA chip and Experion Instrument (Bio-Rad, Hercules, CA) and Qubit DNA HS assay (Invitrogen, Carlsbad, CA). Libraries were sequenced as 100 bp single-end reads (50 bp for MII and GV samples) with a Hi-Seq 2000 apparatus at the Vincent J. Coates Genomics Sequencing Laboratory (University of California, Berkeley).

Read mapping and analysis of differential expression

All read processing and mapping was performed using the CLC Genomics Workbench 6.0.2 (Qiagen, Valencia, CA). Reads were

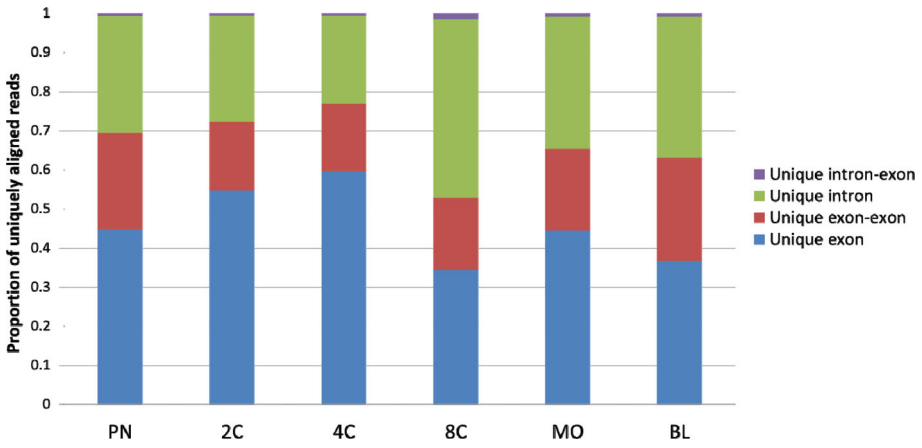

Figure 3. Proportion of reads aligning to different gene regions during early development. Different types of unique read alignment as a proportion of all uniquely aligned reads at specific stages. An increase in the proportion of intron aligned reads (both intron and intron-exon) at the 8-cell stage suggests an increase transcriptional activity as compared to other stages.

Table 1. Gene ontology enrichment for genes not differentially expressed from PN to BL (padj > 0.05) and with significant differences in intron alignment between 4C and 8C samples ($P < 0.01$).

	GO term	Gene count	*P*-value
Biological process	Steroid hormone receptor signaling pathway (transcription factors)	12	3.70E-06
	Protein localization	33	1.10E-04
	Cell cycle	48	1.10E-04
	Regulation of pigmentation during development	5	3.80E-04
	Positive regulation of macromolecule metabolic process	47	1.70E-03
	Response to DNA damage stimulus	24	5.00E-03
	Cellular macromolecule catabolic process	39	6.40E-03
	Regulation of transcription	106	6.70E-03
	Vesicle docking	5	1.10E-02
	Posttranscriptional regulation of gene expression	15	1.40E-02
Molecular function	DNA binding	190	9.80E-07
	Steroid hormone receptor binding (transcription factors)	10	6.90E-06
	Transition metal ion binding	127	4.30E-03
	DNA-dependent ATPase activity	8	4.60E-03
	Microtubule binding	8	1.60E-02
	Protein serine/threonine kinase activity	4	2.10E-02
Cellular component	Intracellular organelle lumen (nucleoplasm)	93	9.40E-06
	Spliceosome	12	4.80E-03
	Intracellular nonmembrane-bounded organelle (cytoskeleton)	108	9.50E-03
	Cytoplasmic membrane-bounded vesicle	30	1.10E-02
	Mitochondrion	49	2.90E-02
	Eukaryotic translation initiation factor 4F complex	3	3.40E-02

processed based on quality (quality limit = 0.05) and GC content at the beginning of the read (10 bp cut from 5' end) using tools previously described [24]. Mapping was done to the reference sequences for *Macaca mulatta* obtained from Ensembl and annotated with a .gtf file provided as part of RhesusBase [25]. A gene was considered expressed if the reads per kilobase of transcript model per million mapped reads was >0.4. Differential gene expression between sample groups was analyzed using the DESeq2 package in R [26] using total exon counts for each gene. Genes were considered differentially expressed if the fold change between stages was >2 or <0.5 and had an adjusted *P*-value (adjp) <0.05. In addition, DESeq2 was used for hierarchical clustering of samples based on all expressed genes and principal component analysis.

Network and cluster analyses

The self-organizing tree algorithm (SOTA) [27] function from the clValid package in R was used to identify clusters of co-expressed genes across developmental stages (from PN to BL). The number of clusters generated was set at 16.

Groups of genes were identified as belonging to co-expression clusters, and network analysis was performed using the weighted gene co-expression network analysis (WGCNA) package [28] in R. Variance-stabilized expression values for all annotated genes from DESeq2 were used as input for block formation and module detection. A single block was used for module detection. The soft threshold for the correlation matrix of co-expression was 12, and a correlation cutoff of 0.7 was used for merging highly correlated modules. Modules were assigned to developmental stages based on

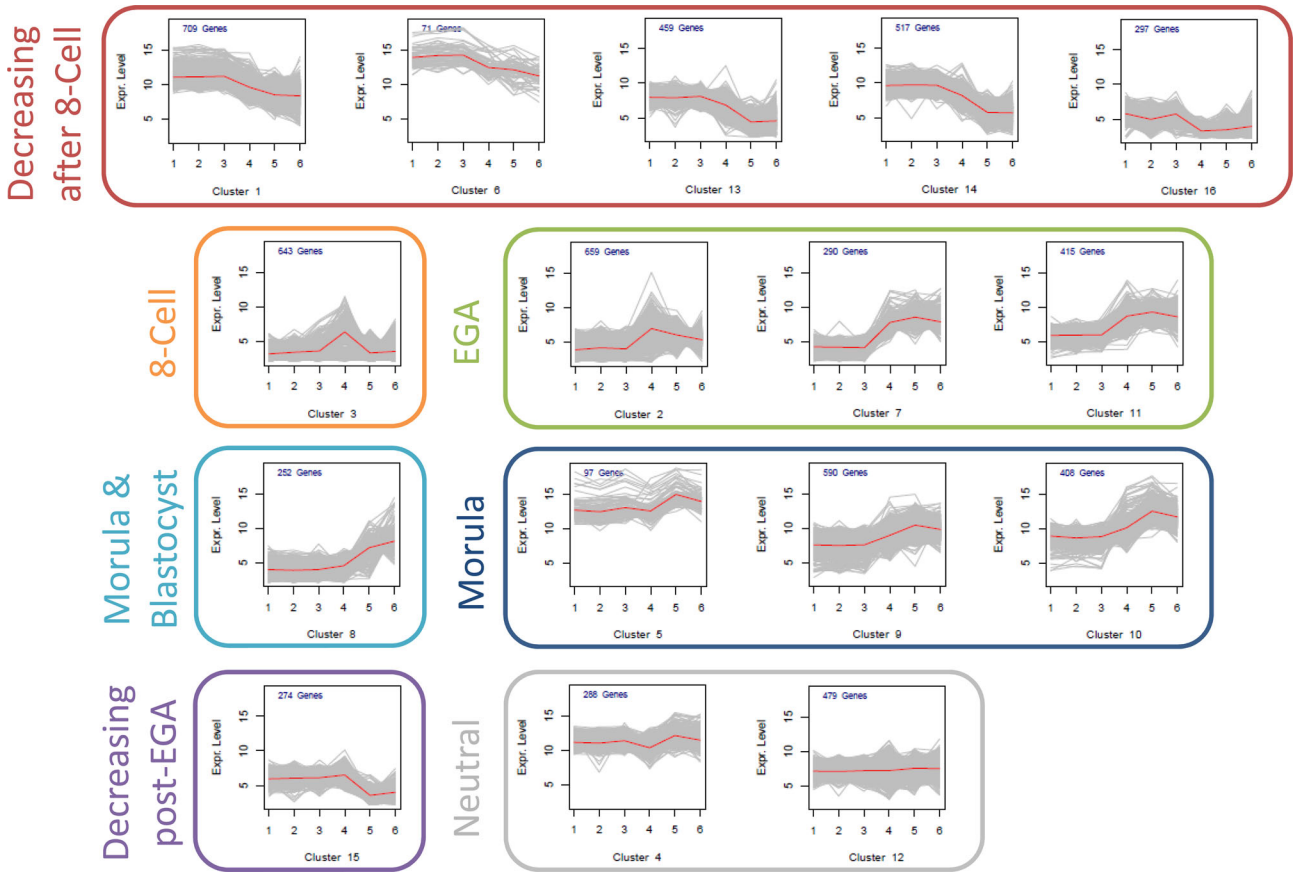

Figure 4. Clustering of genes by sequential stages of development. Genes expressed from PN to BL were analyzed for co-expression using the SOTA clustering (x-axis labels: 1 = PN, 2 = 2C, 3 = 4C, 4 = 8C, 5 = MO, 6 = BL). Clusters were assigned an expression profile based on fluctuations between stages: decreasing after 8C (5 clusters; red), 8C elevated (1 cluster; orange), EGA (3 clusters; green), MO and BL elevated (1 cluster; cyan), MO elevated (3 clusters; navy), decreasing post-EGA (1 cluster; purple), and neutral (2 clusters; grey).

high correlation and statistically significant *P*-values to specific sample groups. Hub genes within modules were identified using a cutoff value of kME > 0.95 for module membership. Connectivity between genes within a module was visualized using VisANT [29].

Gene ontology (GO) enrichment of modules and clusters was performed using the Functional Annotation tool from the DAVID Bioinformatics Database [30,31]. Enrichment clusters were deemed significant for $P < 0.05$. Enrichment for GO terms was performed with medium stringency settings.

Identification of key regulatory factors

For each pair of consecutive stages, promoters of upregulated differentially expressed genes (DEGs) were evaluated for presence or absence of transcription factor binding site (TFBS) for each TF family. To determine if a given TFBS is over-represented in a particular set of DEG promoters as compared to promoters of all genes, we applied the hypergeometric distribution as a model. Thus, *P*-values were generated for each TFBS (corresponding to a TF family) in each pairwise comparison, and then adjusted for multiple testing using the Benjamini-Hochberg correction. Only TF families with an adjusted *P*-value < 0.05 that appeared more frequently in DEG promoters than all promoters were considered significantly over-represented. Genes encoding TFs that belonged to over-represented TF families were then examined to determine if they were also upregulated DEGs, which classified them as putative key regulatory factors (Figure 4A). Gene ontology and Swiss-Prot Protein Information Resource (SP-PIR) term enrichment were performed using the Functional Annotation tool from the DAVID Bioinformatics Database [30,31], using medium stringency settings. Generated *P*-values were corrected using the Benjamini-Hochberg correction and terms were deemed significant if the adj-$P < 0.05$ and gene count >30.

Results

Sequencing and sample mapping statistics

Individual rhesus macaque oocytes and embryos were prepared for RNA-Seq using oligo(dT)-based cDNA synthesis and amplification prior to sequencing library construction. Sequencing of 21 samples from nine distinct developmental stages yielded a total of 508 million reads, with an average of 23 million reads per sample. All raw sequencing data are available through the Gene Expression Omnibus (GEO) under accession number (GSE103313). Of 53,788 genes in the annotation, 14,527 (27% of the total in the reference) were expressed in all replicates of at least one developmental stage, with an average of 11,101 genes expressed per developmental stage. Of the embryonic samples, the 8C stage showed the highest number of expressed genes (n = 12,600).

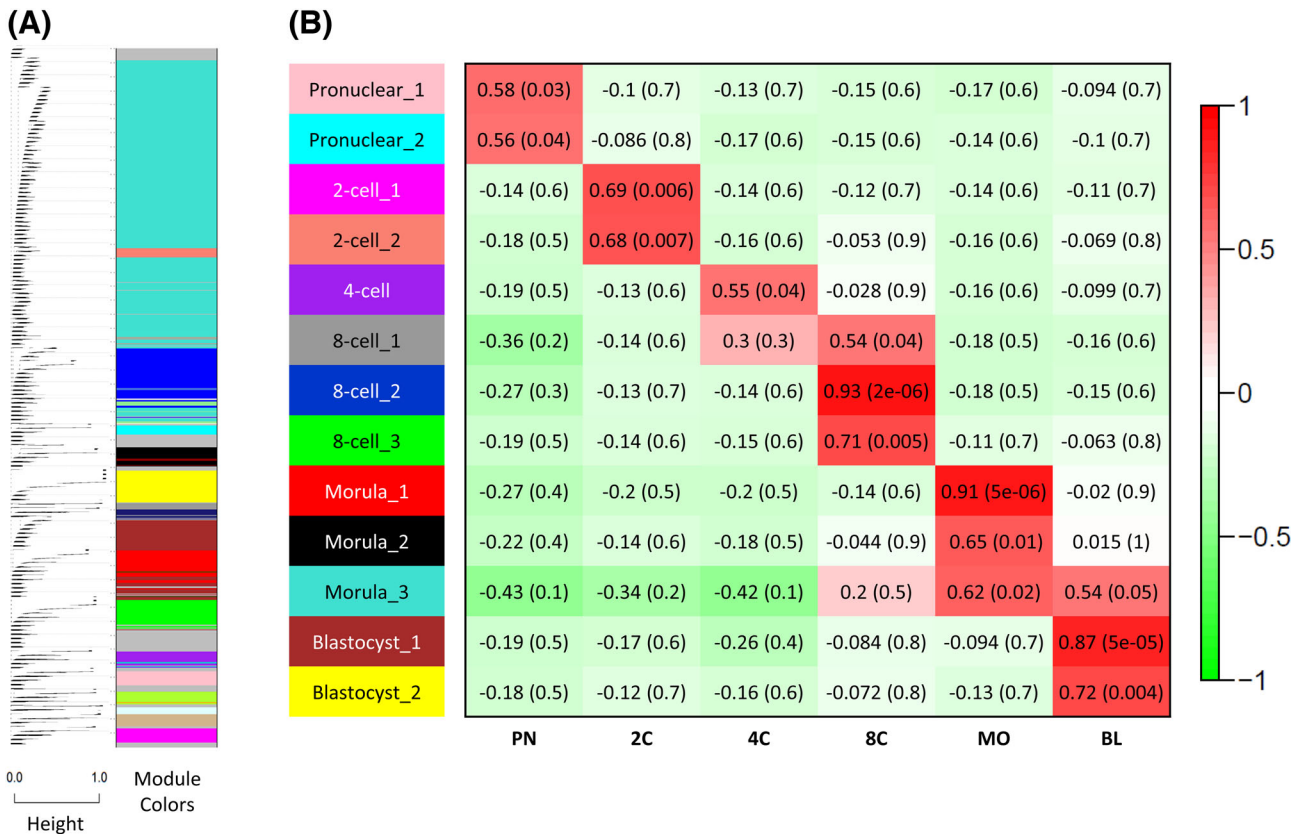

Figure 5. Clusters of genes with similar co-expression throughout multiple developmental stages were identified using WGCNA. (**A**) Nineteen unique co-expression groups, shown in clustering dendrogram, were each assigned a unique module color. (**B**) Correlation and statistical significance is shown for each co-expression group (y-axis) that was significantly correlated to a developmental stage (x-axis).

Oocytes (GV, MI, and MII) expressed an average of 11,385 genes, with 7613 genes expressed in all three stages and maintained in cleavage stage embryos. These transcripts likely correspond to maternal transcripts that are necessary to sustain development up until EGA. Most transcripts expressed uniquely during embryonic stages (PN to BL in at least one sample per group, n = 2823) started expression at the 8C stage (42%), while 39% began expression at morula (MO) or BL stage, and 18% started expression between the PN and 4-cell (4C) stages. These data indicate that the most significant increase in diversity of gene expression begins at the 8C stage and continues through to the BL stage, while the pre-EGA transcriptome is approximately half as diverse.

Relation of stage-specific transcriptome profile to developmental phases

Principal component analysis (PCA) of global transcript levels segregated different developmental stages into unique clusters (Figure 1A). The three stages representing oocytes (GV, MI, MII) formed one distinct cluster while pre-EGA embryos (PN, 2C, 4C) formed another. These two clusters were more proximal to each other than the later stage embryos, forming two subclusters of a larger pre-EGA group. Embryos at the 8C stage formed a unique cluster, demonstrating the unique transcriptional environment of what is most likely major EGA. Morula and blastocyst samples formed a fourth cluster that was closer to the 8C embryos than any others, but which was considerably less tightly grouped together, suggesting that post-EGA transcriptomes may exhibit greater variation than other stages of preimplantation development. Unsupervised hierarchical clustering of expression profiles supported the PCA and revealed clusters with similar relationships to one another (Figure 1B).

Differential gene expression indicates major embryonic genome activation at the 8-cell stage

DEGs were identified through all possible pairwise comparisons between the sample groups (Figure 2A). The incidence of DEG was most prominent between 4C and 8C embryos, with 3063 genes upregulated and 2635 downregulated (Figure 2B), suggesting that EGA in the macaque is at the 8C stage. Significant up- and downregulation of transcript levels were also observed when comparing maturing oocytes and zygotes, specifically the MI-MII and MII-PN transitions, which likely reflects changes in the transcriptome resulting from oocyte maturation and fertilization. Because the oocyte is mostly transcriptionally quiescent prior to major EGA, the changes in gene expression observed during these stages likely reflect post-transcriptional regulation of transcript polyadenylation and selective degradation of maternal transcripts [32,33].

Comparisons across nonsequential stages identified large shifts in the transcriptome from maternal to embryonic stage, with large number of DE genes between MII-stage oocytes to 8C, MO, and BL stage embryos (8965, 9635, and 9100 DEG, respectively). Also, many genes showed up- and downregulation between MII samples and PN, 2C, and 4C embryos (average of 2073 upregulated and 1614 downregulated genes). This activity could correspond to minor EGA as reported in mice [34] and cattle [35], but ascertaining

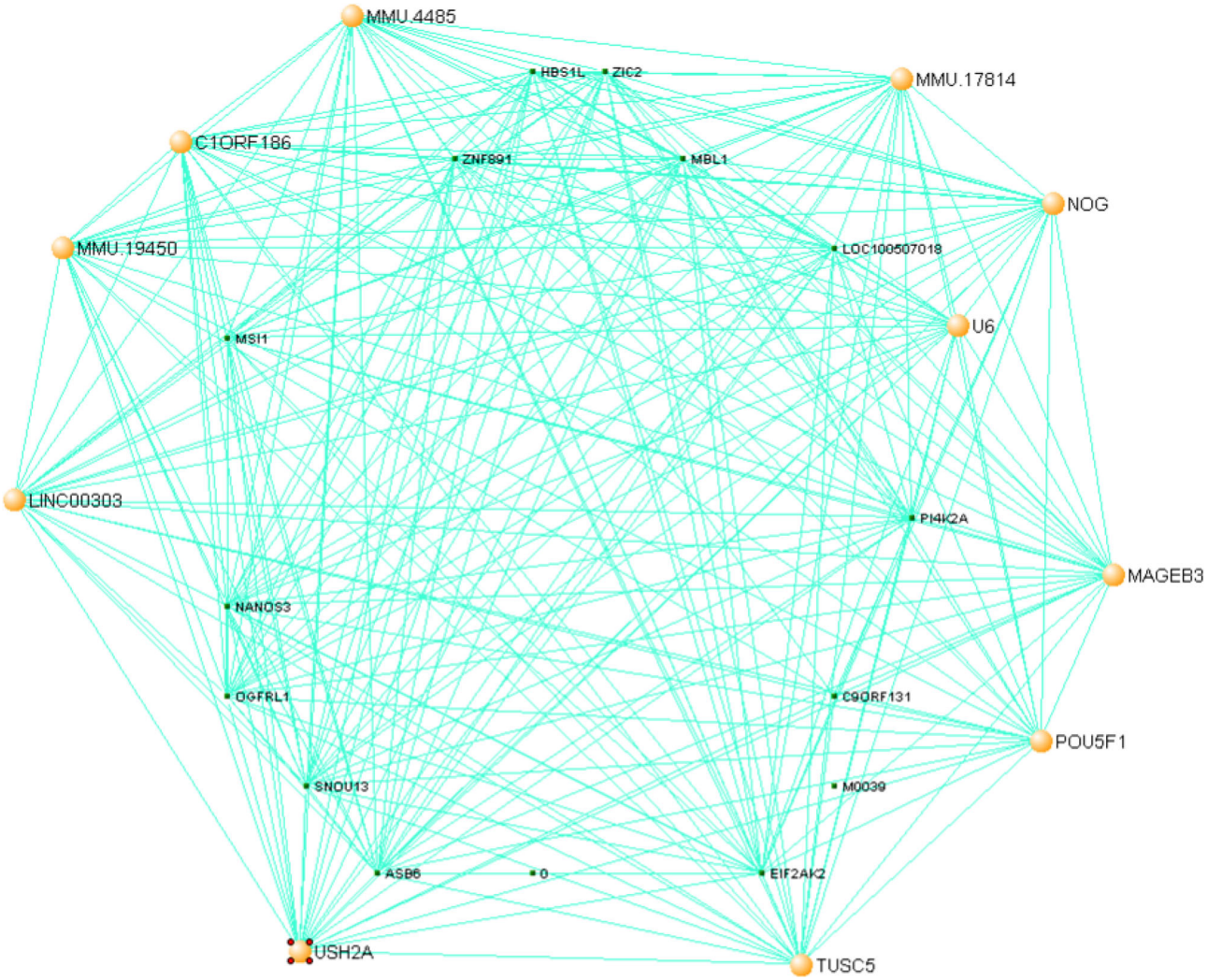

Figure 6. Gene network visualization of a morula-specific gene expression module. VisANT was used to graph the cluster of genes most highly correlated with the morula stage (Morula_1, $P = 5e\text{-}06$, $R = 0.91$). The 30 genes with the highest connectivity within the network (kME) are charted, and hub genes within the network (kME > 0.95) are highlighted with yellow dots.

if these changes are related to onset of a possible minor EGA is confounded by post-transcriptional regulation of poly(A) tails and the use of a poly(A) amplification method in the workflow that precludes assessing whether genes are being regulated at the expression or post-transcriptional level. Furthermore, use of oligo-dT priming in these experiments means that any maternal RNA in a de-adenylated state, or transcripts with poly(A) tails of limited length, would not be captured in the transcriptome profile.

The proportion of transcribed intronic sequences increases at the 8-cell stage

Transcription of genes initially includes intronic sequences that are spliced from the mature transcript during processing. Increases in alignment of transcript reads to introns could indicate higher levels of transcription [36] between developmental stages due to the increased presence of de novo transcripts yet to undergo splicing. The proportion of reads aligning uniquely to intronic sequence increased significantly between the 4C and 8C stages, indicating a probable spike in transcriptional activity coinciding with major EGA (Figure 3). Transcripts with higher intron content at the 8C stage were enriched for biological processes related to transcription and post-transcriptional regulation, DNA binding, and spliceosome and eukaryotic translation initiation factor 4 (*EIF4*) (Table 1). Together, these enrichment terms strongly indicate that the onset of regulatory programs facilitating large-scale transcription occurs at the 8C stage.

Genes form co-expression clusters specific to developmental stage and process

Transcript expression levels were analyzed for co-expression throughout early development using two forms of clustering analysis. A SOTA was used to identify clusters based on similarity of changes in gene expression from PN to BL (Figure 4). Each cluster was assigned to a broad developmental phase depending on the stage-specific pattern of expression shown: 8C elevated (1 cluster, 643 genes), 8C, MO, and BL elevated, or "EGA" (3 clusters, 1364 genes), MO and BL elevated (1 cluster, 252 genes), MO-elevated (3 clusters, 1095 genes), decreasing at 8C (5 clusters, 2053 genes), and decreasing post-EGA (1 cluster, 274 genes).

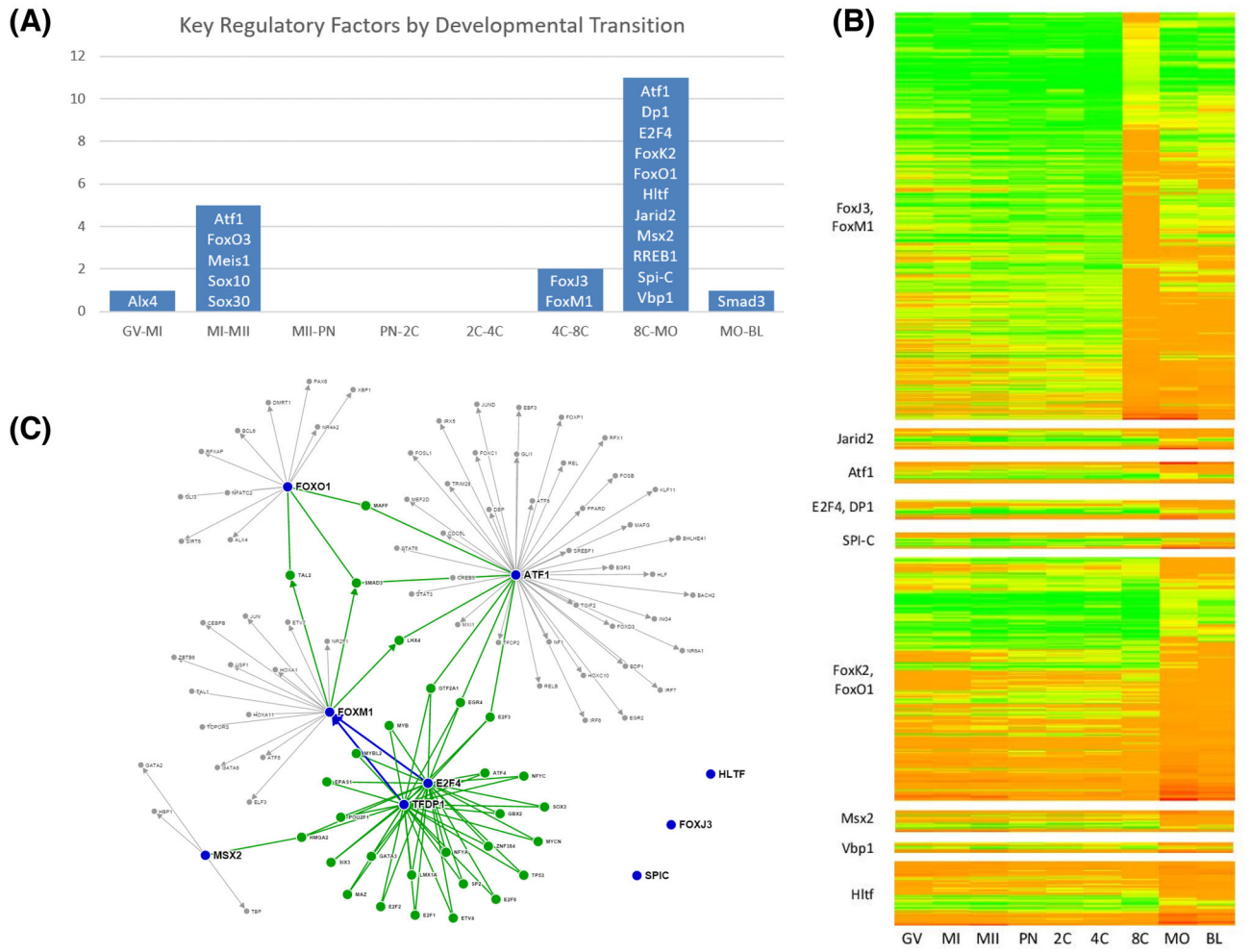

Figure 7. The majority of key TFs identified were associated with the 8-cell stage. (**A**) An increased abundance of regulatory factors was observed at EGA (4C-8C and 8C-MO transitions); many of these TFs have been implicated in early embryonic development and EGA. (**B**) Normalized expression of upregulated DEGs (4C-8C and 8C-MO pairwise comparisons; padj < 0.05 and FC > 8) containing the TFBS of key regulatory factors within their promoters. (**C**) Key regulatory factors of macaque EGA are shown in the context of the regulatory network established for human embryonic stem cells (Homo Sapiens v1.2 cell type hESCT0). Interactome includes interactions between key TFs (blue) and targets (green if regulated by multiple key TFs; grey if regulated by one). Network was generated at www.regulatorynetworks.org [62,63].

A cluster of genes with increased expression only at the 8C stage had enrichment for embryonic development and RNA binding ontology terms. The three clusters showing increased expression at the 8C, MO, and BL stages were enriched for DNA-dependent transcription, translation, and ribonucleoprotein complex related GO terms. MO and BL gene clusters showing increased expression in either one or both stages were enriched for nucleotide processing (RNA processing, ncRNA processing, nucleotide binding, mRNA binding), translation and protein metabolism (translation and translational elongation, ribonucleoprotein complex biogenesis, cellular protein metabolic processes, translation initiation factor activity), and energy metabolism (regulation of cellular protein metabolic process, ATPase activity, respiratory electron chain transport chain). Together, genes belonging to 8C-elevated and "EGA" clusters strongly relate to an increase in transcription and metabolism associated with EGA and the formation of the morula and blastocyst. Clusters exhibiting pronounced decreases in transcript levels starting at the 8C stage (5 of 16 clusters) cumulatively accounted for the majority of genes included in this analysis. These clusters were functionally enriched primarily for gross cellular functions (maintenance of location in cell, protein localization) and regulation of cell growth. These results suggest a drastic shift in the developing transcriptome as embryonic genes begin expression and maternal transcripts undergo degradation.

WGCNA [28,37] was performed to further understand gene relationships and identify potential regulators of major co-expression clusters. A total of 19 expression clusters were identified (Figure 5A), with 13 showing a statistically significant correlation to a single developmental stage between the PN and BL (Figure 5B). The number of genes within clusters varied widely, between 34 and 10,579. The cluster correlated most highly with the ML stage (Morula_1) contains 815 genes (average number per cluster is 1369). The gene *POU5F1* (also known as *OCT4*) was one of 16 hub genes for this cluster, demonstrating the presence and importance of this well-known regulator of development and lineage specification at the ML stage (Figure 6). Expression of *CDX2* was closely tied to this cluster as well. Another gene associated with lineage differentiation and maintenance of pluripotency (*SOX2*) was a hub gene within a cluster specific to the blastocyst (Blastocyst_2). In that same cluster,

FGF4 was found to be a hub gene as well. Transcripts for *KLF4*, *LIN28*, and *SALL4* were found in a cluster correlated to morulae and strongly associated with blastocysts (Morula_3). Importantly, *Polymerase II A* (*POLR2A*) was also found in Morula_3, suggesting that this gene is most strongly correlated to genes expressed post-EGA.

The proportion of key embryonic genome activation regulatory factors increases at 8-cell stage

Considering that clustering analyses depend almost singularly on gene co-expression to determine "hub" genes, we sought to identify key regulatory factors for transitions between developmental states using the TFBS method [38], which involves promoter analysis of DEG for TFBS of known transcription factors. Using the MatInspector tool (available at genomatix.de), we obtained a list of 70,740 macaque promoters, belonging to 16,264 genes sequences orthologously conserved between the macaque and human genomes. We also obtained a list of TFBS, corresponding to 165 known macaque transcription factor families found within these promoters. Of these, 675,059 high confidence TFBS with core and matrix similarities of 1, corresponding to 100% confidence, were selected for analysis. Transcription factor families with enriched TFBS in promoters of multiple upregulated DEG were identified as key regulatory factors for the given comparison of developmental stages. No key regulatory factors were identified for the MII-PN, PN-2C, or 2C-4C comparisons, suggesting that increases in transcript abundance at these early stages may not be accomplished by canonical transcriptional regulation. The lack of key factors in these stages is also expected due to the fact that pre-EGA eggs and embryos are transcriptionally quiescent. Two transcription factors were identified for the 4C-8C comparison (FOXJ3 and FOXM1), 11 for the 8C-MO comparison (ATF1, DP1, E2F4, FOXK2, FOXO1, HLTF, JARID2, MSX2, RREB1, SPI-C, and VBP1), and 1 for the MO-BL comparison (SMAD3) (Figure 7A), indicating their potential role in regulating transcriptional programs at these stages. Putative targets of the key regulatory factors identified for the 4C-8C and 8C-MO transitions (Figure 7C) were functionally enriched for terms related to epigenetic regulation, such as acetylation and transferases (Table 2).

Using a pre-established human embryonic stem cell (ESC) regulatory network, possible interactions between transcription factors identified at the 8C stage were analyzed (Figure 7B), revealing strong interconnectivity—particularly between ATF1, DP1, E2F4, FOXM1, and FOXO1.

Discussion

Comprehensive cataloging and characterization of the early mammalian transcriptome help elucidate some of the crucial biological processes that take place during this critical developmental period. Other studies have successfully addressed this question in humans and mice using next generation sequencing technology [39,40], but until now the only characterization of macaque oocyte and early embryo transcriptomes has been conducted with smaller-scale and pathway-oriented RT-PCR [41] or quantitative assay by dot blotting [42–44]. The global perspective of the transcriptome given by RNA-seq provides valuable insights into molecular mechanisms and serves as a useful adjunct to pre-existing transcriptome databases for the rhesus macaque that do not include NGS sequence from early embryonic tissues [25].

Table 2. Gene ontology and SP-PIR term enrichment for differentially upregulated genes (4C-8C and 8C-MO) with putative TFBS of key regulatory factors in their promoters.

	Term	Gene count	Adjusted *P*-value
Gene ontology	Mitochondrion	97	4.10E-10
	Mitochondrial part	61	4.10E-08
	Membrane-enclosed lumen	130	2.10E-07
	Organelle lumen	125	1.30E-06
	Intracellular organelle lumen	121	3.40E-06
	Nucleolus	56	3.10E-04
	Ribonucleoprotein complex	45	3.40E-04
	Mitochondrial envelope	38	6.60E-04
	Envelope	50	7.10E-04
	Organelle envelope	50	7.40E-04
	RNA binding	54	2.50E-03
	Nuclear lumen	90	3.80E-03
	Mitochondrial membrane	34	4.10E-03
	Translation	35	6.70E-03
	Organelle membrane	68	2.50E-02
SP-PIR	Mitochondrion	72	2.00E-07
	Transit peptide	41	4.20E-04
	Acetylation	148	4.40E-04
	Transferase	87	1.10E-03
	RNA binding	42	2.50E-03
	Nucleus	205	2.70E-02
	Transport	91	3.50E-02

Defining the time of EGA in the macaque has largely been accomplished by examining the onset of nucleolar transcription [3,45], which places macaque EGA at the 6–8 cell stage, agreeing closely with the 4–8 cell stage timing reported in humans [14]. Although it is important to note that earlier studies examining nucleogenesis and uridine incorporation indicated that human and bovine EGA occurs at the 8–16 cell stages [46,47], EGA can be characterized in any number of ways at different time points. We define EGA as the onset of transcription from a large number of genes at high levels.

The use of oligo(dT) amplification of cDNA for library preparation in this study should be taken into consideration as it mostly encompasses mature transcripts. Many other RNA species with regulatory roles such as miRNA [48,49] would be unaccounted for with our sequencing methodology. Also, due to the unfeasibility of acquiring in vivo embryos for this study, in vitro maturation (IVM) oocytes and in vitro fertilization (IVF) embryos were used. It is well established that in vitro culture methods can have influences on gene expression that could result in over-representation of certain molecular pathways (e.g. stress response) and the culture methodology should be considered when examining this data. Those considerations notwithstanding, our data reflect the timing of EGA in the macaque through several findings: (1) a significant increase in the number of differentially expressed genes (both up- and downregulated) between the 4C and 8C stages; (2) an increased proportion of intron alignment beginning at the 8C stage, indicating global increases in pre-spliced mRNA; (3) an increase in the number of expressed key regulatory factors at the 8C stage; and (4) transcription GO enrichment within gene clusters associated with the 8C stage.

The significant correlation of three gene co-expression modules to the 8C stage revealed the presence of many biological processes and molecular functions that can be attributed to large-scale transcription. Indeed, the two most significant terms enriched in the "8-cell_3" module were for transcription regulation and transcription

associated with zinc ions. Following that, GO terms relating to embryonic development were enriched in the morula- and blastocyst-specific expression module ("Morula_3") and in utero embryonic development was enriched in "Morula_2." In a blastocyst-specific module ("Blastocyst_2"), enriched GO terms included differentiation events, such as tube development and appendage morphogenesis. The overall processes depicted by the cluster genes and enrichment terms indicate the onset of EGA and beginning of lineage differentiation as the morula and blastocyst form.

The majority of key regulatory factors that were identified using the TFBS method are associated with the 8C stage, further indicating that EGA occurs at this stage of development in the macaque. The key TFs associated with the 8C stage are highly interconnected within the regulatory network established for human ESC. The interactions between DP1 and E2F4 are particularly striking. DP1 regulates the cell cycle through dimerization with E2F family members and transactivation of E2F target genes required for DNA replication [50,51]. Within the E2F family, E2F4 is a master regulator of histone genes and likely plays a role in histone biogenesis [52]. In addition, since many E2F-bound regions in vivo do not contain a binding motif [53], it is possible that E2F4, and other E2F family members, may play an even greater role in EGA than we were able to detect.

The other highly connected nodes of this regulatory network—ATF1, FOXM1, and FOXO1—have all been strongly associated with the pluripotent state in other mammals. ATF1 has been identified as a key regulator of EGA in the mouse [54], and FOXO1 has been shown to activate transcription of *SOX2* and *OCT4* in vivo [55]. Overexpression of *FOXM1* in human fibroblasts stimulates expression of *SOX2*, *NANOG*, and *OCT4* [56], while knockdown of *FoxM1* in murine ESC decreases expression of *Nanog*, *Oct4*, *Utf1*, and *Esrrb* [57]. Altogether, the components of this network indicate functions related to cell cycle regulation, DNA replication, proliferation, and expression of canonical pluripotency factors.

The roles of FOXJ3, HLTF, and SPI-C in this network remain more unclear, and may indicate a degree of divergence between the regulatory programs governing EGA in the macaque as compared to humans. In the mouse, Spi-C regulates histone modifications such as H3K9me2 [58], and has been linked to expression of eIF-1a and Oct4 in zygotes [59]. FoxJ3 appears necessary for proliferation, and may regulate an uncharacterized group of zinc finger proteins [60].

Generally, key TFs associated with the 8C stage are enriched for functions relevant to EGA (cell cycle regulation, DNA replication, proliferation, and expression of canonical pluripotency factors such as SOX2, OCT4, and NANOG), suggesting that these TFs may be integral to macaque preimplantation development. Overall, these data provide a useful preliminary regulatory network that may be interrogated in the future to establish functional relationships, and determine which of these key regulatory factors are crucial for successful preimplantation development.

The data presented here constitute a useful atlas of spatiotemporal gene expression throughout preimplantation development in the macaque. Important trends that take place during the time points analyzed (fertilization, EGA, morula and blastocyst development) are captured and described through gene co-expression and network interconnectivity. Additionally, these data will serve as a valuable comparison to molecular characterizations of human developmental stages, given the vast array of studies previously conducted regarding macaque embryology and reproductive physiology as models of human reproduction [16]. In particular, these data may be applied to studies of evolutionary divergence [61] at a biological time where the most basal differences between species emerge. Finally, the presence of a number of uncharacterized hub genes within stage-specific expression modules presents an opportunity for discovery of additional players in these processes, improving our canonical understanding of early development and providing new targets for markers of developmental competence or new possibilities for improving assisted reproductive technologies in humans and other species.

Acknowledgments

Dr Chuan-Yun Li's assistance in obtaining an annotation file of Rhesusbase transcripts is much appreciated.

References

1. Lee YS, Vandevoort CA, Gaughan JP, Midic U, Obradovic Z, Latham KE. Extensive effects of in vitro oocyte maturation on rhesus monkey cumulus cell transcriptome. *Am J Physiol: Endocrinol Metab* 2011; **301**:E196–E209.
2. Schultz RM. The molecular foundations of the maternal to zygotic transition in the preimplantation embryo. *Hum Reprod Update* 2002; **8**:323–331.
3. Schramm RD. Causes of developmental failure of in-vitro matured rhesus monkey oocytes: impairments in embryonic genome activation. *Hum Reprod* 2003; **18**:826–833.
4. Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Scholer H, Smith A. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* 1998; **95**:379–391.
5. Scholer HR, Hatzopoulos AK, Balling R, Suzuki N, Gruss P. A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. *EMBO J* 1989; **8**:2543–2550.
6. Niwa H, Toyooka Y, Shimosato D, Strumpf D, Takahashi K, Yagi R, Rossant J. Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* 2005; **123**:917–929.
7. Cross JC. Genes regulating embryonic and fetal survival. *Theriogenology* 2001; **55**:193–207.
8. Copp AJ. Death before birth: clues from gene knockouts and mutations. *Trends Genet* 1995; **11**:87–93.
9. Wadhwa P, Buss C, Entringer S, Swanson J. Developmental origins of health and disease: brief history of the approach and current focus on epigenetic mechanisms. *Semin Reprod Med* 2009; **27**:358–368.
10. Aagaard-Tillery KM, Grove K, Bishop J, Ke X, Fu Q, Mcknight R, Lane RH. Developmental origins of disease and determinants of chromatin structure: maternal diet modifies the primate fetal epigenome. *J Mol Endocrinol* 2008; **41**:91–102.
11. Cui X, Li X, Yin X, Kong IK, Kang J, Kim N. Maternal gene transcription in mouse oocytes: genes implicated in oocyte maturation and fertilization. *J Reprod Dev* 2007; **53**:405–418.
12. Flach G, Johnson MH, Braude PR, Taylor RA, Bolton VN. The transition from maternal to embryonic control in the 2-cell mouse embryo. *EMBO J* 1982; **1**:681–686.
13. Misirlioglu M, Page GP, Sagirkaya H, Kaya A, Parrish JJ, First NL, Memili E. Dynamics of global transcriptome in bovine matured oocytes and preimplantation embryos. *Proc Natl Acad Sci USA* 2006; **103**:18905–18910.
14. Braude P, Bolton V, Moore S. Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature* 1988; **332**:459–461.
15. Canovas S, Cibelli JB, Ross PJ. Jumonji domain-containing protein 3 regulates histone 3 lysine 27 methylation during bovine preimplantation development. *P Natl Acad Sci USA* 2012; **109**:2400–2405.
16. Wilson EK. Modeling man: the monkey colony at the Carnegie Institution of Washington's Department of Embryology, 1925–1971. *J Hist Biol* 2012; **45**:213–251.

17. Burruel V, Klooster KL, Chitwood J, Ross PJ, Meyers SA. Oxidative damage to rhesus macaque spermatozoa results in mitotic arrest and transcript abundance changes in early embryos. *Biol Reprod* 2013;**89**(3):72. http://doi.org/10.1095/biolreprod.113.110981.
18. Sarason RL, VandeVoort CA, Mader DR, Overstreet JW. The use of nonmetal electrodes in electroejaculation of restrained but unanesthetized macaques. *J Med Primatol* 1991; **20**:122–125.
19. Burruel V, Klooster K, Barker CM, Pera RR, Meyers S. Abnormal early cleavage events predict early embryo demise: sperm oxidative stress and early abnormal cleavage. *Sci Rep* 2015; **4**:6598.
20. Wolf DP. Assisted reproductive technologies in rhesus macaques. *Reprod Biol Endocrinol* 2004; **2**:37.
21. Wolf DP, Thormahlen S, Ramsey C, Yeoman RR, Fanton J, Mitalipov S. Use of assisted reproductive technologies in the propagation of rhesus macaque offspring1. *Biol Reprod* 2004; **71**:486–493.
22. Meyers SA, Li M, Enders AC, Overstreet JW. Rhesus macaque blastocysts resulting from intracytoplasmic sperm injection of vacuum-dried spermatozoa. *J Med Primatol* 2009; **38**:310–317.
23. Schramm RD, Al-Sharhan M. Fertilization and early embryology: development of in-vitro-fertilized primate embryos into blastocysts in a chemically defined, protein-free culture medium. *Hum Reprod* 1996; **11**:1690–1697.
24. Chitwood JL, Rincon G, Kaiser GG, Medrano JF, Ross PJ. RNA-seq analysis of single bovine blastocysts. *BMC Genomics* 2013; **14**:350.
25. Zhang S, Liu C, Shi M, Kong L, Chen J, Zhou W, Zhu X, Yu P, Wang J, Yang X, Hou N, Ye Z et al. RhesusBase: a knowledgebase for the monkey research community. *Nucleic Acids Res* 2013; **41**:D892–D905.
26. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol* 2010; **11**:R106.
27. Handl J, Knowles J, Kell DB. Computational cluster validation in post-genomic data analysis. *Bioinformatics* 2005; **21**:3201–3212.
28. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; **9**:559.
29. Hu Z, Mellor J, Wu J, Delisi C. VisANT: an online visualization and analysis tool for biological interaction data. *BMC Bioinformatics* 2004; **5**:17.
30. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009; **37**:1–13.
31. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2008; **4**:44–57.
32. Reyes JM, Chitwood JL, Ross PJ. RNA-Seq profiling of single bovine oocyte transcript abundance and its modulation by cytoplasmic polyadenylation. *Mol Reprod Dev* 2015; **82**:103–114.
33. Reyes JM, Ross PJ. Cytoplasmic polyadenylation in mammalian oocyte maturation. *WIREs RNA* 2016; **7**:71–89.
34. Latham KE. Embryonic genome activation. *Front Biosci* 2001; **6**:d748–d759.
35. Kanka J, Kepkova K, Nemcova L. Gene expression during minor genome activation in preimplantation bovine development. *Theriogenology* 2009; **72**:572–583.
36. Ameur A, Zaghlool A, Halvardson J, Wetterbom A, Gyllensten U, Cavelier L, Feuk L. Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat Struct Mol Biol* 2011; **18**:1435–1440.
37. Horvath S, Dong J, Miyano S. Geometric interpretation of gene coexpression network analysis. *PLoS Comput Biol* 2008; **4**:e1000117.
38. Gu Q, Nagaraj SH, Hudson NJ, Dalrymple BP, Reverter A. Genome-wide patterns of promoter sharing and co-expression in bovine skeletal muscle. *BMC Genomics* 2011; **12**:23.
39. Xue Z, Huang K, Cai C, Cai L, Jiang C, Feng Y, Liu Z, Zeng Q, Cheng L, Sun YE, Liu J, Horvath S et al. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* 2013; **500**:593–597.
40. Tang F, Barbacioru C, Nordman E, Bao S, Lee C, Wang X, Tuch BB, Heard E, Lao K, Surani MA, Pesce M. Deterministic and stochastic allele specific gene expression in single mouse blastomeres. *PLoS One* 2011; **6**:e21208.
41. Vassena R, Dee Schramm R, Latham KE. Species-dependent expression patterns of DNA methyltransferase genes in mammalian oocytes and preimplantation embryos. *Mol Reprod Dev* 2005; **72**:430–436.
42. Latham KE. The Primate Embryo Gene Expression Resource in embryology and stem cell biology. *Reprod Fertil Dev* 2006; **18**:807–810.
43. Zheng P, Patel B, Mcmenamin M, Reddy SE, Paprocki AM, Schramm RD, Latham KE. The primate embryo gene expression resource: a novel resource to facilitate rapid analysis of gene expression patterns in non-human primate oocytes and preimplantation stage embryos1. *Biol Reprod* 2004; **70**:1411–1418.
44. Zheng P, Vassena R, Latham K. Expression and downregulation of WNT signaling pathway genes in rhesus monkey oocytes and embryos. *Mol Reprod Dev* 2006; **73**:667–677.
45. Schramm RD, Bavister BD. Onset of nucleolar and extranucleolar transcription and expression of fibrillarin in macaque embryos developing in vitro1. *Biol Reprod* 1999; **60**:721–728.
46. Tesarik J, Kopecny V, Plachot M, Mandelbaum J. High-resolution autoradiographic localization of DNA-containing sites and RNA synthesis in developing nucleoli of human preimplantation embryos: a new concept of embryonic nucleologenesis. *Development* 1987; **101**:777–791.
47. Camous S, Kopecny V, Flechon JE. Autoradiographic detection of the earliest stage of [3H]-uridine incorporation into the cow embryo. *Biol Cell* 1986; **58**:195–200.
48. Mondou E, Dufort I, Gohin M, Fournier E, Sirard M-A. Analysis of microRNAs and their precursors in bovine early embryonic development. *Mol Hum Reprod* 2012; **18**:425–434.
49. Mineno J. The expression profile of microRNAs in mouse embryos. *Nucleic Acids Res* 2006; **34**:1765–1771.
50. DeGregori J, Leone G, Miron A, Jakoi L, Nevins JR. Distinct roles for E2F proteins in cell growth control and apoptosis. *Proc Natl Acad Sci USA* 1997; **94**:7245–7250.
51. Muller H. E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis. *Genes Dev* 2001; **15**:267–285.
52. Gokhman D, Livyatan I, Sailaja BS, Melcer S, Meshorer E. Multilayered chromatin analysis reveals E2f, Smad and Zfx as transcriptional regulators of histones. *Nat Struct Mol Biol* 2013; **20**:119–126.
53. Bieda M. Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome. *Genome Res* 2006; **16**:595–605.
54. Jin XL, O'Neill C. The regulation of the expression and activation of the essential ATF1 transcription factor in the mouse preimplantation embryo. *Reproduction* 2014; **148**:147–157.
55. Zhang X, Yalcin S, Lee D, Yeh TJ, Lee S, Su J, Mungamuri SK, Rimmele P, Kennedy M, Sellers R, Landthaler M, Tuschl T et al. FOXO1 is an essential regulator of pluripotency in human embryonic stem cells. *Nat Cell Biol* 2011; **13**:1092–1099.
56. Xie Z, Tan G, Ding M, Dong D, Chen T, Meng X, Huang X, Tan Y. Foxm1 transcription factor is required for maintenance of pluripotency of P19 embryonal carcinoma cells. *Nucleic Acids Res* 2010; **38**:8027–8038.
57. Tan G, Cheng L, Chen T, Yu L, Tan Y, Singh SR. Foxm1 mediates LIF/Stat3-dependent self-renewal in mouse embryonic stem cells and is essential for the generation of induced pluripotent stem cells. *PLoS One* 2014; **9**:e92304.
58. Schweitzer BL, Huang KJ, Kamath MB, Emelyanov AV, Birshtein BK, Dekoter RP. Spi-C has opposing effects to PU.1 on gene expression in progenitor B cells. *J Immunol* 2006; **177**:2195–2207.
59. Kageyama S, Liu H, Nagata M, Aoki F. The role of ETS transcription factors in transcription and development of mouse preimplantation embryos. *Biochem Biophys Res Commun* 2006; **344**:675–679.
60. Grant GD, Gamsby J, Martyanov V, Brooks L, George LK, Mahoney JM, Loros JJ, Dunlap JC, Whitfield ML. Live-cell monitoring of periodic gene expression in synchronous human cells identifies Forkhead genes involved in cell cycle control. *Mol Biol Cell* 2012; **23**:3079–3093.

61. Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, Remington KA, Strausberg RL, Venter JC, Wilson RK, Batzer MA, Bustamante CD et al. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 2007; **316**:222–234.
62. Neph S, Stergachis AB, Reynolds A, Sandstrom R, Borenstein E, Stamatoyannopoulos JA. Circuitry and dynamics of human transcription factor regulatory networks. *Cell* 2012; **150**:1274–1286.
63. Stergachis AB, Neph S, Sandstrom R, Haugen E, Reynolds AP, Zhang M, Byron R, Canfield T, Stelhing-Sun S, Lee K, Thurman RE, Vong S et al. Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature* 2014; **515**:365–370.