doi:10.1038/nature12364

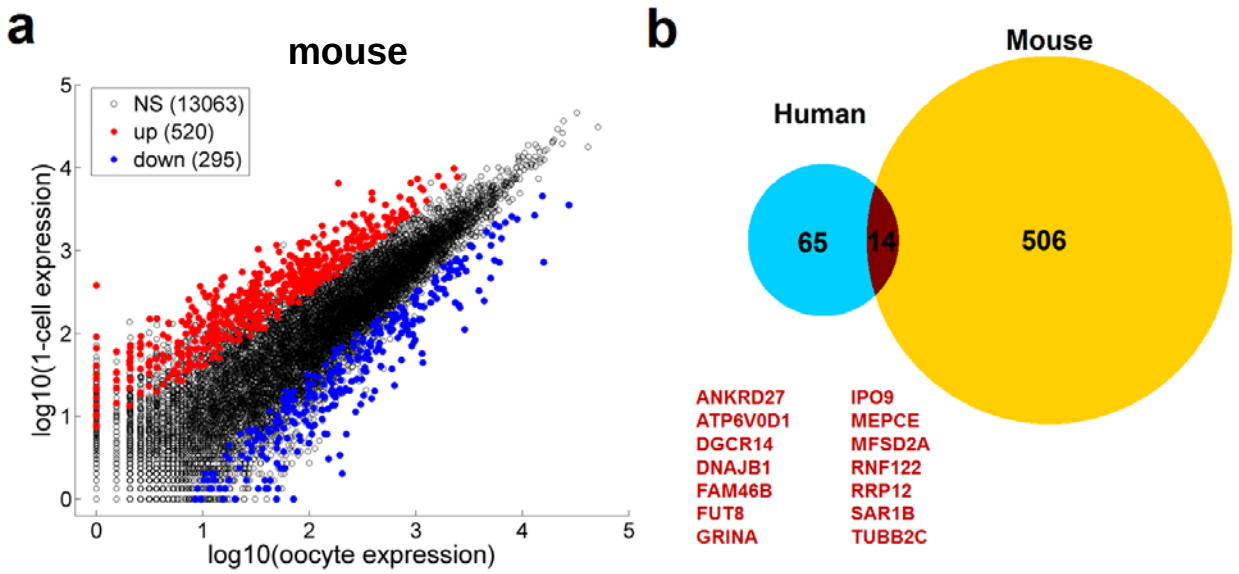
Supplementary text

Collectively, we were able to detect ~14,000 expressed genes with RPKM (reads per kilobase per million) > 1 or ~16,000 with RPKM > 0.1 in at least one cell type from oocyte to the morula stage (Supplementary Table 1). We use a cutoff RPKM > 0.5 in the current analysis, which is equivalent to RPM (reads per million) > 0.05 (both metrics have been used in the literature[16,25]). In practice, both metrics provide the same exact information when comparing between samples because transcript length is constant.

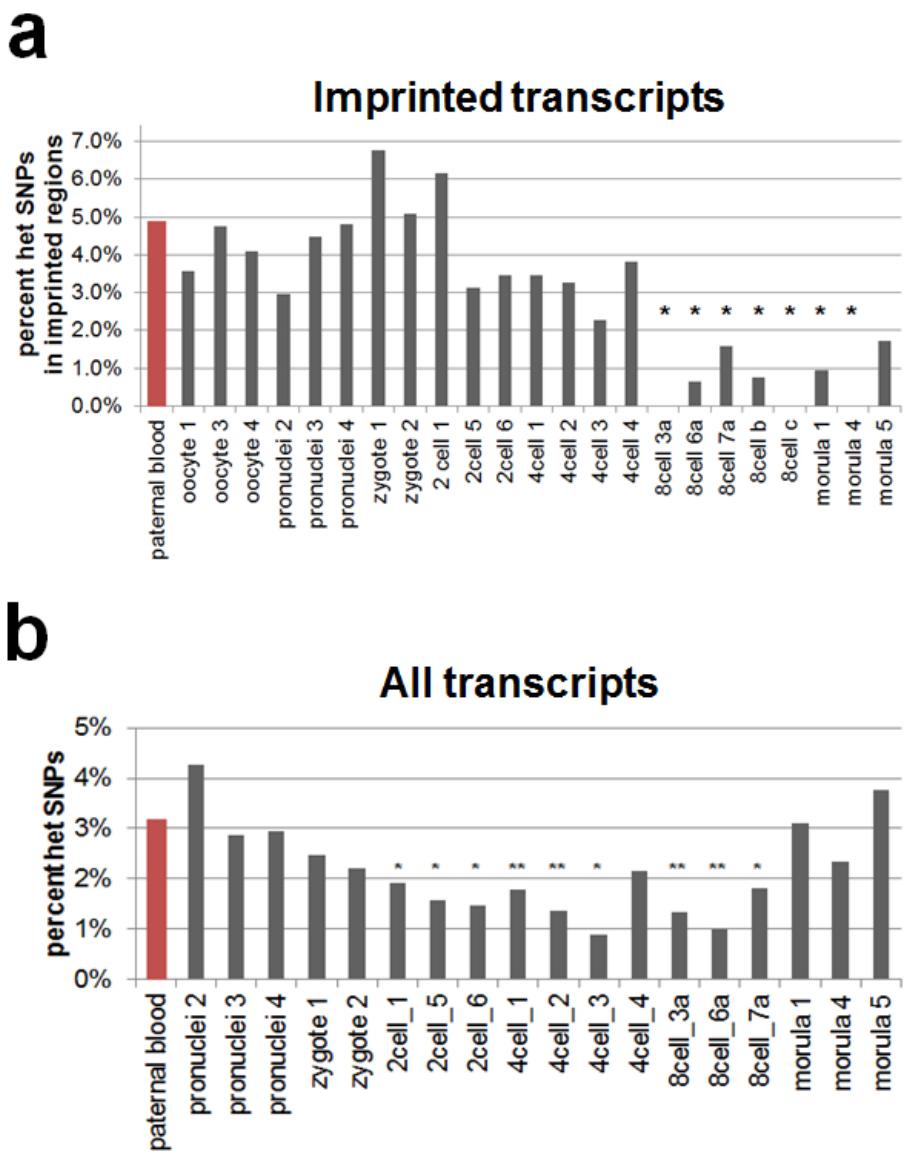
Pairwise-comparisons between successive stages of human preimplantation transcriptomes revealed gene activation occurring as early as the 1-cell stage, as discussed in the main text. Further comparisons revealed a relatively transcriptional quiescence in the 1-2 and 2-4 cell divisions, but massive changes between the 4- to 8-cell transitions, consistent with the previously characterized major EGA period[38-40]. Of the 3,151 altered genes between 4- and 8-cell embryos, 1,399 and 1,752 are up- and down-regulated in 8-cell embryos (Supplementary Fig. 4a). Although the 8-cell to morula stage embryos also exhibited large transcriptional changes, only half as many genes were altered (n=1,633). Consistently, mouse ZGA was most pronounced in the 1-2 cell transition, but subsequent genome activation were progressively less dramatic (Supplementary Fig. 4b). Together, this result indicates the initial wave of EGA (or ZGA) has the greatest magnitude of transcriptional changes.

We analyzed the status of imprinting genes that exhibit known parent-of-origin expression patterns[41,42]. Based on the presence of heterozygous SNPs, we found that many known imprinted genes (on average 15 detected per single cell) show bi-allelic expression in oocytes, likely contributed by the presence of maternally-stored transcripts deposited during oogenesis. In contrast, 8-cell embryos exhibit significantly depleted (p<0.05, binomial test)

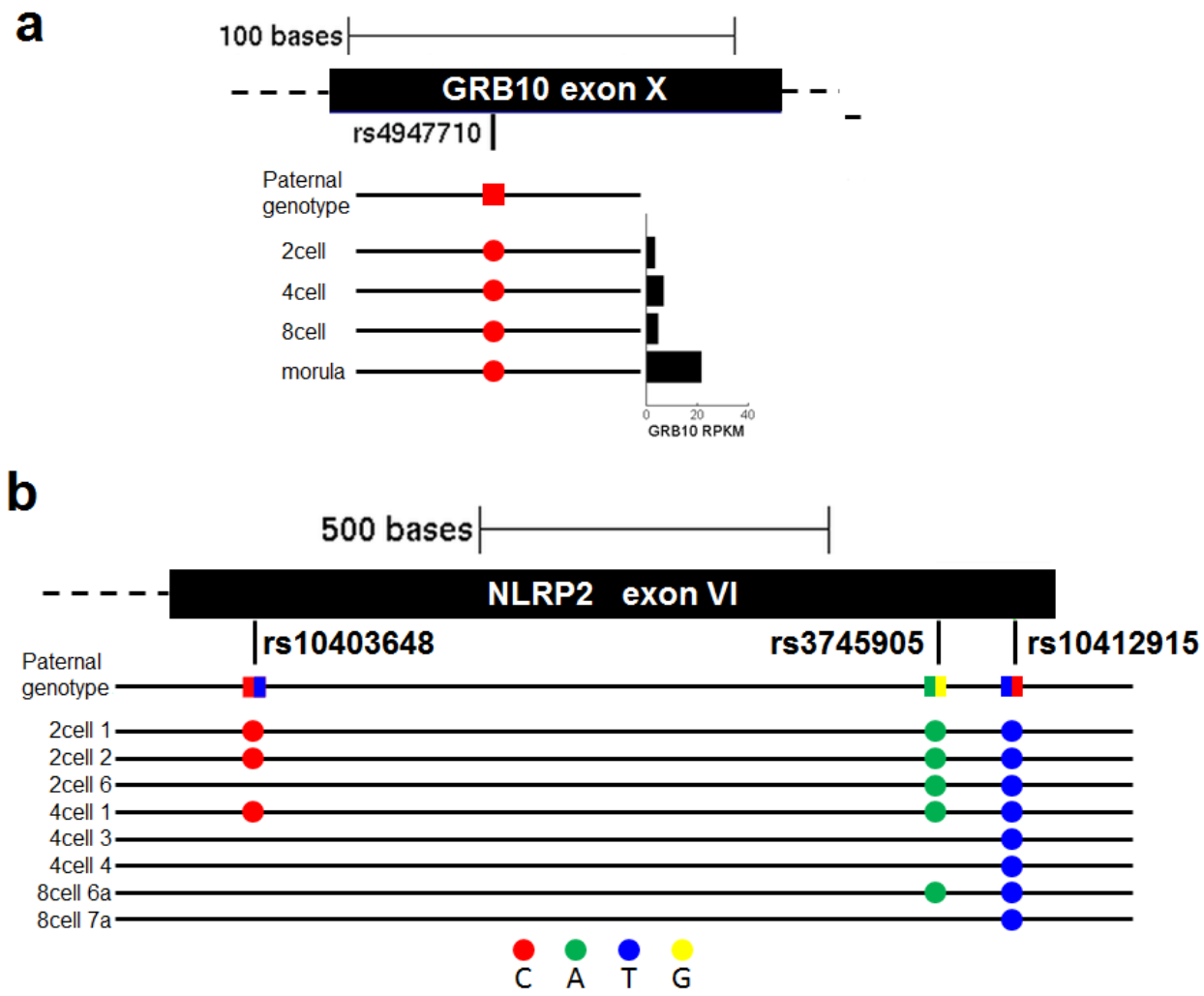
heterozygous SNPs, suggesting complete decay of maternally-stored imprinted gene transcripts (Supplementary Fig. 2a). In fact, global maternal transcript degradation was observed as early as the 2-cell stage (Supplementary Fig. 2b). Next, we monitored the activation of paternally expressed (maternally silenced) genes during each preimplantation stage by again leveraging paternal SNP patterns. For example, within the paternally expressed GRB10 gene locus, all embryos express the minor variant allele inherited from the father who is homozygous for the minor allele (Supplementary Fig. 3a). This gene shows 4-fold upregulation at the morula stage, indicating that this imprinted gene undergoes paternal-allelic specific activation.



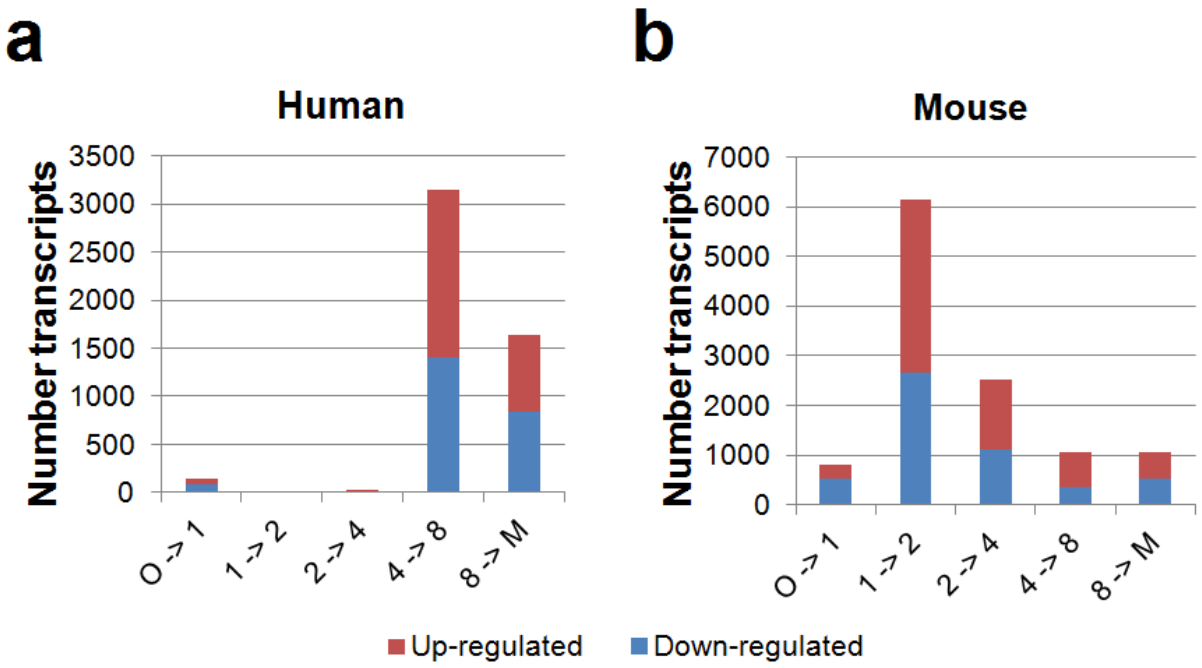
Supplementary Fig. 1. Differentially gene expression between mouse oocyte and pronuclei. a) scatterplot of mean expression in mouse oocyte (n=2) and 1-cell pronuclei (n=3). Significantly up- or down-regulated genes were determined by FDR < 5% and minimum 2-fold change. b) Venn diagram showing a significant overlap of genes between minor ZGA genes in human and mouse ($p < 10^{-5}$, hypergeometric test). Overlapped gene names are listed in red.



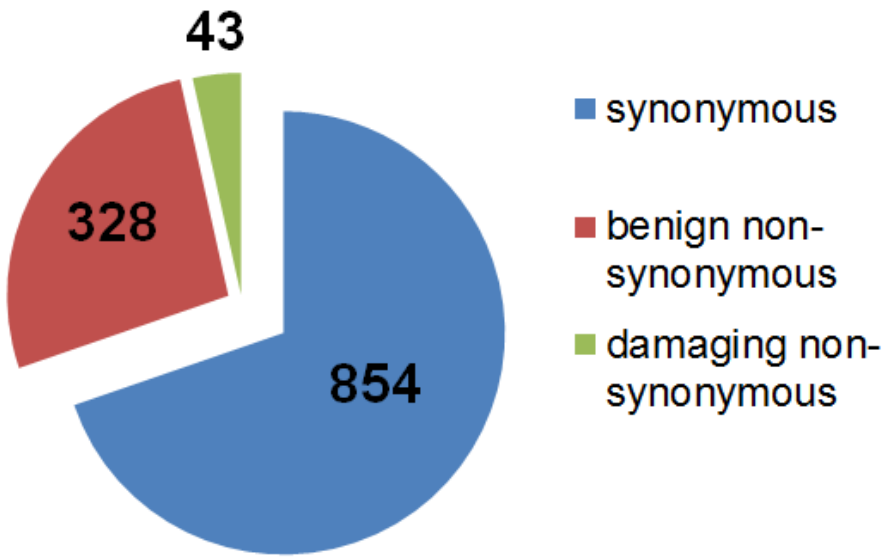
Supplementary Fig. 2. Distribution of heterozygous SNPs in human embryos for all developmental stages. Bar graph showing the proportion of heterozygous SNPs found within a) imprinted transcripts (on average 15 per cell) or b) all transcripts (n=14,765). Exome-seq of paternal blood was used as an expected rate of heterozygous SNPs. Embryos with significantly depleted heterozygous SNPs are demarked by * ($p< 0.05$) and ** ($p< 0.01$, binomial test).



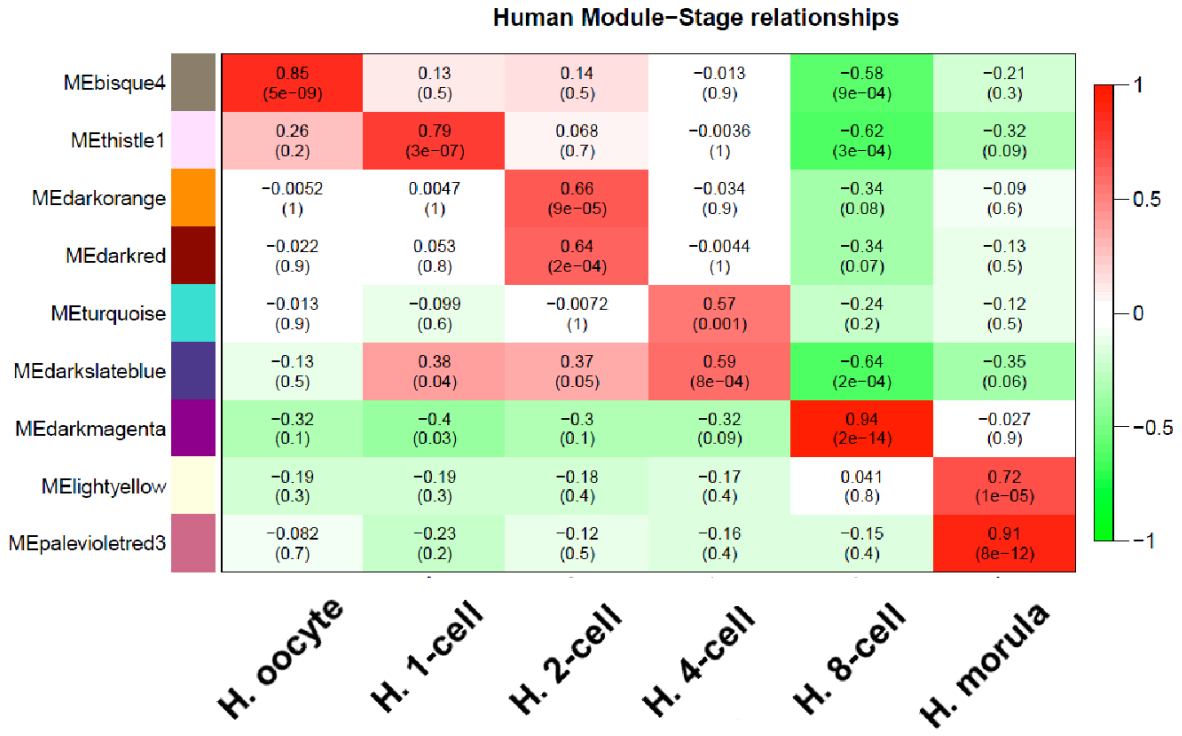
Supplementary Fig. 3. Example of paternally (a) and maternally (b) expressed human imprinting gene. a) GRB10 is a known paternally expressed gene. Reads from the same embryo were pooled and RPKM represents the average expression from pooled embryos. Only 1 morula embryo could detect SNPs at this locus. B) NLRP2 is a known maternally expressed gene and the above observed pattern may represent a maternally phased SNP block.



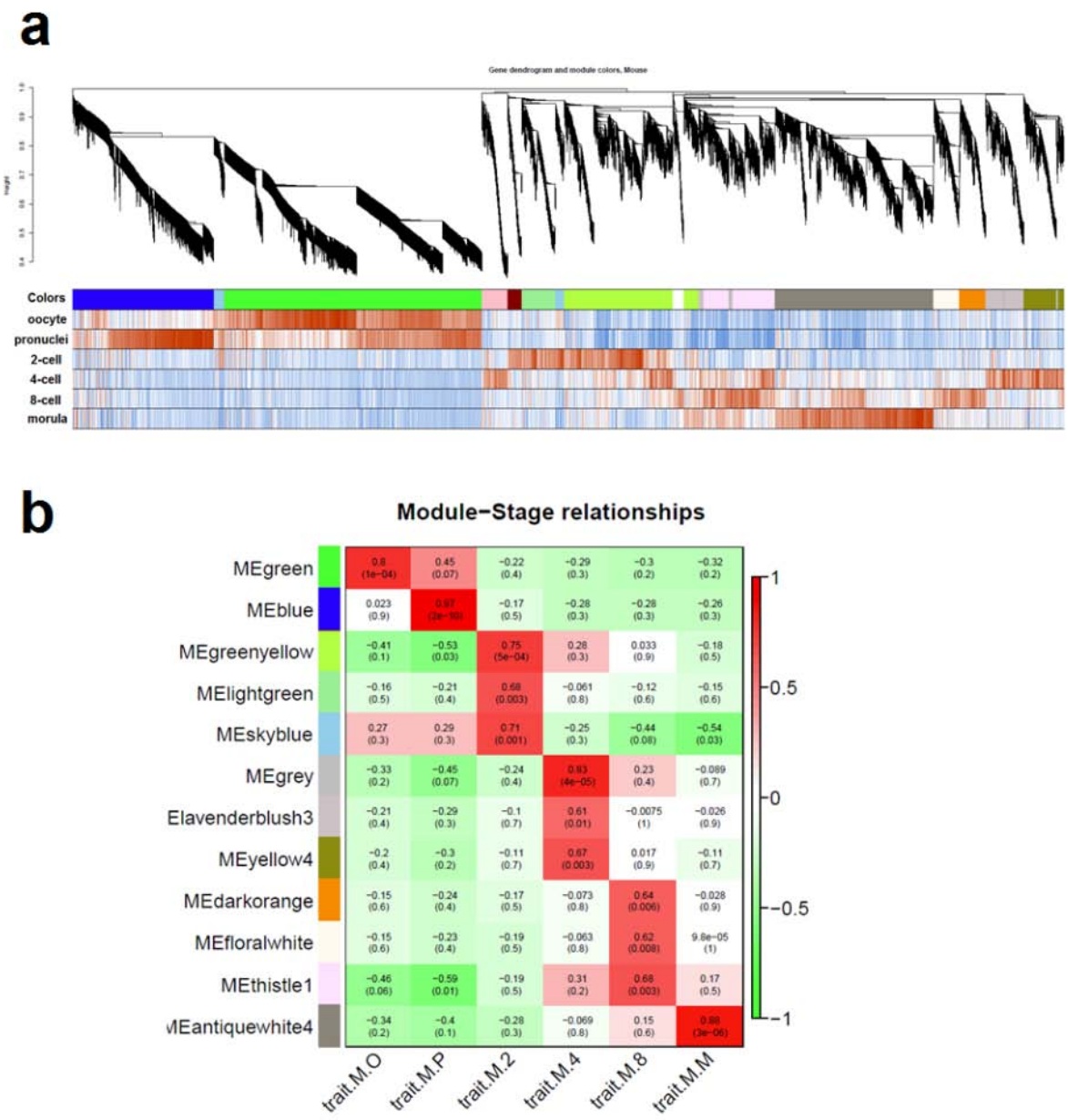
Supplementary Figure 4. Pairwise comparison of gene expression between consecutive stages of human and mouse preimplantation development. Bar graph showing the total number of differentially expressed genes between successive developmental stages (FDR<5% and minimum 2-fold change). Number of oocytes and embryos are as summarized in Supplementary Table 1. Up- and down-regulated genes are relative to the previous stage. O: oocyte; 1: 1-cell (includes pronuclear and zygotes); 2: 2-cell, 4: 4-cell; 8: 8-cell; M: morula.



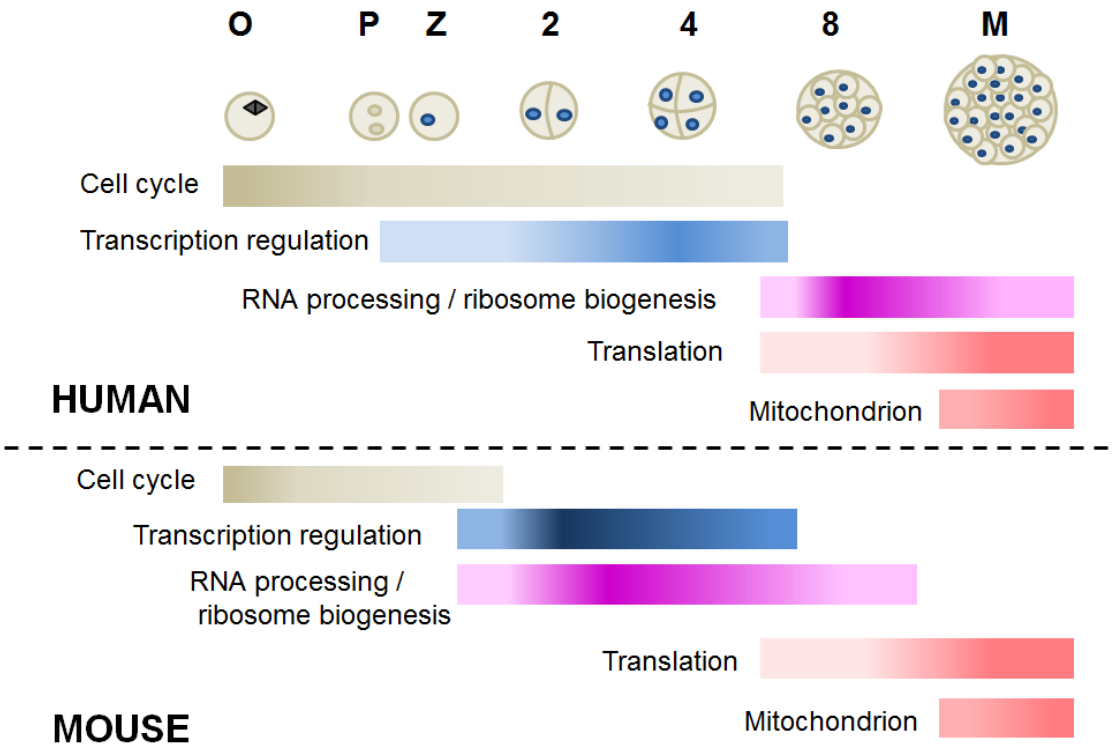
Supplementary Fig. 5. Distribution of SNP function in a single 8-cell human embryo. This pie chart show distribution of SNP function in 1,225 SNPs that are homozygous for the variant, and validated in two other single cells from the same embryo. SNP variant functional change was predicted using the SIFT program[23].



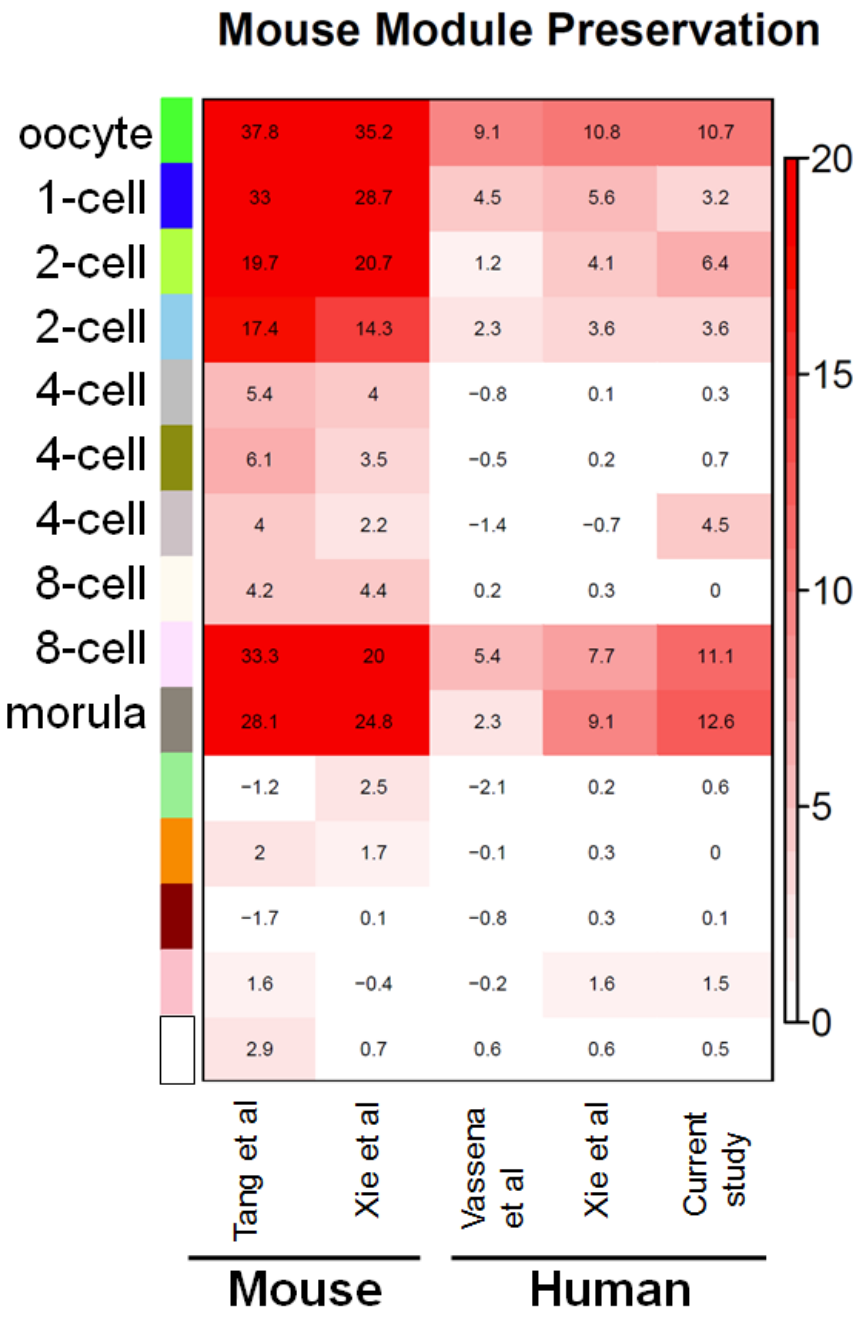
Supplementary Fig. 6. Heatmap reporting correlations (and corresponding p-values) between modules and stage indicators. Each module is represented by its module eigengene (which can be interpreted as weighted average). The reported modules (rows) are often highly correlated (i.e. over-expressed) with distinct stages (columns). Colour legend indicates the level of correlation between gene co-expressions with stage-specific expression.



Supplementary Fig. 7. Network analysis of mouse pre-implantation development.
a) Hierarchical cluster tree showing co-expression modules identified using WGCNA. Modules correspond to branches and are labeled by colours as indicated by the first colour band underneath the tree. The remaining colour bands reveal correlations between transcripts and stage indicators. Red and green indicates that the transcript is highly correlated or anti-correlated for the particular stage, respectively. Note the high overlap of red bars between multiple developmental stages. b) Heatmap reporting correlations (and corresponding p-values) between modules and stage indicators. Each module is represented by its module eigengene (which can be interpreted as weighted average). Strikingly, the reported modules (rows) are often highly correlated (i.e. over-expressed) with distinct stages (columns). Colours indicate the level of correlation between gene co-expressions with stage-specific expression.



Supplementary Fig. 8. Schematic summarizing functional terms in stage-specific modules for human and mouse preimplantation development. Both human and mouse networks were constructed independently and gene ontology was performed on modules that showed stage-specific patterns as seen in Supplementary Fig. 6-7.



Supplementary Fig. 9. Mouse module preservation in available datasets. Heatmap of preservation scores between mouse modules (n=15) in available human and mouse datasets (x-axis). Color legend represents Zsummary scores. Module colors are as defined in Supplementary Fig. 7. Data presented are from refs 10, 11, and 25.

References

38 Braude, P., Bolton, V. & Moore, S. Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. Nature 332, 459-461, doi:10.1038/332459a0 (1988).

39 Dobson, A. T. et al. The unique transcriptome through day 3 of human preimplantation development. Human molecular genetics 13, 1461-1470, doi:10.1093/hmg/ddh157 (2004).

40 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 101, 777-791 (1987).

41 Gregg, C. et al. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. Science 329, 643-648, doi:10.1126/science.1190830 (2010).

42 Wagner, J. R. et al. Computational analysis of whole-genome differential allelic expression data in human. PLoS computational biology 6, e1000849, doi:10.1371/journal.pcbi.1000849 (2010).