SCIENTIFIC
REPORTS
natureresearch

OPEN

An integrated genome-wide multi-omics analysis of gene expression dynamics in the preimplantation mouse embryo

Steffen Israel[1], Mathias Ernst[2,3], Olympia E. Psathaki[1,4], Hannes C. A. Drexler [1], Ellen Casser[1], Yutaka Suzuki[5], Wojciech Makalowski[6], Michele Boiani [1], Georg Fuellen[2] & Leila Taher[2,3]

Early mouse embryos have an atypical translational machinery that consists of cytoplasmic lattices and is poorly competent for translation. Hence, the impact of transcriptomic changes on the operational level of proteins is predicted to be relatively modest. To investigate this, we performed liquid chromatography–tandem mass spectrometry and mRNA sequencing at seven developmental stages, from the mature oocyte to the blastocyst, and independently validated our data by immunofluorescence and qPCR. We detected and quantified 6,550 proteins and 20,535 protein-coding transcripts. In contrast to the transcriptome – where changes occur early, mostly at the 2-cell stage – our data indicate that the most substantial changes in the proteome take place towards later stages, between the morula and blastocyst. We also found little to no concordance between the changes in protein and transcript levels, especially for early stages, but observed that the concordance increased towards the morula and blastocyst, as did the number of free ribosomes. These results are consistent with the cytoplasmic lattice-to-free ribosome transition being a key mediator of developmental regulation. Finally, we show how these data can be used to appraise the strengths and limitations of mRNA-based studies of pre-implantation development and expand on the list of known developmental markers.

It has been about *100 years* since the *mouse* became a premier *model* organism. This status has been reinforced by the arrival of high-throughput RNA sequencing technologies, making it possible to investigate the regulatory circuits underlying development in detail. However, it is uncertain how closely RNA changes correlate with the operational level of the proteins. In fact, work in plants, yeast, lower vertebrates[1] and mammalian cell lines[2] has revealed a modest correlation. Mouse oocytes and early embryos feature an atypical translational machinery regarded to be poorly competent for mRNA translation ('cytoplasmic lattices' in place of free ribosomes[3]). Thus, the impact of transcriptional changes on the embryo proteome is expected to be limited. Indeed, in some cases the mRNA is detected throughout preimplantation development, but the protein is only observed from a certain pre-implantation stage onward[4]; or the mRNA is degraded soon after fertilization, while the protein persists through the blastocyst stage[5–7]. Unfortunately, conventional tools for protein analysis such as antibodies (immunofluorescence, immunocytochemistry, western blotting) do not scale well to genome-wide investigations.

[1]Max-Planck-Institute for Molecular Biomedicine, Roentgenstr. 20, 48149, Muenster, Germany. [2]Institute for Biostatistics and Informatics in Medicine and Ageing Research, Rostock University Medical Center, Ernst-Heydemann Str. 8, 18057, Rostock, Germany. [3]Division of Bioinformatics, Department of Biology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Staudtstr. 5, 91058, Erlangen, Germany. [4]University of Osnabrück, Center for Cellular Nanoanalytics Osnabrück (CellNanOs), Integrated Bioimaging Facility Osnabrück (iBiOs), Barbarastr. 11, 49076, Osnabrück, Germany. [5]Department of Medical Genome Sciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Chiba, 277-8562, Japan. [6]Institute of Bioinformatics, Faculty of Medicine, University of Münster, Niels Stensen Str. 14, 48149, Münster, Germany. Steffen Israel and Mathias Ernst contributed equally. Correspondence and requests for materials should be addressed to M.B. (email: mboiani@mpi-muenster.mpg.de) or G.F. (email: fuellen@uni-rostock.de) or L.T. (email: leila.taher@fau.de)

Large-scale qualitative and quantitative proteomic technologies have matured over the past two decades. In particular, direct measurement of proteins using mass spectrometry (MS) holds great promise as a complement to transcriptomics. Still, current high-throughput protein quantification methods are less sensitive than those for mRNA. Because mammalian oocytes and embryos are small and the size of the detected proteome is directly related to the amount of input material, the analysis of the mammalian oocyte and embryo proteomes with MS was effectively prohibitive until a few years ago. This is in contrast to *Xenopus* or *Drosophila*, in which a single or a few oocytes are sufficient to detect ~5,000 proteins[1,8–10]. Even in the case of relatively large mammalian oocytes and embryos, such as those of bovines, 100 of them[11,12] only enabled the detection of ~1,000 and 1,500 proteins. Mouse oocytes and embryos are smaller and, thus, 7,000 oocytes/zygotes were required to identify ~3,000 proteins up to the 1-cell stage in 2010[13], while 3,000 blastocysts were necessary to determine ~2,500 proteins in 2014[14]. Very recently, Gao *et al.* collected samples consisting of 4,000 to 8,000 embryos to distinguish ~5,000 proteins across six developmental stages, from the 1-cell stage to the blastocyst[15]. Hence, refraining from mass-killing oocyte donors or producing oocytes from stem cells *in vitro*[16], mouse embryologists are forced to achieve more with less. Gradual and continuous improvement of our protocols[17–20] over several years, including the optimization of buffers and sample collection conditions, have substantially improved our yields.

We combined high-throughput liquid chromatography-tandem mass spectrometry (LC-MS/MS) with mRNA sequencing to generate datasets encompassing seven stages of mouse development spanning from the oocyte to the blastocyst. We anticipate that this resource will be key to gaining a greater understanding of the oocyte to embryo transition, and provide two examples of its varied applications: (1) how to query the 'rule' of weak transcript/protein correlation in order to expose exceptions to the rule; and (2) how to expand the list of markers in order to follow the oocyte-to-embryo transition. Our dataset enriches the status of the mouse as a model system in developmental biology with the protein dimension, enabling a better understanding of the gene expression cascade that leads to the phenotype.

Results

Ultrastructural data underscore the relevance of a direct examination of the embryonic proteome.
To systematically investigate the relationship between the proteome and the transcriptome in the developing mouse, we chose the paradigm of recovering fertilized oocytes *in vivo* after ovarian stimulation and culturing them *in vitro* in KSOM(aa) medium under 5% CO_2 in air (see Methods). This made it possible to continuously monitor the progression of the embryos, to identify and collect stages more precisely, and to allay concerns over the quality of embryos developing inside a hormonally stimulated genital tract[21,22]. In a separate group of embryos used to test for developmental quality, 89.5% (N = 258) of the fertilized oocytes developed to blastocyst and, of these, 42.3% (N = 104) progressed to term (embryo transfer). Typical features of early mouse development, including changes in endoplasmic reticulum (ER) architecture[23] and in ribosome morphology[24–26] were recapitulated, supporting the use of our *in vitro* system to yield embryos that are representative of normal development. In particular, we noted that hexagonal-shaped free ribosomes enabling efficient protein synthesis[24–26] are rare prior to the morula stage (see Fig. 1A). Nevertheless, developmental progression was impeded when cycloheximide (CHX) – an inhibitor of protein synthesis – was added to the culture medium (see Fig. 1B). Briefly, the number of embryos that were able to develop to the next stage was always smaller in the presence of cycloheximide, independently of the developmental stage. Although the numbers have been reported to be sensitive to the exact time when CHX is added to the culture medium and to its concentration, our results are in agreement with previous studies[27,28] and indicate that protein synthesis is essential for further development of the early embryo.

Together, these data suggest that the impact of transcriptional changes on the proteome may be small, calling for a direct examination of the embryonic proteome.

A high-quality proteome of mouse oocytes and preimplantation embryos to a depth of 6,550 proteins.
For the proteome analysis we collected and processed a total of ~12,600 oocytes or embryos, in three biological replicates of ~600 oocytes/embryos per developmental stage: unfertilized oocytes, fertilized oocytes with pronuclei, and preimplantation embryos at the 2-, 4-, 8-cell, advanced morula and blastocyst stages (see Methods). The detected proteome comprised 6,550 proteins. Among these, 5,217 proteins were detected in at least two replicates of one or more developmental stages, and 1,709 proteins were detected in all replicates of all developmental stages. Protein abundance measurements (L/H ratios, see Methods) were highly reproducible, with minimum Spearman's rank correlation coefficients between replicates in the range of 0.67 to 0.76 (for the oocyte and 2-cell stage, respectively, see Supplemental Fig. S1). Compared to the theoretical proteome (see Supplemental Methods), the 6,550 detected proteins are mainly involved in RNA processing, organelle organization, intracellular transport and cellular metabolism. Although these processes are not exclusive to preimplantation development, they are consistent with the nature of embryonic cleavage as a phase of development during which biomass is not produced *de novo* but rather reorganized.

The complete proteome of the mouse preimplantation embryo is unknown. In order to estimate the completeness of our dataset, we computed the fraction of members of 233 known mammalian protein complexes that are present in our dataset (based on[29], see Supplemental Methods). Since all protein members are required for the function of a complex, undetected members hint at a technical limitation rather than genuine biological absence. The overall median for the fractions of complex members detected in at least one replicate was 0.80, and ranged from 0.75 to 0.80, depending on the developmental stage (see Supplemental Fig. S2). In addition, we directly compared our dataset to a very recently published dataset[15] in which 4,830 different proteins were identified in at least one of two replicates from six developmental stages (1-cell to blastocyst). We found that 4,028 (83%) of these proteins are contained in a reduced version of our dataset comprising the same six developmental stages (see Supplemental Fig. S3), and that our reduced dataset contains an additional 2,369 proteins (not present in the
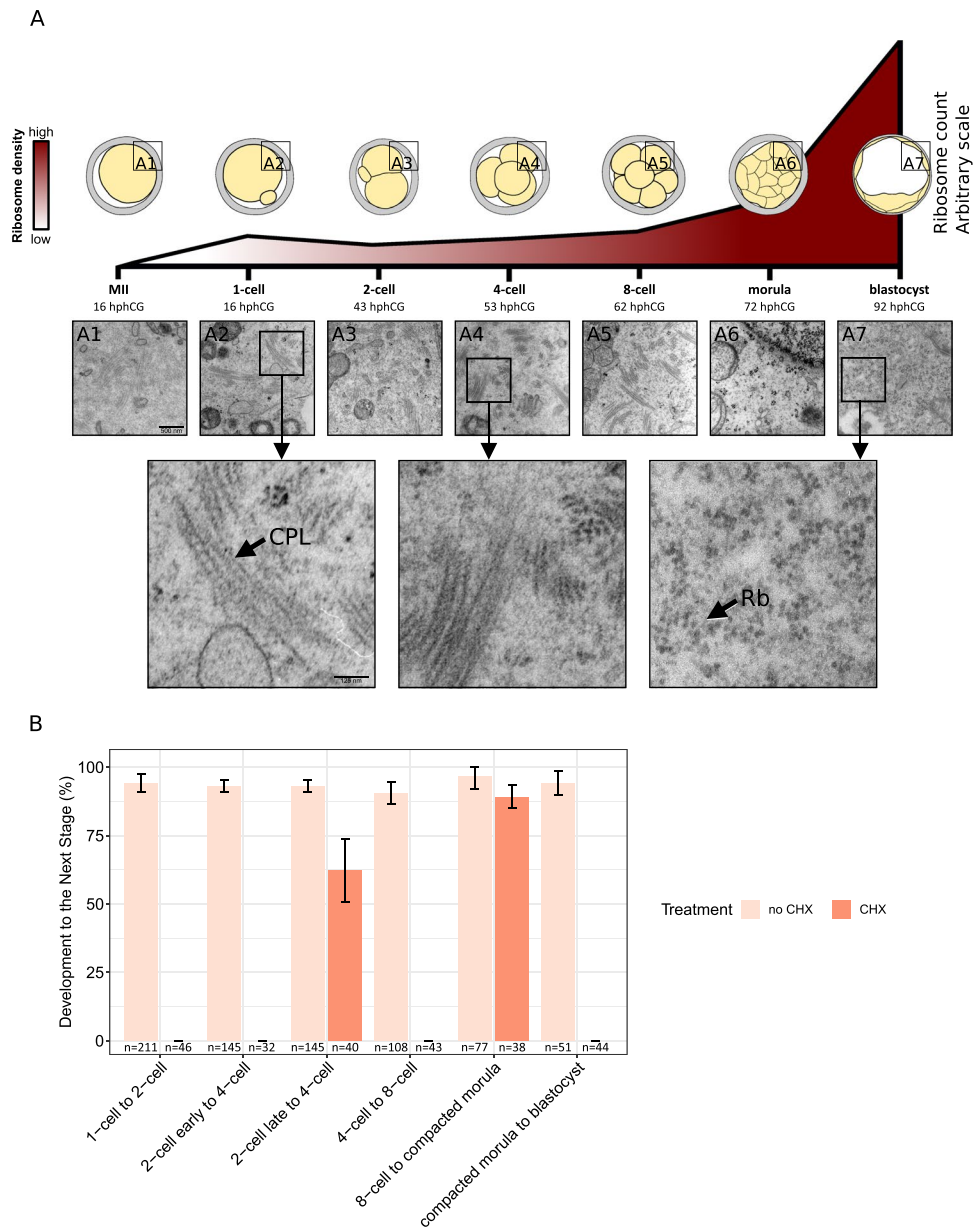
Figure 1. *In vivo*-fertilized, *in vitro*-cultured mouse oocytes as a source of embryonic material for proteomic analysis. (**A**) Oocytes and developmental stages were examined in ultrastructure. The density of hexagonal-shaped free ribosomes increases over time during preimplantation development (estimates are based on three sections from three different embryos of each stage). hCG was injected at 5 pm. Indicated under each stage is the number of hours post hCG (hphCG). Micrographs of cytoplasmic lattices (black arrow, "CPL") and free ribosomes (black arrow, "Rb") are shown. (**B**) The treatment of embryos with cycloheximide (CHX) documents that protein synthesis is necessary for developmental progression to the next stage. For each stage transition, the height of the bars denotes the percentage of embryos developing to the next stage without adding CHX; orange bars denote the same percentage (if any) after treatment with CHX. The numbers under the bars indicate the total number of embryos examined.

alternative dataset[15]). These findings suggest that we have achieved a high coverage – of up to 80% – of the mouse preimplantation proteome.

The dynamics of protein expression orchestrating preimplantation development is complex. As described numerous times on the mRNA level, fertilization is followed by extensive gene expression reprogramming. Nevertheless, the impact of transcriptional changes on the proteome is uncertain. Thus, it has been hypothesized that once activated, a gene continues to be transcribed during later developmental stages, resulting in product accumulation[30] that extends into the proteins. On the other hand, early protein studies of the mouse embryo based on radioactive gel electrophoresis support the hypothesis that protein expression occurs in phases[31]. To identify proteins whose expression significantly fluctuates as a function of the developmental
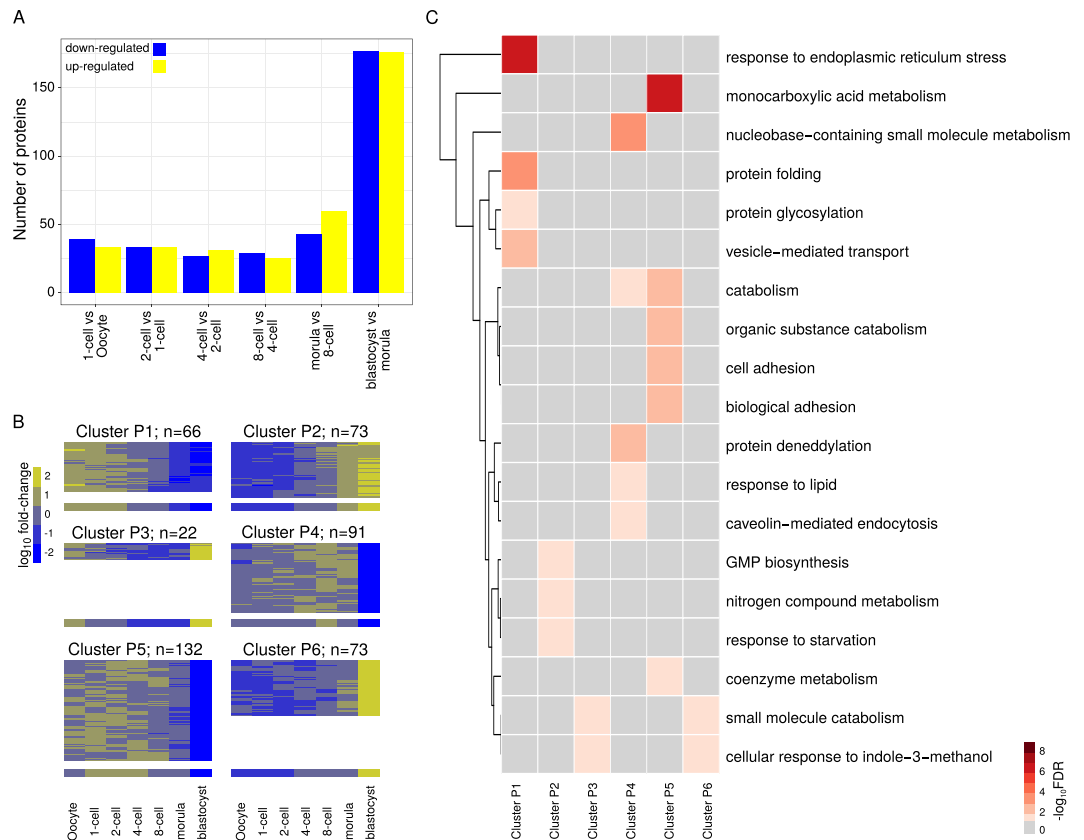
Figure 2. Differentially expressed proteins across preimplantation development and their functional enrichment. (**A**) Number of differentially expressed proteins between pairs of consecutive developmental stages (fold-change ≥ 2 or ≤ 0.5 between any two developmental stages, P-value ≤ 0.05 from ANOVA). (**B**) Expression profile of protein clusters. The heatmaps show fold-changes relative to the oocyte scaled using the z-score transformation. The height of the heatmaps is proportional to the number of proteins in each cluster, which is also indicated. The median fold-change across all cluster members for each developmental stage is represented below the heatmaps. (**C**) Annotation of protein clusters. Gene Ontology (GO) terms associated (FDR ≤ 0.05) with each cluster were summarized with REVIGO[89]. REVIGO "representatives" for the individual GO terms are listed on the right of the heatmap. Statistical significance (sum of the $-\log_{10}$ FDR of the individual GO terms) for the annotation of each of the clusters is represented using a color gradient. Non-significant associations are represented in gray. Details are presented in Supplemental Table S2.

stage, we subjected the 5,217 proteins detected in at least two replicates of at least one developmental stage to an analysis of variance (ANOVA, see Methods). This revealed a total of 1,290 (25%) differentially expressed proteins (P-value ≤ 0.05). Among these, 905 proteins exhibited fold changes ≥ 2 or ≤ 0.5 between any two developmental stages and 488 proteins did between consecutive developmental stages (see Fig. 2A). A relatively large amount of the latter – 253 proteins – only featured such fold changes during the transition from the morula to the blastocyst (see Supplemental Fig. S4). Compared to the detected proteome, the 488 proteins were associated with small molecule and carboxylic acid/carbohydrate metabolism, enzymatic activity; and (extracellular) exosome production (FDR ≤ 0.05, see Supplemental Methods and Table S1). These terms are consistent with a sequence of landmark events in mouse preimplantation development, such as the enzymatic transition from metabolic usage of pyruvate to usage of glucose[32], and the paracrine communication between embryos[33] as well as between the embryos and the maternal genital tract[34].

Fuzzy clustering of the 772 proteins detected in at least two replicates of each developmental stage that were differentially expressed (P-value ≤ 0.05) and showed a fold change ≥ 2 or ≤ 0.5 between any two developmental stages revealed six clusters (see Supplemental Methods and Fig. 2B,C). The two largest clusters (P5 and P4) comprise proteins whose expression decreases sharply between the morula and blastocyst stages and, compared to the detected proteome, are primarily enriched in monocarboxylic acid metabolism (P5), and nucleobase-containing small molecule metabolism (P4, see Fig. 2C and Supplemental Table S2). Clusters P5 and P4 are approximately mirror images of clusters P6 and P3, respectively. Nevertheless, the proteins in clusters P6 and P3 have their own functional profiles; thus, both clusters are connected to small molecule catabolism and cellular response to indole-3-methanol. The two remaining clusters (P1 and P2) are mirror images of each other and comprise proteins that steadily decrease or increase towards the blastocyst. Proteins in cluster P1 are mainly associated with
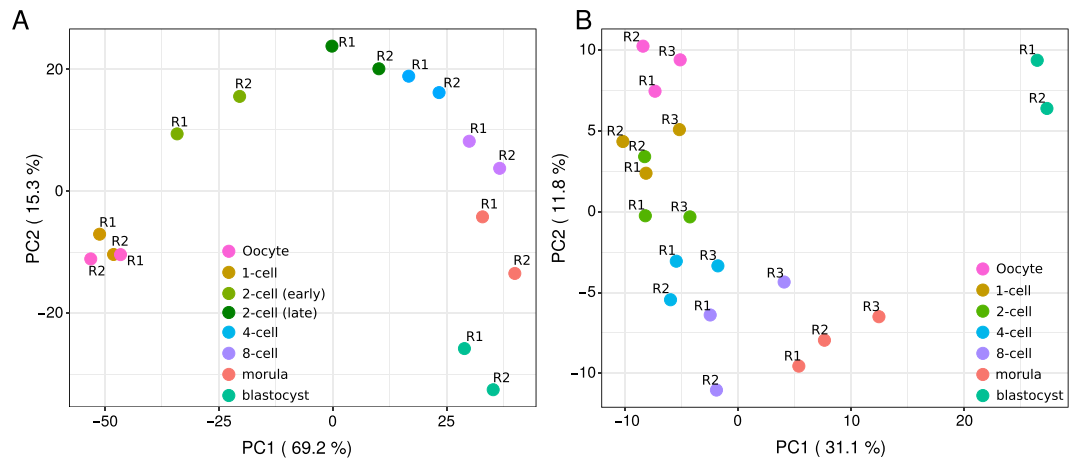
Figure 3. The proteome and the transcriptome develop differently in time. (**A**) Principal component analysis (PCA) of the expression values of the detected transcripts that are the cognates of the proteins detected in all replicates at all developmental stages (see B). The first two PCs are shown, with sample points colored by developmental stage. (**B**) Principal-component analysis of the ($\log_2$) L/H ratios of the 1,709 proteins detected in all replicates of all developmental stages.

response to endoplasmic reticulum stress and protein folding, whereas those in cluster P2 are related to GMP biosynthesis, nitrogen compound metabolism and response to starvation.

We independently validated our proteomics measurements using immunofluorescence assays. Among the proteins that are present throughout development (albeit more abundant at the beginning) we selected Ddx6, which is associated with processing bodies (P-bodies) involved in the storage and degradation of mRNAs[35]. Among the proteins that are detected in oocytes and early stages but become undetected later on, we selected Rc3h1 (roquin), which is an element of a post-transcriptional repression pathway and whose mutation leads to the *sanroque* phenotype[36]. Among the proteins that are not detected in oocytes and early stages but become detected later on, we selected Alppl2, known for its role in the placenta and expressed in the trophectoderm of the preimplantation embryo[37–39]. The immunofluorescence profiles of Ddx6, Rc3h1 and Alppl2 matched the corresponding proteomics profiles (see Supplemental Fig. S5). We further validated our proteomics measurements with enzymatic/immunofluorescence data collected from the literature and/or obtained in the past by our own laboratory on 33 proteins (37 sets of measurements across multiple developmental stages, see Supplemental Table S3). Specifically, we quantified the similarity between the expression profiles as determined by enzymatic/immunofluorescence assays and our proteomics pipeline by computing the Spearman's rank correlation. We observed strong correlations (Spearman's rank correlation coefficients between 0.6 and 0.79) for seven proteins (and seven sets of measurements) and very strong correlations (Spearman's rank correlation coefficients between 0.8 and 1.00) for five proteins (seven sets of measurements). The results are significant compared to the random expectation (empirical P-value < 0.006, see Methods).

Taken together, these results reveal systematic changes of the proteome of the embryo as it develops. Furthermore, these changes are complex and unlikely to reflect a mere alternative between monotonic accumulation and stage-specific expression[30,31].

Changes in protein abundances become more prominent as development progresses, and so does the concordance with changes in transcript expression values.

Previous transcriptome-based studies of mouse embryonic development have shown that the transcriptomes of oocytes and early embryos can be clearly divided into two groups: prior and after the 2-cell stage[40–42]. However, the most conspicuous morphological changes during preimplantation development – compaction and cavitation – occur well after the 2-cell stage, in the morula[41]. To directly compare the oocyte-to-embryo transition on the protein and mRNA levels, we generated our own transcriptome using RNA-seq. For this purpose, we collected and processed a total of 3,424 oocytes or embryos in two biological replicates of 214 oocytes/embryos per developmental stage (see Methods). Anticipating major differences between the early and late 2-cell stage, we considered these separately. We identified a total of 20,535 protein-coding transcripts with at least one read count in any of the samples.

As expected, principal component analysis (PCA) of the expression values of the transcripts showed that developmental stages can be distinguished based on their transcriptomes and that most of the variance in the data is contributed by changes at early developmental stages, (see Fig. 3A). PCA performed on the abundances of the cognate proteins clearly distinguished the developmental stages based on their proteomes, albeit with most of the variance in the data being contributed by changes between the morula and the blastocyst. Indeed, the progression from the 4-cell to the blastocyst embryo aligned almost perfectly with the increase of the first principal component (PC) and explained 31.1% of the variance, while the progression from the oocyte to the 4-cell stage aligned with the decrease of the second PC and explained only 11.8% of the variance (see Fig. 3B). These findings indicate that in contrast to the transcriptome, in the proteome the oocyte-to-embryo transition is less connected to the 2-cell stage, with the protein expression signature of the blastocyst being particularly further apart from those of the other developmental stages considered. This is in agreement with the establishment of two blastocyst cell
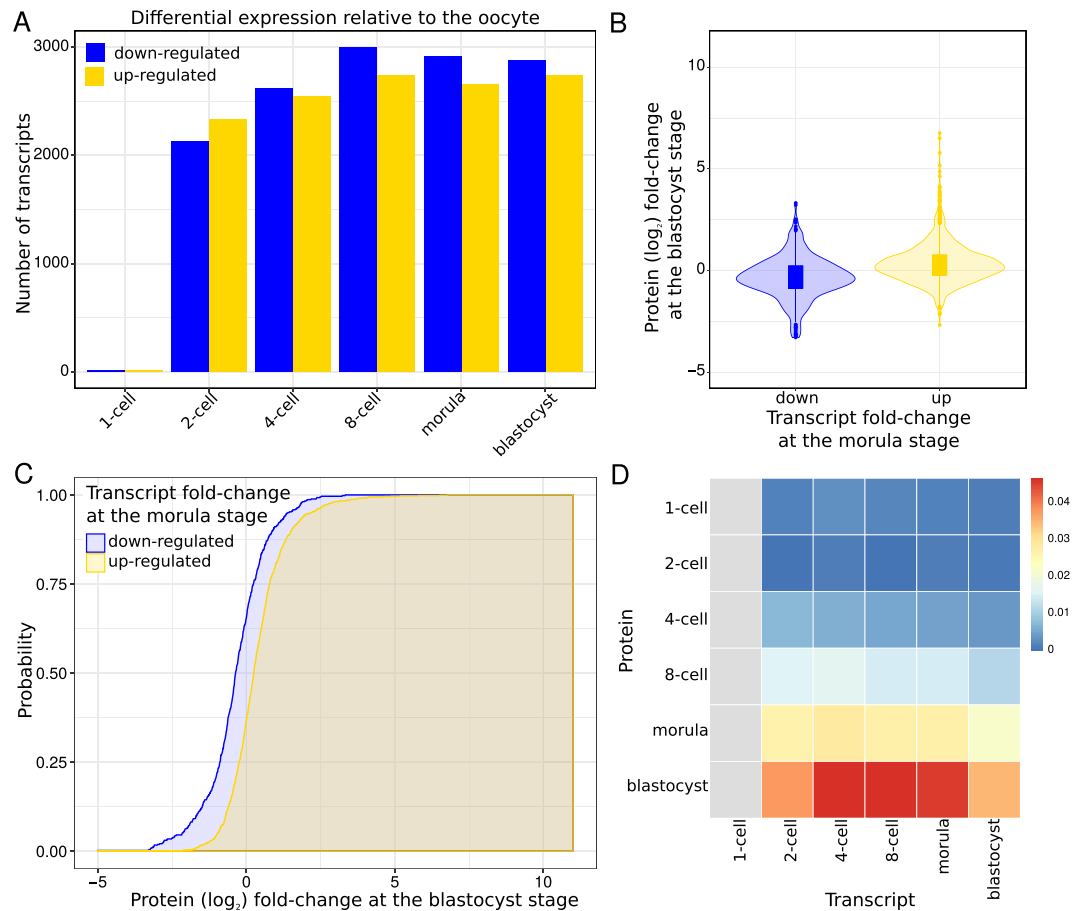
Figure 4. Changes in the transcriptome are reflected at the proteome level from the morula stage onwards. (**A**) Number of differentially expressed transcripts at each developmental stage relative to the oocyte (FDR $\leq 1 \times 10^{-5}$ and fold-change ≥ 2 or ≤ 0.5 between any two developmental stages). (**B**) Violin plot showing the distribution of protein ($\log_2$) fold-changes at the blastocyst stage relative to the oocyte for proteins whose cognate transcripts are down- (blue) or up- (yellow) regulated at the morula stage relative to the oocyte. (**C**) Cumulative distribution functions (CDFs) of protein ($\log_2$) fold-changes at the blastocyst stage relative to the oocyte for proteins whose cognate transcripts are down- (blue) or up- (yellow) regulated at the morula stage relative to the oocyte. 1,617 proteins and their cognate transcripts were used to estimate the CDF: 5,565 transcripts were found differentially expressed between the morula and the oocyte; 1,617 of the cognate proteins were detected in both the blastocyst and the oocyte. The CDF for the proteins whose transcripts are up-regulated is shifted to the right compared to the CDF for the proteins whose transcripts are down-regulated. We used the difference between the two areas bounded by the CDFs (shaded) to quantify the shift and, thereby, the impact of the transcript expression changes at the morula stage on the protein expression changes at the blastocyst stage. (**D**) Heatmap representing the difference in the area of the two CDFs for all pairs of stages. The x-axis of the grid corresponds to transcripts; the y-axis corresponds to proteins. The colors indicate the differences between the two areas bounded by the corresponding CDFs normalized using the entire protein fold-change range. Red indicates a large area, and hence a considerable shift between the distributions, while blue indicates the opposite. Gray indicates the pairs of stages for which we did not estimate the CDFs because less than 25 proteins/transcripts were detected and/or found differentially expressed, respectively.

populations that differ radically in their metabolic and cell cycle parameters: polarized external cells (the future trophectoderm) and apolar internal cells (the future inner cell mass)[43,44].

Consistent with the PCA, we found a strikingly weak correlation between the changes in protein abundances and the changes in transcript expression values relative to the oocyte, with Spearman's rank correlation coefficients in the range of −0.06 (2-cell early versus 1-cell and 2-cell early versus 2-cell) to 0.41 (morula versus blastocyst, see Supplemental Fig. S6). Thus, to explore the relationship between the transcriptome and the proteome in the course of preimplantation development, we divided the proteins into two disjoint groups according to the direction of change in expression of their cognate transcripts relative to the oocyte (see Fig. 4A). More precisely, given developmental stages S_i and S_j, we separated the proteins into two groups: (i) those with transcripts up-regulated at S_i; and (ii) those with transcript down-regulated at S_i. Then, for each of the two groups of proteins, we estimated the probability of observing a certain protein ($\log_2$) fold-change at S_j relative to the oocyte (see Fig. 4B,C and Supplementary Methods). For any ($\log_2$) fold-change x, if the protein expression changes at
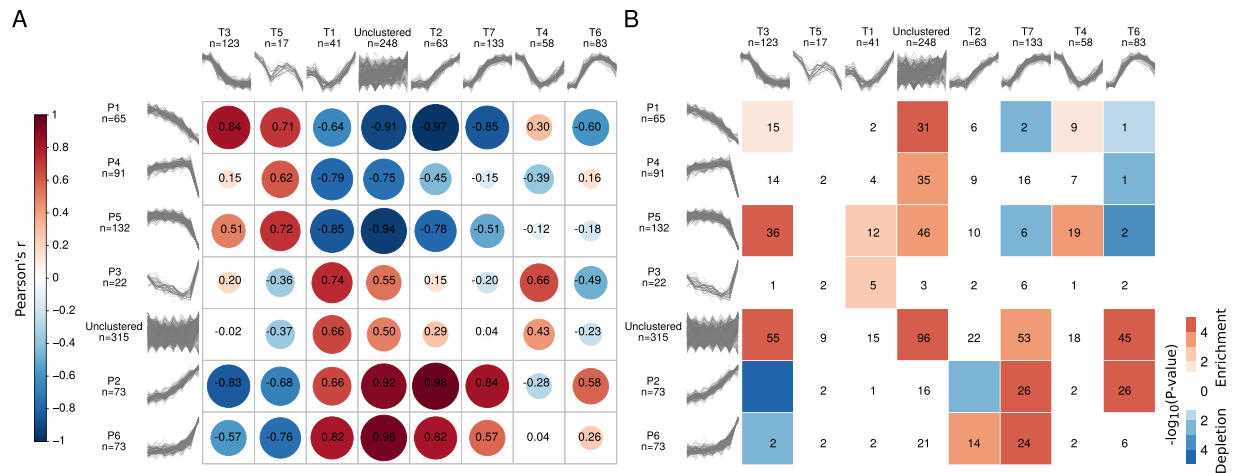
Figure 5. Comparison between protein and transcript clusters. (**A**) Pearson correlation coefficients (in black) computed between the medians of the expression profiles of pairs of protein and transcript clusters. The correlation matrix based on the Pearson correlation coefficients between the median (log2) fold changes across the members of protein and transcript clusters was visualized using the R corrplot package[90]. The size and color of the circles are both indicators of the magnitude and sign of the correlation. The matrix was reordered based on hierarchical clustering using the complete linkage method. (**B**) Significance of the overlap (−log10 P-value) calculated using Fisher's Exact test) between the members of protein (rows) and transcript (columns) clusters. Significant enrichments are depicted in red and depletions in blue. The numbers indicate the size of the overlap; overlaps of size zero are not indicated.

S_j reflect the transcript expression changes at S_i, the probability of observing a protein with a (log$_2$) fold-change of x or less at S_j is expected to be greater for those proteins whose transcripts are down-regulated than for those whose transcripts are up-regulated at S_i. Hence, we quantified the concordance between protein and transcript expression changes by measuring the difference between the areas bounded by the two implicit cumulative distribution functions (CDFs, see Supplementary Methods). This analysis revealed little or no concordance between protein and transcript expression changes at early developmental stages (see Fig. 4D and Supplemental Fig. S7). The concordance, however, increased towards later developmental stages, with expression changes at the morula and blastocyst stages exhibiting the overall highest concordances. Moreover, the concordance for the transcript expression changes at the 4-cell, 8-cell and morula stages was highest for the protein changes at the blastocyst stage, and higher than that between transcript expression changes at the blastocyst stage and protein changes at the blastocyst stage (see Fig. 4D).

Altogether, these results are in agreement with the increase in the density of free ribosomes that enable efficient protein synthesis only starting at the morula stage (see Fig. 1A). Despite some *de novo* transcript synthesis beginning at the 1-cell stage, the paucity of a conventional translation machinery (i.e., the paucity of free ribosomes) prevents transcripts from being robustly translated until the morula stage. Furthermore, despite the steep increase of the free ribosomes, a delay between transcription and translation is still evident at the blastocyst stage. Overall, the majority of the proteins do not match the previously described[45] stage-specific groups of transcripts that support a 'hit and run cascade' model for early embryonic development. Instead, our results document only a moderate amount of change in the proteome, suggesting a steady basal translation of transcripts into proteins, and a role for subcellular compartmentalization and storage in order to make the proteins available when and where required.

Exceptions to the rule of weak transcript-protein correlation define a special class of genes with distinct developmental functions. To exemplify how our dataset can be analyzed to better understand the relationship between the transcriptome and the proteome during preimplantation development, we applied fuzzy clustering to the transcripts of the 772 proteins that we had clustered before, and compared the transcript and protein clusters. Specifically, we clustered the (log$_2$) fold-changes of the transcripts relative to the oocyte, and found seven clusters (see Supplemental Methods, Fig. S8 and Table S4), which, in contrast to the protein clusters, are often characterized by expression profiles with evident inflection points either at the early or late 2-cell-embryo stage. Next, to quantify the similarity between the protein and transcript clusters, we computed the Pearson's correlation coefficient (r) between their expression profiles in a pairwise manner (see Methods). Out of a total of 42, 14 pairs had a Pearson's correlation coefficient greater than or equal to 0.5, indicating high similarity (see Fig. 5A). In addition, we assessed the overlap between the members of all pairs of protein and transcript clusters and found that only ten shared more proteins/transcripts than expected by chance (P-value ≤ 0.05, one-sided Fisher's exact test, see Fig. 5B). The overlap was particularly high among pairs of protein and transcript clusters with similar expression profiles ($r \geq 0.5$), with an odds-ratio of 7.8 (P-value = 0.008, one-sided Fisher's exact test), highlighting the fact that, despite the little overall concordance between protein and transcript expression changes, the expression of some proteins indeed mirrors that of their cognate transcripts. Compared to the

772 differentially expressed proteins and their cognate transcripts considered for clustering, the 146 genes that overlap among the pairs of clusters with similar expression profiles were enriched in positive regulation of secretion, reflecting the increasing role of the embryo-derived 'secretome' as development progresses in preparing the ground for the molecular dialogue between the embryo and the maternal endometrium[34].

For the purpose of identifying and characterizing the genes with either strongly correlated or anticorrelated protein and transcript expression profiles, we perused the correlation between the expression profiles across the developmental series of the aforementioned 772 differentially expressed proteins and their cognate transcripts. Despite the expected weak overall correlation (with a median Spearman's rank correlation coefficient across all genes of 0.18), we observed that the distribution of Spearman's rank correlation coefficients was relatively broad (see Supplemental Fig. S9). To enhance confidence in the observed profiles, we independently validated our proteomics and transcriptomics data using immunofluorescence and TaqMan (qPCR) respectively (see Methods). We selected three genes among those exhibiting strong negative correlations during preimplantation development and particularly from the 1- to the 4-cell stage: Pdia3, Top1, and DNAjb11. These genes are particularly appropriate in the context of our study, because mutations of Pdia3, Top1 and Dnajb11 interfere with development and prove lethal in homozygosis[46–49]. The results of the TaqMan assay for Pdia3, Top1 and Dnajb11 correlated positively with those of RNA-seq, as did the results of the immunofluorescence with those of LC-MS/MS (see Supplemental Fig. S10), confirming the existence of genes with strongly anticorrelated protein and transcript expression profiles.

Finally, with the comfort of the validation data, we moved on to analyze the features of the genes at the extremes of the distribution of Spearman's rank correlation coefficients. Indeed, 7% of the proteins and transcripts exhibited very strong positive correlations (≥ 0.8) and 3% showed very strong negative correlations (≤ -0.8). Among the former are genes involved in ubiquitin metabolism and ubiquitination (Dcun1d5, Uspx9, Dcaf8, Gabarapl2, Rnf114, Stt3b, Ube2g1), signal transduction (Arhgap12, Gna13, Pdpk1), synthesis and modification of DNA (Ctc1, Rrm2, Hmces), splicing and storage of mRNA, including translational initiation (Paip1, Rbm8a, C1qbp, Igf2bp3, Igf2bp2, Xab2, Nhp2). Among the latter are genes involved in membrane vesicle trafficking (Dynlrb1, Epn2, Vta1, Napa, Eea1), chaperoning (Hypk, Fkbp2), and protein glycosylation in association with ribosome binding (Rpn1, Rpn2). Known genes with established roles in development are found in both groups (e.g., Rrm2[50]; Igf2bp2[51]; Igf2bp3[52]; Epn2[53]). Overall, the proteins with very strong positive correlations are implicated in dynamic processes, while those with very strong negative correlations represent maintenance systems, with a convergence on signaling. Thus, the release and uptake of vesicles supported by the anticorrelated genes is one way to modulate the concentration of signaling molecules supported by the highly correlated genes, as exemplified by the case of Epn2[53].

Proteomic profiles suggest new markers to better follow the oocyte-to-embryo transition.
To show how our dataset can be applied to the identification of new candidate developmental markers, thereby broadening the options offered by morphology/morphokinetics or metabolic markers secreted into the culture medium, we examined the molecular basis of morphological staging. As an illustration, we uncovered new candidate markers to follow the oocyte-to-embryo transition, and thus compared the proteomes of early (oocyte, 1- and 2-cell embryos) and late (4-cell to blastocyst embryos) developmental stages. In particular, we trained and tested linear discriminant analysis (LDA) classifiers. Our results show that protein expression can be used to perfectly separate between early and late developmental stages, with an area under the Receiver Operator Characteristic (ROC) curve of 1.00 (see Supplemental Methods and Fig. S11). Samples from the 4-cell stage embryos were close to the decision boundary of the classifier, indicating at this stage the coexistence of features from both previous and later stages, and characterizing the 4-cell stage as a transitional stage. Further, we inferred twenty candidate markers for early and late developmental stages by ranking the proteins according to their relevance for the classification (see Supplemental Methods and Fig. 6A). These proteins include enzyme modulators, hydrolases and ligases (see Fig. 6B and Supplemental Fig. S12). In particular, Ddx6 is an RNA helicase that has been found in P-bodies[54] and is involved in translation repression and in 2-cell stage embryonic arrest[35]. Moreover, some of these proteins (e.g., Ppm1a and Wtap) are mediators of TGF-β and Wnt signaling[55,56]. This finding is compatible with the aforementioned overrepresentation of 'exosome production' among differentially expressed proteins, since signaling pathways rely in part on exosome-mediated mobilization. Interestingly, five of the twenty candidate markers (Calr, Hyou1, Pdia3, Pdia4 and Txndc5) are involved in the protein processing in endoplasmic reticulum (ER) pathway (KEGG identifier mmu04141[57–59], odds-ratio = 6.7, P-value = 0.002, Fisher's Exact test, see Fig. 6C), enlightening the molecular basis of the changes in ER architecture that take place during the transition from oocyte to embryo[60] and that are concomitant to the increase in protein synthesis and folding after EGA[61]. We independently validated the expression profiles of the candidate makers using a recently published dataset study[15] as well as additional SILAC data (see Supplemental Methods, Table S5 and Fig. S13). These twenty marker proteins constitute good candidates for further molecular studies of mammalian preimplantation development.

Discussion
In this study, we used MS-based proteomics to generate a proteome dataset with three biological replicates for the preimplantation stages of mouse development, from the oocyte to the blastocyst. This proteome was compared to the cognate transcriptome generated by RNA-seq. With 6,550 detected proteins, ours is the largest developmental proteome of a mammalian species characterized to date, and yet substantially smaller than the number of 20,535 protein-coding transcripts found in the same samples. A similarly conceived, recently published study conducted with a different workflow (TMT instead of SILAC) revealed nearly 5,000 proteins despite the much higher amount of input material used[15]. While neither of these datasets is complete, we found that our proteome
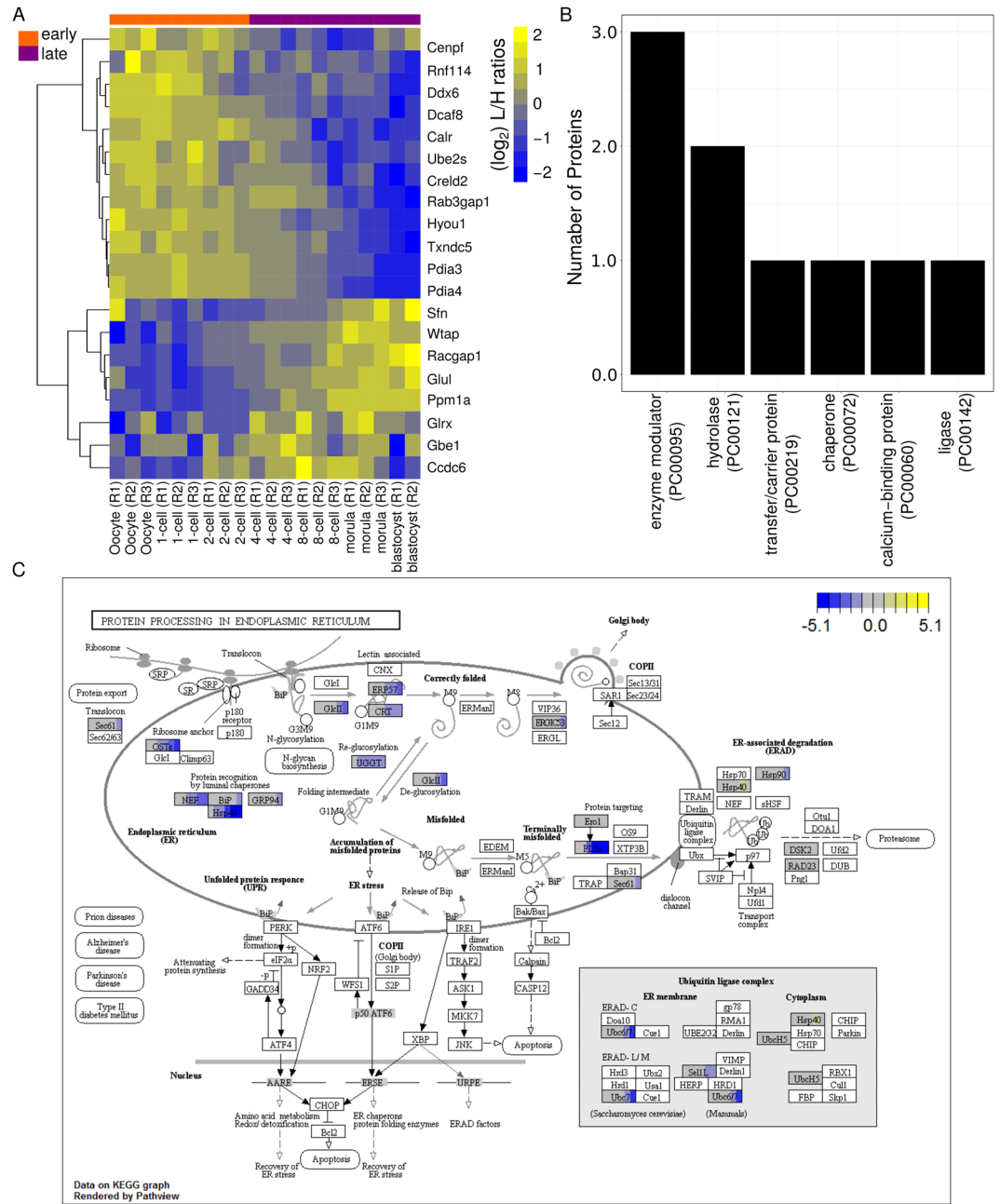
www.nature.com/scientificreports/

Figure 6. Classification of early and late preimplantation developmental stages based on protein abundances using Linear Discriminant Analysis (LDA). (**A**) Heatmap of ($\log_2$) L/H ratios for 20 candidate protein markers for distinguishing between early and late preimplantation developmental stages. The 20 samples are sorted chronologically according to developmental stage and replicate number. Row clustering was performed with a Pearson correlation-based distance using the complete linkage method. The package pheatmap in R was used for visualization[91]. (**B**) PANTHER14.1[92,93] protein classification available for seven of the twenty candidate proteins markers (see A). (**C**) Pathway analysis of genes differentially expressed in the murine "protein processing in endoplasmic reticulum" KEGG[57–59] pathway (mmu04141). The boxes representing the proteins/genes are uniformly divided by the number of developmental stages. Replicate averages are laid out chronologically from left to right across all developmental stages considered. The ($\log_2$) of the fold-change relative to the oocyte is indicated in yellow (up-regulated) or blue (down-regulated). Only proteins/genes among the 764 that were found differentially expressed across the developmental series are colored. The graphical representation of the KEGG[57–59] pathway was visualized using the "Pathview" R/Bioconductor package[94].

coverage is in the order of magnitude of up to 80%. Clearly, most mRNAs are stored and only translated when needed, and MS-based proteomics of developmental stages is not solely a matter of input amount: it is largely a matter of sample preparation and preprocessing (e.g., prefractionating) and of the experimental procedures and equipment used.

SCIENTIFIC REPORTS | *(2019) 9:13356* | https://doi.org/10.1038/s41598-019-49817-3

9

Our main finding when taking the sole proteome into consideration is that the majority of detected proteins change only moderately in abundance during the development from oocyte to morula. Accordingly, we hypothesize that the oocyte-to-embryo transition may last until the morula stage, in contrast to the swifter transition at the transcriptome level, largely accomplished between 2-cell and 4-cell stage. The blastocyst's proteome stands out as markedly different from the proteomes of the preblastocyst stages. This distinction is consistent with the formation of the first epithelium, the trophectoderm. Translation in the preimplantation embryo is limited by the availability of free ribosomes, which are the most active players of a cell's translational machinery, but poorly represented in pre-morula-stage mouse embryos. This explains why an impaired translational machinery does not affect blastocyst formation, but causes blastocyst implantation failure in mice[62].

Our main finding when comparing the protein abundance profiles with their cognate transcript profiles is that the projection of the proteome onto the developmental time axis differs from the prediction based on the transcriptome, with the correlation improving as development progresses. While most changes at the protein level explain the transition between the morula and the blastocyst, most changes at the mRNA level explain the transition between the oocyte and the 2- to 4-cell stage. Although the overall protein-mRNA correlation is weak, for a small subset (7%) of the detected proteome, the proteins and their cognate mRNAs have very similar profiles. Moreover, for another small subset (3%) of the detected proteome, the correlation is even negative, with protein levels increasing as transcript levels decrease. These cases may be explained, for example, by the packaging of RNA in granules, such as P-bodies[63,64], whereby the mRNA broken free from these granules becomes available for both translation and degradation. Notably, we observed a decrease of the P-body protein Ddx6 from oocyte to blastocyst, which together with the increase in free ribosomes would explain the improving protein-mRNA correlation as development progresses. These covariates make the anti-correlated proteins virtually impossible to predict from their transcripts. From our data it is now clear that these anti-correlations are no exceptions, but manifestations of a non-negligible phenomenon in mouse development.

Two limitations of our study, apart from artifacts that may occur in our *in vitro* setting as well as in the *in vivo* situation (caused by the hormonal status of the genital tract[15]), are the following. First, it is difficult to determine whether we failed to detect important proteins. However, our coverage estimates are in the order of magnitude of up to 80%, suggesting that the number of false negatives is bounded. Second, it is not known how the genotype of the gametes influences the composition of the developmental proteome. However, as reported by us[19], the proteomes of the oocytes of different inbred strains (129/Sv, C57Bl/6J, C3H/HeN, DBA/2J), while not identical, only differ in a minor proportion of detected proteins. A third limitation is that our ability to detect proteins in oocytes and embryos depends on the reference we used for SILAC. For example, trophectodermal markers seemed to be underrepresented in our dataset, although several of these proteins were also underrepresented in a study that did not use SILAC[15].

In summary, while there is still a long journey ahead until the proteome of mouse preimplantation development is exhaustively enumerated, our dataset constitutes a substantial contribution to closing the gap between 'predicted' phenotype (based on mRNA) and 'actual' phenotype (based on protein) of the mouse embryo. Except for a small subset of genes, proteins and mRNAs have discordant profiles, and this is in agreement with the paucity of free ribosomes observed at early stages of development. Hence, our proteome dataset enables a more direct investigation of mammalian developmental processes. The range of applications of our resource is broad. For instance, it facilitates the molecular definition of embryo quality, which has a major impact on the course of gestation and yet is insufficiently accounted for on the molecular level. While morphological/morphokinetic markers commonly used to predict an embryo's chances to develop can be subjective, our proteomic resource offers specific and measurable molecular candidates to complement the non-molecular markers. Thus, our LDA classifier was able to attain perfect separation between early and late developmental stages based solely on protein abundances. Also, since mammalian oocytes and embryos are produced in the gonads in comparatively small numbers (compared to e.g. *Xenopus*) and their availability can be subject to ethical and legal restrictions (e.g. in humans), knowing which gene products can be reliably predicted from mRNA has diagnostic value: these mRNA markers allow to make predictions that are backed by the proteins, and they do not require to consume the whole oocyte or embryo since cytoplasmic biopsies can be amplified for mRNA. For example, the cases of anti-correlation in which the mRNA is rapidly degraded after fertilization whereas the protein persist throughout the blastocyst stage, may be cases of candidate maternal genes. In any event, the biological implications of our findings are enormous: although virtually all studies of mammalian preimplantation development rely on transcriptomic data, we show that the predictive value of mRNAs for protein abundances – which are closer to the phenotype – is, at most, modest.

Methods
Ethics statement. This mouse study was performed in accordance with the recommendations of the Federation of Laboratory Animal Science Associations (FELASA) and with the ethical permit issued by the Landesamt fuer Natur, Umwelt und Verbraucherschutz (LANUV) of the state of North Rhine Westphalia, Germany (permit number: LANUV 81-02.04.2017.A432).

Metaphase II oocyte collection. Metaphase II (MII) oocytes of B6C3F1 mice aged 8–10 weeks were collected from the oviductal ampullae after gonadotropin priming with 10 IU of each PMSG and hCG, injected 48 hours apart, at 5 pm, as described[20,65].

***In vivo* oocyte fertilization and *in vitro* embryo production.** Gonadotropin-primed B6C3F1 females were mated to CD1 males (see Supplemental Fig. S14). Pronuclear oocytes were collected from oviductal ampullae at 10am on the day of the copulation plug. By 11am they had been freed of expanded cumulus cells in 50 U/mL hyaluronidase in HZCB medium, and placed in culture in 500 microliters KSOM(aa) medium[66] in 4-well plates (Nunc) under an atmosphere of 5% CO_2 in air at 37 degrees Celsius. All embryos were staged carefully based on morphology and time spent in culture (beginning at 11am on the day of isolation from the oviduct).

Transmission electron microscopy (TEM). Mouse embryos were fixed 2 h at room temperature in 2,5% glutaraldehyde (Merck, Darmstadt, Germany) in 0.1 M cacodylate buffer, pH 7,4 subsequently post-fixed for 2 h in 1% aqueous osmium tetroxide (Plano, Germany), dehydrated stepwise in a graded ethanol series and afterwards embedded in Epon 812 (Fluka, Buchs, Switzerland). Ultrathin (70-nm) sections were prepared with an ultramicrotome (EM UC6, Leica, Wetzlar, Germany), stained for 30 min with 1% uranyl acetate and 20 min in 3% lead citrate. Sections were examined at 50 kV in a Zeiss 109 transmission electron microscope (Zeiss, Oberkochen, Germany).

Sample preparation for LC-MS/MS. For the proteome analysis we collected and processed a total of ~12,600 oocytes or embryos from May 2014 to October 2016. During this time, mouse housing conditions, including diet (Teklad 2020SX), did not change. Specifically, we lysed, in triplicate, an average of ~600 oocytes/embryos per developmental stage: unfertilized oocytes, fertilized oocytes with pronuclei, and preimplantation embryos at the 2-, 4-, 8-cell, advanced morula and blastocyst stages. The samples were true biological replicates that were handled independently from start to end. Protein quantification was performed with our established spike-in SILAC-based labeling pipeline[17,19,20]. Briefly, oocytes and embryos were deprived of the zona pellucida by pipetting in warm acidic Tyrode solution for 30–60 seconds and then rinsing in protein-free HCZB medium (BSA replaced through *polyvinylpyrrolidone* 40 kDa). Each sample lysate was then mixed with an equal amount of isotopically labeled (heavy) lysate from F9 embryonic carcinoma (EC) cells[67], digested with trypsin, and subjected to MS analysis.

F9 EC cells were grown for several passages in RPMI 1640 medium (PAA, Cölbe, Germany), supplemented with 10% dialyzed fetal calf serum (Sigma, Deisenhofen, Germany), heavy amino acids $^{13}C_6^{15}N_2$-L-Lysine (K8) and $^{13}C_6^{15}N_4$-L-Arginine (R10; Silantes, Martinsried, Germany) as well as Glutamine and the antibiotics penicillin and streptomycin (Gibco, Darmstadt, Germany). The extent of labeling was 97.8%.

The F9 EC cell line was originally isolated by Berstine *et al.*[67] as a subline of the teratocarcinoma OTT6050 established by implanting a 6 day-old embryo in the testis of a 129/J mouse. F9 EC cells, have many characteristics of early mouse embryonal cells and can differentiate into almost all cell types[68–70], are grown without feeders[71], and are expected to provide a labeled counterpart for a large share of the proteins present in early embryos, making them a very appropriate SILAC reference for our purposes.

LC-MS/MS analysis of SILAC mixtures. Subsequent to the tryptic digest, the peptide mixtures were offline fractionated by high pH reversed phase chromatography with fraction concatenation. The resulting peptide pools were analyzed by MS on a Q-Exactive mass spectrometer. The MS proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository[72] with the accession number PXD007082 and are summarized in Supplemental Table S6.

Basic processing of raw LC-MS/MS data (MaxQuant, Perseus). Raw data were processed by MaxQuant Software (v1.5.3.8, Martinsried, Bavaria, Germany) involving the built-in Andromeda search engine[73,74]. MS/MS spectra were searched against the mouse UniprotKB database (version from Dec. 2015) concatenated with reversed sequence versions of all entries and supplemented with common contaminants (see Supplemental Methods). Primary quantification was performed using the heavy F9 lysate mix as an internal standard, and ratios between corresponding light (L) and heavy (H) peptide versions were normalized to correct for unequal protein amounts and expressed as L/H (i.e., light/heavy: sample/SILAC internal standard). All these protein ratios are the means of at least two (light and heavy) peptide ratios from the raw spectra. Quality control determined that the sample corresponding to the blastocyst stage for replicate 3 was of low quality; this sample was therefore omitted from all analyses. The ID mapping procedure in some cases returned more than one gene name for a given peptide group; those may or may not correspond to distinct genes. To avoid ambiguities, we excluded such entries from the dataset.

Protein data normalization and batch correction. We log₂-transformed and quantile-normalized the L/H ratios of all proteins detected at least in two developmental stages in at least two replicates. To correct for the batch effect (see Supplemental Fig. S15), we performed an ANOVA for each protein, using the log₂-transformed L/H ratios as response variable and the replicate identifier as categorical explanatory variable X_i:

$$\log_2\frac{L}{H} = \mu + X_i + \epsilon$$

where μ is the global mean for the protein and ϵ denotes the error. The residuals of the model were used as the corrected L/H ratios for each protein, after adding to each value the global mean μ for the given protein as a constant. Batch-corrected, normalized L/H ratios were used to express protein abundance throughout this study.

RNA isolation and RNA sequencing. For the transcriptome analysis we collected and lysed, in duplicate, an average of 214 oocytes/embryos per developmental stage: unfertilized oocytes, fertilized oocytes with pronuclei and preimplantation embryos at the (early and late) 2-, 4-, 8-cell, advanced morula and blastocyst stages, on which we then performed RNA sequencing (RNA-seq). Total RNA was converted to cDNA using the Smarter system (Takara) and sequencing libraries were prepared using the Nextera kit (Illumina). Libraries were sequenced on Illumina HiSeq3000 platform to obtain ~43 million 36-base-single-end reads per library. The raw data are available at the DNA Databank of Japan (DDBJ) Sequence Read Archive (DRA005956 and DRA006335).

RNA-seq trimming and mapping. Low quality reads were filtered using Trimmomatic (version 0.36[75]) with the following parameters: HEADCROP:15 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15

MINLEN:20. The remaining reads were mapped to the *Mus musculus* Ensembl GRCm38 assembly using TopHat (version 2.1.1[76]) and Bowtie (version 2.2.9[77]). As the only non-default parameter for TopHat, we provided the GRCm38 Ensembl 87 (version 1) GTF annotation with the "-G" option. The number of reads mapped to each gene was quantified with with HTSeqCount (version 0.6.1[78]) using standard parameters.

RNA differential expression analysis. A matrix containing the number of reads mapped to each protein-coding gene for each sample was used as input for differential expression analysis with the DESeq2 R/Bioconductor package[79,80]. The P-values obtained from DESeq2 were adjusted with Benjamini-Hochberg's method to control the false discovery rate (FDR)[81]. Genes were considered significantly differentially expressed on the basis of ($\log_2$) fold-change (($\log_2$) fold-change ≥ 1 or ≤ -1 between the two developmental stages considered) and FDR $\leq 1 \times 10^{-5}$. Expression values of protein-coding transcripts were calculated using DESeq2 using the regularized log-transformation[79,80].

Protein differential expression analysis. For each protein detected at least in two developmental stages in at least two replicates we computed a linear model:

$$\log_2 \frac{L}{H} = \mu + T_i + \epsilon$$

where μ is the global mean for the gene, T_i is a categorical explanatory variable representing the developmental stage, and ϵ denotes the error. For 1,290 proteins, the ANOVA P-value corresponding to T_i was ≤ 0.05.

Validation of proteins by enzymatic assays and immunofluorescence. Results of enzymatic assays for G6PD (EC 1.1.1.49) and HPRT (EC 2.4.2.7) were retrieved from the literature[82–86].

Additional proteins including proteins without enzymatic activity were verified by immunofluorescence, using commercial antibodies. For each target gene, at least 5 MII oocytes or embryos per stage were examined using the following antibodies, all rabbit polyclonal: anti-DNAJB11 (Sigma-Aldrich cat.no. HPA010814), anti-PDIA3 (Abcam cat.no. ab228789), anti-TOP1 (Sigma-Aldrich cat.no. HPA019039), anti-Rc3h1 (Thermo Scientific catalog no. PA5-34519), anti-Alppl2 (Thermo Scientific catalog no. PA5-22336), anti-DDX6 (Thermo scientific catalog no. PA5-55012). Secondary antibodies were Alexa-Fluor conjugates reactive against the species of the primary antibody. Following our standard fixation, permeabilization, incubation and washing protocol[87], samples were imaged using a 20X objective on an inverted motorized Nikon TiE2000 microscope fitted with an Andor Dragonfly spinning disc confocal unit Scanning System. Immunofluorescent signals were quantified using Image-J[88].

For each protein, we calculated the Spearman's rank correlation coefficient between the immunofluorescent signals or enzymatic measurements and the average L/H ratios in our dataset for all available developmental stages. For proteins for which multiple sets of measurements were available we computed and considered as many correlation coefficients. An empirical P-value was computed by randomly associating each of the protein measurements from the literature with one of the corresponding sets of measurements in our dataset (see Supplemental Table S3) and repeating this 10,000 times. The reported empirical P-value is the number of times in which we obtained the same number of correlation coefficients greater or equal than 0.6 as with the original data out of the 10,000 attempts, expressed as a relative frequency.

TaqMan validation of RNAseq. For each target gene, the cDNA equivalent of 10 MII oocytes or embryos per stage was used. Total RNA was isolated from large pools (>100 oocytes or embryos) using Quick-RNA™ MicroPrep (Zymo Research) following the manufacturer's instructions and was reverse-transcribed on a GeneAmp® PCR System 9700 (Applied Biosystems). Real-time quantitative PCR reactions were performed on cDNA on a 7900 HT FAST Realtime PCR System (Applied Biosystems). PrimeTime® Predesigned qPCR Assay (6-FAM/ZEN/IBFQ) from Integrated DNA Technologies were used. Assay IDs: Dnajb11_Mm.PT.58.9272431, Pdia3_Mm.PT.8194853; Top1_Mm.PT.58.6752545. All samples were processed as technical duplicates/replicates. Data were analyzed using the Applied Biosystems RQ Manager (Version 1.2.2) and Microsoft Excel.

Data access. The proteomic data from this study have been submitted to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository[72] under accession number PXD007082. The sequence data generated for this study have been submitted to DNA Databank of Japan (DDBJ, http://www.ddbj.nig.ac.jp/) under the accession numbers DRA005956 and DRA006335.

References

1. Smits, A. H. *et al.* Global absolute quantification reveals tight regulation of protein expression in single Xenopus eggs. *Nucleic Acids Res* **42**, 9880–9891, https://doi.org/10.1093/nar/gku661 (2014).
2. Schwanhausser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342, https://doi.org/10.1038/nature10098 (2011).
3. Yurttas, P. *et al.* Role for PADI6 and the cytoplasmic lattices in ribosomal storage in oocytes and translational control in the early mouse embryo. *Development* **135**, 2627–2636, https://doi.org/10.1242/dev.016329 (2008).
4. Vinot, S. *et al.* Asymmetric distribution of PAR proteins in the mouse embryo begins at the 8-cell stage during compaction. *Dev Biol* **282**, 307–319, https://doi.org/10.1016/j.ydbio.2005.03.001 (2005).
5. Ohsugi, M., Zheng, P., Baibakov, B., Li, L. & Dean, J. Maternally derived FILIA-MATER complex localizes asymmetrically in cleavage-stage mouse embryos. *Development* **135**, 259–269, https://doi.org/10.1242/dev.011445 (2008).
6. Li, L., Baibakov, B. & Dean, J. A subcortical maternal complex essential for preimplantation mouse embryogenesis. *Dev Cell* **15**, 416–425, https://doi.org/10.1016/j.devcel.2008.07.010 (2008).
7. Coonrod, S. *et al.* Testis-specific lactate dehydrogenase (LDH-C4; Ldh3) in murine oocytes and preimplantation embryos. *J Androl* **27**, 502–509, https://doi.org/10.2164/jandrol.05185 (2006).

8. Sun, L. *et al*. Quantitative proteomics of Xenopus laevis embryos: expression kinetics of nearly 4000 proteins during early development. *Sci Rep* **4**, 4365, https://doi.org/10.1038/srep04365 (2014).

9. Kronja, I. *et al*. Quantitative proteomics reveals the dynamics of protein changes during Drosophila oocyte maturation and the oocyte-to-embryo transition. *Proc Natl Acad Sci USA* **111**, 16023–16028, https://doi.org/10.1073/pnas.1418657111 (2014).

10. Casas-Vila, N. *et al*. The developmental proteome of Drosophila melanogaster. *Genome research* **27**, 1273–1285, https://doi.org/10.1101/gr.213694.116 (2017).

11. Demant, M., Deutsch, D. R., Frohlich, T., Wolf, E. & Arnold, G. J. Proteome analysis of early lineage specification in bovine embryos. *Proteomics* **15**, 688–701, https://doi.org/10.1002/pmic.201400251 (2015).

12. Deutsch, D. R. *et al*. Stage-specific proteome signatures in early bovine embryo development. *J Proteome Res* **13**, 4363–4376, https://doi.org/10.1021/pr500550t (2014).

13. Wang, S. *et al*. Proteome of mouse oocytes at different developmental stages. *Proc Natl Acad Sci USA* **107**, 17639–17644, https://doi.org/10.1073/pnas.1013185107 (2010).

14. Fu, Z. *et al*. Integral proteomic analysis of blastocysts reveals key molecular machinery governing embryonic diapause and reactivation for implantation in mice. *Biol Reprod* **90**, 52, https://doi.org/10.1095/biolreprod.113.115337 (2014).

15. Gao, Y. *et al*. Protein Expression Landscape of Mouse Embryos during Pre-implantation Development. *Cell Rep* **21**, 3957–3969, https://doi.org/10.1016/j.celrep.2017.11.111 (2017).

16. Hikabe, O. *et al*. Reconstitution *in vitro* of the entire cycle of the mouse female germ line. *Nature* **539**, 299–303, https://doi.org/10.1038/nature20104 (2016).

17. Pfeiffer, M. J. *et al*. Proteomic analysis of mouse oocytes reveals 28 candidate factors of the "reprogrammome". *J Proteome Res* **10**, 2140–2153, https://doi.org/10.1021/pr100706k (2011).

18. Schwarzer, C. *et al*. Maternal age effect on mouse oocytes: new biological insight from proteomic analysis. *Reproduction* **148**, 55–72, https://doi.org/10.1530/REP-14-0126 (2014).

19. Pfeiffer, M. J. *et al*. Differences in embryo quality are associated with differences in oocyte composition: a proteomic study in inbred mice. *Proteomics* **15**, 675–687, https://doi.org/10.1002/pmic.201400334 (2015).

20. Wang, B., Pfeiffer, M. J., Drexler, H. C., Fuellen, G. & Boiani, M. Proteomic Analysis of Mouse Oocytes Identifies PRMT7 as a Reprogramming Factor that Replaces SOX2 in the Induction of Pluripotent Stem Cells. *J Proteome Res* **15**, 2407–2421, https://doi.org/10.1021/acs.jproteome.5b01083 (2016).

21. Van der Auwera, I. & D'Hooghe, T. Superovulation of female mice delays embryonic and fetal development. *Hum Reprod* **16**, 1237–1243, https://doi.org/10.1093/humrep/16.6.1237 (2001).

22. Ertzeid, G. & Storeng, R. The impact of ovarian stimulation on implantation and fetal development in mice. *Hum Reprod* **16**, 221–225, https://doi.org/10.1093/humrep/16.2.221 (2001).

23. Cech, S. & Sedlackova, M. Ultrastructure and morphometric analysis of preimplantation mouse embryos. *Cell Tissue Res* **230**, 661–670, https://doi.org/10.1007/bf00216209 (1983).

24. Bachvarova, R., De Leon, V. & Spiegelman, I. Mouse egg ribosomes: evidence for storage in lattices. *J Embryol Exp Morphol* **62**, 153–164 (1981).

25. Piko, L. & Clegg, K. B. Quantitative changes in total RNA, total poly(A), and ribosomes in early mouse embryos. *Dev Biol* **89**, 362–378, https://doi.org/10.1016/0012-1606(82)90325-6 (1982).

26. van Blerkom, J. & Brockway, G. O. Qualitative patterns of protein synthesis in the preimplantation mouse embryo. I. Normal pregnancy. *Dev Biol* **44**, 148–157, https://doi.org/10.1016/0012-1606(75)90382-6 (1975).

27. Kidder, G. M. & McLachlin, J. R. Timing of transcription and protein synthesis underlying morphogenesis in preimplantation mouse embryos. *Dev Biol* **112**, 265–275, https://doi.org/10.1016/0012-1606(85)90397-5 (1985).

28. Latham, K. E., Garrels, J. I., Chang, C. & Solter, D. Quantitative analysis of protein synthesis in mouse embryos. I. Extensive reprogramming at the one- and two-cell stages. *Development* **112**, 921–932 (1991).

29. Ori, A. *et al*. Spatiotemporal variation of mammalian protein complex stoichiometries. *Genome Biol* **17**, 47, https://doi.org/10.1186/s13059-016-0912-5 (2016).

30. Kidder, G. M. The genetic program for preimplantation development. *Dev Genet* **13**, 319–325, https://doi.org/10.1002/dvg.1020130502 (1992).

31. Levinson, J., Goodfellow, P., vadeboncoeur, M. & McDevitt, H. Identification of stage-specific polypeptides synthesized during murine preimplantation development. *Proc Natl Acad Sci USA* **75**, 3332–3336, https://doi.org/10.1073/pnas.75.7.3332 (1978).

32. Dumollard, R., Ward, Z., Carroll, J. & Duchen, M. R. Regulation of redox metabolism in the mouse oocyte and embryo. *Development* **134**, 455–465, https://doi.org/10.1242/dev.02744 (2007).

33. Saadeldin, I. M., Kim, S. J., Choi, Y. B. & Lee, B. C. Improvement of cloned embryos development by co-culturing with parthenotes: a possible role of exosomes/microvesicles for embryos paracrine communication. *Cell Reprogram* **16**, 223–234, https://doi.org/10.1089/cell.2014.0003 (2014).

34. Giacomini, E. *et al*. Secretome of *in vitro* cultured human embryos contains extracellular vesicles that are uptaken by the maternal side. *Sci Rep* **7**, 5210, https://doi.org/10.1038/s41598-017-05549-w (2017).

35. Hu, J. *et al*. Mouse ZAR1-like (XM_359149) colocalizes with mRNA processing components and its dominant-negative mutant caused two-cell-stage embryonic arrest. *Dev Dyn* **239**, 407–424, https://doi.org/10.1002/dvdy.22170 (2010).

36. Vinuesa, C. G. *et al*. A RING-type ubiquitin ligase family member required to repress follicular helper T cells and autoimmunity. *Nature* **435**, 452–458, https://doi.org/10.1038/nature03555 (2005).

37. Johnson, L. V., Calarco, P. G. & Siebert, M. L. Alkaline phosphatase activity in the preimplantation mouse embryo. *J Embryol Exp Morphol* **40**, 83–89 (1977).

38. Hahnel, A. C. *et al*. Two alkaline phosphatase genes are expressed during early development in the mouse embryo. *Development* **110**, 555–564 (1990).

39. Bai, Q. *et al*. Dissecting the first transcriptional divergence during human embryonic development. *Stem Cell Rev* **8**, 150–162, https://doi.org/10.1007/s12015-011-9301-3 (2012).

40. Wang, Q. T. *et al*. A genome-wide study of gene activity reveals developmental signaling pathways in the preimplantation mouse embryo. *Dev Cell* **6**, 133–144, https://doi.org/10.1016/S1534-5807(03)00404-0 (2004).

41. Hamatani, T., Carter, M. G., Sharov, A. A. & Ko, M. S. Dynamics of global gene expression changes during mouse preimplantation development. *Dev Cell* **6**, 117–131, https://doi.org/10.1016/S1534-5807(03)00373-3 (2004).

42. Zeng, F., Baldwin, D. A. & Schultz, R. M. Transcript profiling during preimplantation mouse development. *Dev Biol* **272**, 483–496, https://doi.org/10.1016/j.ydbio.2004.05.018 (2004).

43. Houghton, F. D. Energy metabolism of the inner cell mass and trophectoderm of the mouse blastocyst. *Differentiation* **74**, 11–18, https://doi.org/10.1111/j.1432-0436.2006.00052.x (2006).

44. MacQueen, H. A. & Johnson, M. H. The fifth cell cycle of the mouse embryo is longer for smaller cells than for larger cells. *J Embryol Exp Morphol* **77**, 297–308 (1983).

45. Ko, M. S. *et al*. Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development. *Development* **127**, 1737–1749 (2000).

46. Li, J. *et al*. Identification and characterization of an oocyte factor required for sperm decondensation in pig. *Reproduction* **148**, 367–375, https://doi.org/10.1530/REP-14-0264 (2014).

47. Wang, Y. *et al.* Impaired bone formation in Pdia3 deficient mice. *PLoS One* **9**, e112708, https://doi.org/10.1371/journal.pone.0112708 (2014).
48. Morham, S. G., Kluckman, K. D., Voulomanos, N. & Smithies, O. Targeted disruption of the mouse topoisomerase I gene by camptothecin selection. *Mol Cell Biol* **16**, 6804–6809, https://doi.org/10.1128/mcb.16.12.6804 (1996).
49. Francisco, A. B. *et al.* Deficiency of suppressor enhancer Lin12 1 like (SEL1L) in mice leads to systemic endoplasmic reticulum stress and embryonic lethality. *J Biol Chem* **285**, 13694–13703, https://doi.org/10.1074/jbc.M109.085340 (2010).
50. Yu, C. *et al.* Oocyte-expressed yes-associated protein is a key activator of the early zygotic genome in mouse. *Cell Res* **26**, 275–287, https://doi.org/10.1038/cr.2016.20 (2016).
51. Dai, N. *et al.* IGF2BP2/IMP2-Deficient mice resist obesity through enhanced translation of Ucp1 mRNA and Other mRNAs encoding mitochondrial proteins. *Cell Metab* **21**, 609–621, https://doi.org/10.1016/j.cmet.2015.03.006 (2015).
52. Forrest, A. R. *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462–470, https://doi.org/10.1038/nature13182 (2014).
53. Chen, H. *et al.* Embryonic arrest at midgestation and disruption of Notch signaling produced by the absence of both epsin 1 and epsin 2 in mice. *Proc Natl Acad Sci USA* **106**, 13838–13843, https://doi.org/10.1073/pnas.0907008106 (2009).
54. Decker, C. J. & Parker, R. P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. *Cold Spring Harb Perspect Biol* **4**, a012286, https://doi.org/10.1101/cshperspect.a012286 (2012).
55. Lin, X. *et al.* PPM1A functions as a Smad phosphatase to terminate TGFbeta signaling. *Cell* **125**, 915–928, https://doi.org/10.1016/j.cell.2006.03.044 (2006).
56. Wu, L. S., Qian, J. Y., Wang, M. & Yang, H. Identifying the role of Wilms tumor 1 associated protein in cancer prediction using integrative genomic analyses. *Mol Med Rep* **14**, 2823–2831, https://doi.org/10.3892/mmr.2016.5528 (2016).
57. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44**, D457–462, https://doi.org/10.1093/nar/gkv1070 (2016).
58. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**, D353–D361, https://doi.org/10.1093/nar/gkw1092 (2017).
59. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27–30, https://doi.org/10.1093/Fnar/F28.1.27 (2000).
60. Kim, B. *et al.* The role of MATER in endoplasmic reticulum distribution and calcium homeostasis in mouse oocytes. *Dev Biol* **386**, 331–339, https://doi.org/10.1016/j.ydbio.2013.12.025 (2014).
61. Michalak, M. & Gye, M. C. Endoplasmic reticulum stress in periimplantation embryos. *Clin Exp Reprod Med* **42**, 1–7, https://doi.org/10.5653/cerm.2015.42.1.1 (2015).
62. Plaks, V. *et al.* Blastocyst implantation failure relates to impaired translational machinery gene expression. *Reproduction* **148**, 87–98, https://doi.org/10.1530/REP-13-0395 (2014).
63. Hogan, D. J., Riordan, D. P., Gerber, A. P., Herschlag, D. & Brown, P. O. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS Biol* **6**, e255, https://doi.org/10.1371/journal.pbio.0060255 (2008).
64. Peshkin, L. *et al.* On the Relationship of Protein and mRNA Dynamics in Vertebrate Embryonic Development. *Dev Cell* **35**, 383–394, https://doi.org/10.1016/j.devcel.2015.10.010 (2015).
65. Casser, E. *et al.* Totipotency segregates between the sister blastomeres of two-cell stage mouse embryos. *Sci Rep* **7**, 8299, https://doi.org/10.1038/s41598-017-08266-6 (2017).
66. Ho, Y., Wigglesworth, K., Eppig, J. J. & Schultz, R. M. Preimplantation development of mouse embryos in KSOM: augmentation by amino acids and analysis of gene expression. *Mol Reprod Dev* **41**, 232–238, https://doi.org/10.1002/mrd.1080410214 (1995).
67. Berstine, E. G., Hooper, M. L., Grandchamp, S. & Ephrussi, B. Alkaline phosphatase activity in mouse teratoma. *Proc Natl Acad Sci USA* **70**, 3899–3903, https://doi.org/10.1073/Fpnas.70.12.3899 (1973).
68. Pierce, G. B. Neoplasms, differentiations and mutations. *Am J Pathol* **77**, 103–118 (1974).
69. Alonso, A., Breuer, B., Steuer, B. & Fischer, J. The F9-EC cell line as a model for the analysis of differentiation. *Int J Dev Biol* **35**, 389–397 (1991).
70. Chen, Y., Du, Z. & Yao, Z. Roles of the Nanog protein in murine F9 embryonal carcinoma cells and their endoderm-differentiated counterparts. *Cell Res* **16**, 641–650, https://doi.org/10.1038/sj.cr.7310067 (2006).
71. Rizzino, A. & Sato, G. Growth of embryonal carcinoma cells in serum-free medium. *Proc Natl Acad Sci USA* **75**, 1844–1848, https://doi.org/10.1073/pnas.75.4.1844 (1978).
72. Vizcaino, J. A. *et al.* The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res* **41**, D1063–1069, https://doi.org/10.1093/nar/gks1262 (2013).
73. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367–1372, https://doi.org/10.1038/nbt.1511 (2008).
74. Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **10**, 1794–1805, https://doi.org/10.1021/pr101065j (2011).
75. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120, https://doi.org/10.1093/bioinformatics/btu170 (2014).
76. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36, https://doi.org/10.1186/gb-2013-14-4-r36 (2013).
77. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359, https://doi.org/10.1038/nmeth.1923 (2012).
78. Anders, S., Pyl, P. T. & Huber, W. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169, https://doi.org/10.1093/bioinformatics/btu638 (2015).
79. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106, https://doi.org/10.1186/gb-2010-11-10-r106 (2010).
80. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq. 2. *Genome Biol* **15**, 550, https://doi.org/10.1186/s13059-014-0550-8 (2014).
81. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B* **57**, 289–300 (1995).
82. Brinster, R. L. Glucose 6-phosphate-dehydrogenase activity in the preimplantation mouse embryo. *Biochem J* **101**, 161–163, https://doi.org/10.1042/bj1010161 (1966).
83. Ayabe, T., Tsutsumi, O. & Taketani, Y. Hexokinase activity in mouse embryos developed *in vivo* and *in vitro*. *Hum Reprod* **9**, 347–351, https://doi.org/10.1093/oxfordjournals.humrep.a138506 (1994).
84. Epstein, C. J. Phosphoribosyltransferase activity during early mammalian development. *J Biol Chem* **245**, 3289–3294 (1970).
85. Epstein, C. J., Wegienka, E. A. & Smith, C. W. Biochemical development of preimplantation mouse embryos: *in vivo* activities of fructose 1,6-diphosphate aldolase, glucose 6-phosphate dehydrogenase, malate dehydrogenase, and lactate dehydrogenase. *Biochem Genet* **3**, 271–281, https://doi.org/10.1007/BF00521142 (1969).
86. Kratzer, P. G. & Gartler, S. M. HGPRT activity changes in preimplantation mouse embryos. *Nature* **274**, 503–504, https://doi.org/10.1038/274503a0 (1978).

14

87. Schwarzer, C. *et al*. ART culture conditions change the probability of mouse embryo gestation through defined cellular and molecular responses. *Hum Reprod* **27**, 2627–2640, https://doi.org/10.1093/humrep/des223 (2012).
88. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671–675, https://doi.org/10.1038/Fnmeth.2089 (2012).
89. Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800, https://doi.org/10.1371/journal.pone.0021800 (2011).
90. R package "corrplot": Visualization of a Correlation Matrix (Version 0.84) (2017).
91. pheatmap: Pretty Heatmaps (2016).
92. Mi, H. *et al*. PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res* **38**, D204–210, https://doi.org/10.1093/nar/gkp1019 (2010).
93. Thomas, P. D. *et al*. PANTHER: a library of protein families and subfamilies indexed by function. *Genome research* **13**, 2129–2141, https://doi.org/10.1101/gr.772403 (2003).
94. Luo, W. & Brouwer, C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **29**, 1830–1831, https://doi.org/10.1093/bioinformatics/btt285 (2013).

Acknowledgements

Author Contributions

L.T., G.F. and M.B. planned the study. H.C.D. performed the proteomics experiments. Y.S. performed the RNA-seq experiments. S.I., M.E., L.T., G.F. and M.B. designed the analytical pipeline, analyzed and interpreted the data. L.T., G.F. and M.B. wrote the manuscript with help from S.I. and M.E. O.E.P. performed the transmission electron microscopy. S.I. and E.C. performed the confocal immunofluorescence imaging. W.M. provided intellectual guidance with RNA-seq analysis and feedback on the experimental design. All authors discussed the results and commented on the manuscript.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.