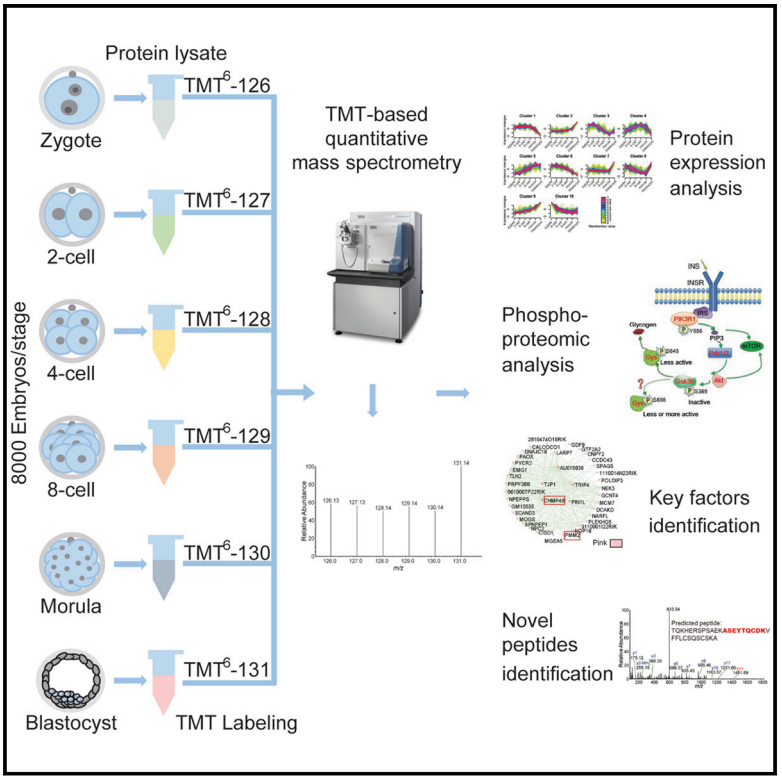
Resource

Cell Reports

Protein Expression Landscape of Mouse Embryos during Pre-implantation Development

Graphical Abstract



Authors

Yawei Gao, Xiaoyu Liu, Bin Tang, ..., She Chen, Lujian Liao, Shaorong Gao

Correspondence

ljliao@bio.ecnu.edu.cn (L.L.), gaoshaorong@tongji.edu.cn (S.G.)

In Brief

Proteomic analysis of pre-implantation embryos has not been explored for the minimal amount of material available. Gao et al. used a quantitative mass spectrometry strategy to elucidate the dynamic changes of the embryo proteome from the zygote to the blastocyst stage, which provides direct insight into the molecular details governing embryonic development.

Highlights

- Comprehensive quantitative proteomic study of mouse pre-implantation embryos
- Phosphoproteomic analysis identified key kinases and important signaling pathways
- Identified poorly studied key factors in early embryos and showed functional importance
- Proteogenomic analysis identified previously undefined exon-junction-derived peptides

Data and Software Availability

PXD003315
GSE70605
GSE106612

Gao et al., 2017, Cell Reports *21*, 3957–3969
December 26, 2017
https://doi.org/10.1016/j.celrep.2017.11.111

CellPress

Cell Reports
Resource

Protein Expression Landscape of Mouse Embryos during Pre-implantation Development

Yawei Gao,[1,4] Xiaoyu Liu,[1,4] Bin Tang,[2,4] Chong Li,[1,4] Zhaohui Kou,[3] Lin Li,[3] Wenqiang Liu,[1] You Wu,[1] Xiaochen Kou,[1] Jingyi Li,[1] Yanhong Zhao,[1] Jiqing Yin,[1] Hong Wang,[1] She Chen,[3] Lujian Liao,[2,*] and Shaorong Gao[1,5,*]

[1]Clinical and Translational Research Center of Shanghai First Maternity and Infant Hospital, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China

[2]Shanghai Key Laboratory of Regulatory Biology, School of Life Sciences, East China Normal University, Shanghai 200241, China

[3]National Institute of Biological Sciences, NIBS, Beijing 102206, China

[4]These authors contributed equally

[5]Lead Contact

*Correspondence: ljliao@bio.ecnu.edu.cn (L.L.), gaoshaorong@tongji.edu.cn (S.G.)

https://doi.org/10.1016/j.celrep.2017.11.111

SUMMARY

Pre-implantation embryo development is an intricate and precisely regulated process orchestrated by maternally inherited proteins and newly synthesized proteins following zygotic genome activation. Although genomic and transcriptomic studies have enriched our understanding of the genetic programs underlying this process, the protein expression landscape remains unexplored. Using quantitative mass spectrometry, we identified nearly 5,000 proteins from 8,000 mouse embryos of each stage (zygote, 2-cell, 4-cell, 8-cell, morula, and blastocyst). We found that protein expression in zygotes, morulas, and blastocysts is distinct from 2- to 8-cell embryos. Analysis of protein phosphorylation identified critical kinases and signal transduction pathways. We highlight key factors and their important roles in embryo development. Combined analysis of transcriptomic and proteomic data reveals coordinated control of RNA degradation, transcription, and translation and identifies previously undefined exon-junction-derived peptides. Our study provides an invaluable resource for further mechanistic studies and suggests core factors regulating pre-implantation embryo development.

INTRODUCTION

Pre-implantation development encompasses the period from fertilization to implantation. After fertilization, the embryo is transcriptionally quiescent, and development is controlled exclusively by maternally provided proteins and RNAs from the egg cytoplasm (Wassarman and Kinloch, 1992; Xue et al., 2013). Subsequently, the newly formed zygotic genome is activated, and maternal factors are subsequently degraded. The switch of this developmental control is called zygotic gene activation (ZGA). In mice, ZGA initiates during the late 1-cell stage (minor ZGA), followed by major gene activation (major ZGA) at the 2-cell stage. (Kanka, 2003; Xue et al., 2013). After the 8-cell stage, blastomeres begin to bind firmly together in a process called compaction, and a tightly organized cell mass of morula is generated. Then, the morula differentiates into a blastocyst composed of the inner cell mass (ICM) and the trophectoderm (TE). Although great efforts have been made in understanding the molecular mechanisms of this process (Leung and Zernicka-Goetz, 2015; Paranjpe and Veenstra, 2015), there are a lot of questions that are yet to be uncovered.

Recent analysis of the transcriptome, the epigenome, and the even-higher-order chromatin structure using single-cell or micro-cell sequencing has greatly expanded our understanding of the regulation mechanism during early embryo development (Dahl et al., 2016; Flyamer et al., 2017; Guo et al., 2013; Liu et al., 2016; Lu et al., 2016; Wu et al., 2016; Xue et al., 2013; Zhang et al., 2016). Proteins are the executors of most biological programs; however, the abundance of proteins cannot be accurately predicted from transcript abundance (Schwanhausser et al., 2013), and the biochemical function of a protein frequently requires post-translational modifications (PTMs) (Snider and Omary, 2014). Therefore, identifying the temporal proteome landscape in early embryos will provide direct insight into the molecular details governing embryonic development.

Mass spectrometry (MS) is a good option to detect and quantitate a large number of proteins and PTMs. There have been a few reports of using MS strategies to identify/quantify proteins in models of embryonic development, including bovine (Deutsch et al., 2014), zebrafish (Lucitt et al., 2008), and *Xenopus* (Peshkin et al., 2015; Wühr et al., 2014) models. The mouse is the most widely used model organism for studying mammalian embryo development, the proteomic analysis of pre-implantation embryos has been technologically challenging due to the scarcity of required materials. Quantitative MS technologies have been developed over the past few years (Ong and Mann, 2005; Zhou et al., 2013). In the current study, we used a tandem mass tag (TMT)-based quantitative MS strategy (Thompson et al., 2003) to elucidate the dynamic changes of the embryo proteome during mouse early development from the zygote to the balstocyst stage. Our work provides a comprehensive analysis of the

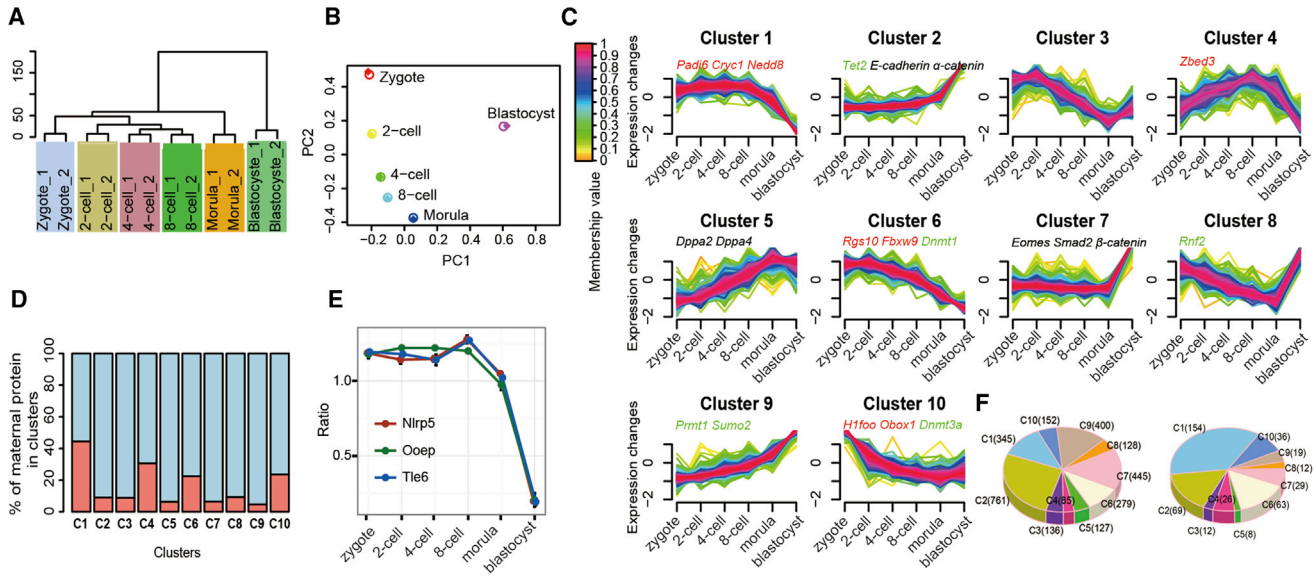


Figure 1. Temporal Profiles of Protein Expression in Early Embryos

(A) Unsupervised clustering of protein expression in six embryonic stages analyzed by quantitative mass spectrometry in biological duplicate.

(B) Principal component analysis of protein expression patterns. The same color with two different shapes represents different replicates of the same stage.

(C) Fuzzy c-means clustering identified ten distinct temporal patterns of proteins expression. The x axis represents six developmental stages, while the y axis represents log2-transformed, normalized intensity ratios in each stage.

(D) The salmon-colored bars show the percentage of maternal proteins in each cluster.

(E) Expression patterns of quantified components in the subcortical maternal complex.

(F) The left pie chart shows the number of proteins in each cluster in C, while the right pie chart shows the number of maternal proteins in each cluster in (C). See also Table S1.

dynamic protein expression and PTMs in mouse embryos during pre-implantation development.

RESULTS

A Dynamic Protein Expression Landscape of Early Embryos

Using TMT-based quantitative MS, we systematically monitored protein expression profiles in six pre-implantation embryonic stages in mice: zygote, 2-, 4-, and 8-cell embryo, morula, and blastocyst. For each stage, 8,000 embryos were used, and the experiment was performed in biological duplicate. In total, 4,608 and 4,590 proteins were quantified in each of the two experiments, and the two experiments had 3,767 proteins in common (Figure S1A; Table S1). All peptides identified in the two datasets are of very high mass accuracy (within 10 parts per million) and have their distributions centered near zero (Figure S1B). The distribution of protein sequence coverage is also highly similar between the two replicate analyses (Figure S1C). We assessed the reproducibility of the quantification by plotting the peptide intensities between biological replicates in each of the six stages and found that the Pearson's correlation coefficient for every stage is over 0.8, indicating very consistent measurement (Figure S1D). Unsupervised hierarchical clustering of the protein expression profiles revealed that the 4-cell and 8-cell embryos closely clustered together, and the cluster expanded to the 2-cell embryo and then to the zygote. The morula and the blastocyst were further separated (Figure 1A). The same trend can be recapitulated by principal component analysis (PCA), a completely different clustering method (Figure 1B). Of note, both methods revealed that the blastocyst differed the most from the rest of the developmental stages, indicating that the protein expression profile encountered a major shift in the blastocyst stage.

We applied the fuzzy c-means algorithm (Futschik and Carlisle, 2005) to cluster protein expression profiles in all developmental stages. In total, we observed 10 distinct clusters of temporal patterns representing proteins that are regulated differently (Figure 1C; Table S1), indicating different expression kinetics. Among these, clusters 1, 6, and 10 represent proteins that are downregulated, clusters 2, 5, and 9 represent proteins that are upregulated, whereas clusters 3, 4, 7, and 8 represent proteins displaying a bi-modal expression pattern. Overlaying our previously published oocyte proteomic data containing primarily maternal proteins (Wang et al., 2010), we found, intriguingly, that maternal proteins are mainly classified within clusters that show a low expression level in the blastocyst stage with distinct kinetics (clusters 1, 4, 6, and 10, Figures 1D and 1F). These proteins include many defined maternal factors that play critical roles regulating ZGA, such as MATER, PADI6, NPM2, ZFP36L2, ATG5, ZAR1, etc. (Xue et al., 2013). Our experiments captured three (MATER, FLOPED, and TLE6) of four known components of the subcortical maternal complex (SCMC), a protein structure that is essential for pre-implantation development(Li et al., 2008), all of which fall into cluster 1 (Figure 1E). This is consistent with the idea that these maternal proteins are

essential for murine zygote progression beyond the first embryonic cell divisions and degraded after ZGA (Bebbere et al., 2016). Some known maternal factors, such as AGO2, are consistently upregulated (cluster 2). RNF2 is continuously downregulated until the morula stage and then upregulated in the blastocyst stage (cluster 8). These observations indicate that although these proteins are of maternal origin, their expression is tightly regulated and their destruction is induced in a timely manner, probably due to different functionalities. Transcription factors that have been identified to regulate early embryo development and cell fate determination were also captured in our data. The embryo undergoes compaction after the 8-cell stage (Riethmacher et al., 1995). E-cadherin and α-catenin, which are responsible for the formation of adherent junctions, were consistently upregulated (cluster 2). The key transcriptional regulators of ICM/TE fate, such as Oct4 and Eomes, were also upregulated in the morula stage and the blastocyst stage, respectively. Gene ontology (GO) analysis of proteins from each cluster (Figure 2A) revealed that the upregulated proteins tend to execute functions, including protein translation, macromolecule metabolism, RNA metabolism, processing, and splicing (clusters 2 and 9), presumably providing energy and materials to support the basic requirements for embryonic growth. On the other hand, the downregulated proteins fall into the categories of protein transport, protein localization, cell cycle, and gamete generation (clusters 1, 6, and 10), suggesting that these molecular events underlie the maternal-to-zygote transition (Xue et al., 2013). The GO terms in the categories "germ cell development" and "translational initiation" are further expanded, with the proteins in each term listed to provide an elaborated view of the proteins participating in these molecular functions (Figures 2B and 2C).

Phosphoproteomic Analysis Identifies Key Kinases and Important Signal Transduction Pathways

We speculate that signal transduction through protein phosphorylation and other post-translational modifications may play an important role during early embryo development, adding more complexity to the signal flow in addition to transcriptional control (Krauchunas et al., 2012). Since strong-cation-exchange fractionation separates peptides based on charge, it has the tendency to enrich phosphopeptides in early fractions (Beausoleil et al., 2004). Based on this, we performed another database search for serine, threonine, and tyrosine phosphorylation as well as methylation, acetylation, and ubiquitination in lysine, even though the low protein amount prevented us from enriching modified peptides before MS analysis. Surprisingly, this analysis identified 1,028 phosphopeptides derived from 650 proteins and 199 lysine modifications, indicating widespread activation of phosphorylation events in early embryos (Tables S2 and S3). The distribution of single-, double-, and triple-phosphorylated peptides, as well as that of serine, threonine, and tyrosine phosphorylation sites (Figure 3A), is consistent with several previously published large-scale stem cell phosphoproteomic studies (Brill et al., 2009; Rigbolt et al., 2011; Van Hoof et al., 2009). The temporal profiles of the changed phosphorylation sites can be grouped into four clusters (Figure S2A), including an increase in 2-cell embryo and a decrease in blastocyst (cluster 1), a gradual decrease (cluster 2 and 3), and an increase only in blastocyst (cluster 4), and GO analysis revealed that clusters 2 and 4 contained proteins that were significantly enriched in biological pathways. The proteins in cluster 2 were enriched in the regulation of cytoskeletal organization, whereas the proteins in cluster 4 were involved in RNA processing and splicing (Figure S2B). The changed phosphorylation sites were further analyzed using Motif-X to extract common amino acid sequence motifs surrounding the phosphorylation sites (Schwartz and Gygi, 2005). This resulted in six motifs that were mostly proline directed, and the kinases that phosphorylate these sites were predicted (Figure 3B). Interestingly, JNK2, CDK2, BARK1 (also known as GRK2), and ERK1 scored as the most probable kinases responsible for the changed phosphorylation events, and our quantitative proteome results show that both ERK1 (MAPK1) and CDK1 maintain high expression throughout the six embryonic stages. In total, we identified 107 phoshopeptides as JNK1 or ERK1 substrates that mapped to 95 proteins (Table S4). Incidentally, the inhibition of JNK or p38 MAP kinase results in inhibition of cavity formation in blastocyst (Maekawa et al., 2005), and inhibition of ERK in late 2-cell embryos results in G2 arrest at the 4-cell stage (Maekawa et al., 2007).

We further investigated several highly conserved signal transduction pathways. For example, we found phosphorylation changes in multiple key enzymes mediating glycogenesis downstream of insulin signaling (Figure 3C). At the protein level, glycogen synthase (Gys) remained stable until the 8-cell stage and then was rapidly degraded in the morula and blastocyst stage (Figure 3D). We found two serine phosphorylation sites on Gys: S645 is known to be phosphorylated by GSK3β to inhibit the syntheses of glycogen and reciprocally correlated with a phosphorylation site on GSK3β (S389), which is also found in our data, and S698 is a newly identified site and exhibits different dynamics from S645 (Figure 3D). In the early study of pre-implantation mouse embryos *in vivo*, the content of glycogen was found to be of a high level from the MII to the morula stage, but decreased during the growth of the blastocyst (Thomson and Brinster, 1966). Our data here provide direct evidence for the enzymatic activity of glycogen synthetase in mouse embryo development and suggest a possible role for the insulin signaling pathway in the regulation of glycogen metabolism. We further found that, similar to Gys, key enzymes in glycogenesis are gradually degraded, and those in glycogenolysis are gradually increased (Ozias and Stern, 1973), whereas Pgm1/2, an enzyme used in both processes, remains stable (Figure 3E). Our data therefore provide clues to explore the regulation mechanism of glycogen levels in mouse embryogenesis.

We also found an additional phosphorylation site (T551) on β-catenin, a key player in the Wnt signaling pathway (Figures S2C and S2D). The most probable kinase for this site is CK1; interestingly, the protein expression pattern of CK1 shares striking similarity with that of T551 phosphorylation on β-catenin (Figures S2E and S2F). Taken together, our analyses suggest that kinases play important roles in governing early embryo development through substrate phosphorylation and orchestrating signal transduction and that our data could be valuable for further mechanistic study of these events.

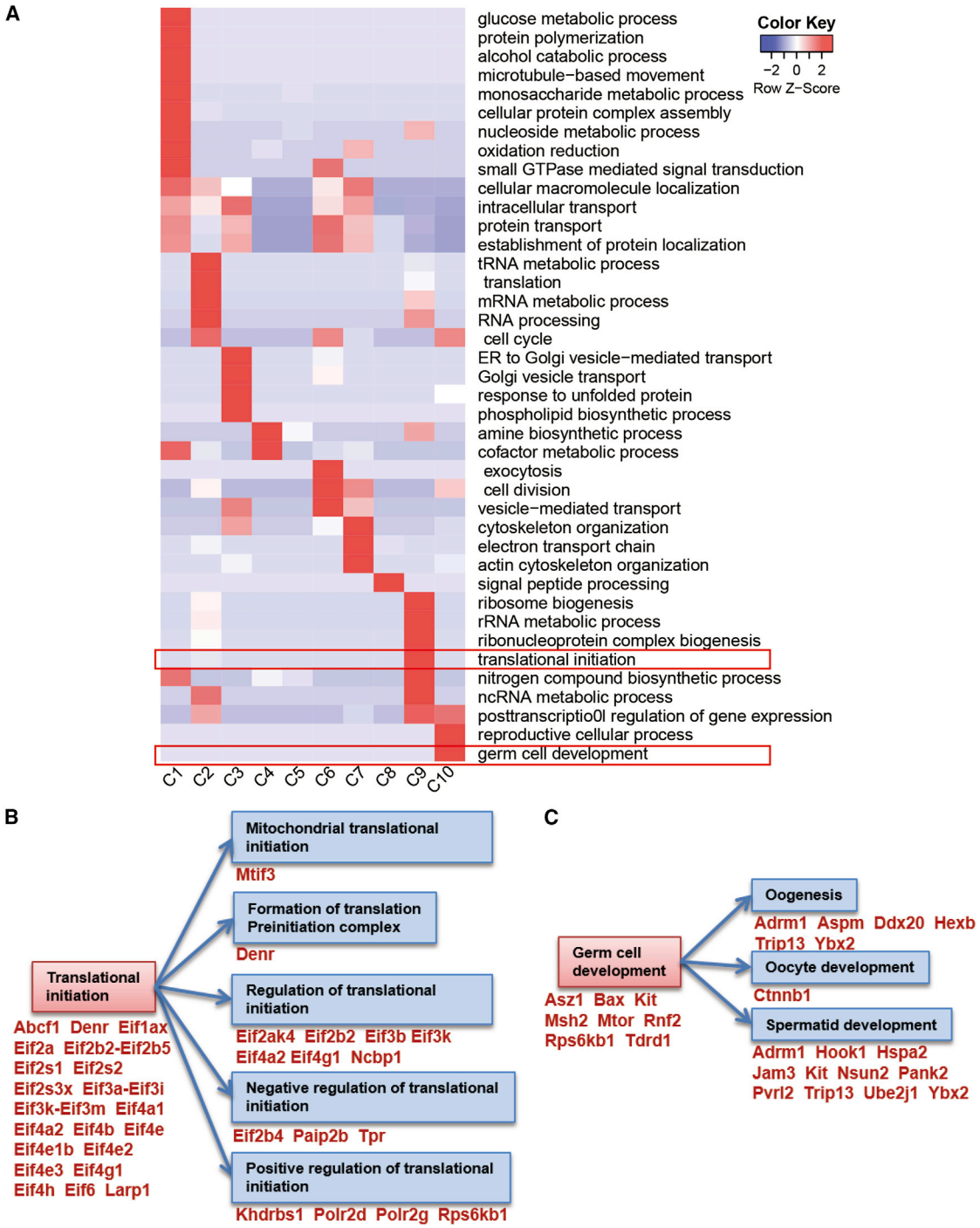


Figure 2. Gene Ontology Analysis of Protein Expression Clusters

(A) The heat map shows the –log10-transformed significance value of the gene ontology terms in the biological processes describing each of the 10 clusters. (B and C) Expanded gene ontology terms and proteins from "translational initiation" (B) and "germ cell development" (C) are displayed.

Network Analysis Identifies Inter-related Functional Modules and Points to Core Mediators

We applied weighted gene co-expression network analysis (WGCNA) (Langfelder and Horvath, 2008) to obtain a system-wide understanding of groups of proteins whose co-expression patterns are highly correlated during embryonic development. The hierarchical clustering dendrogram in Figure 4A represents co-expressed proteins that are highly correlated, and the relative expression levels of these proteins are represented by a heat map. In total eight distinct co-expression modules were

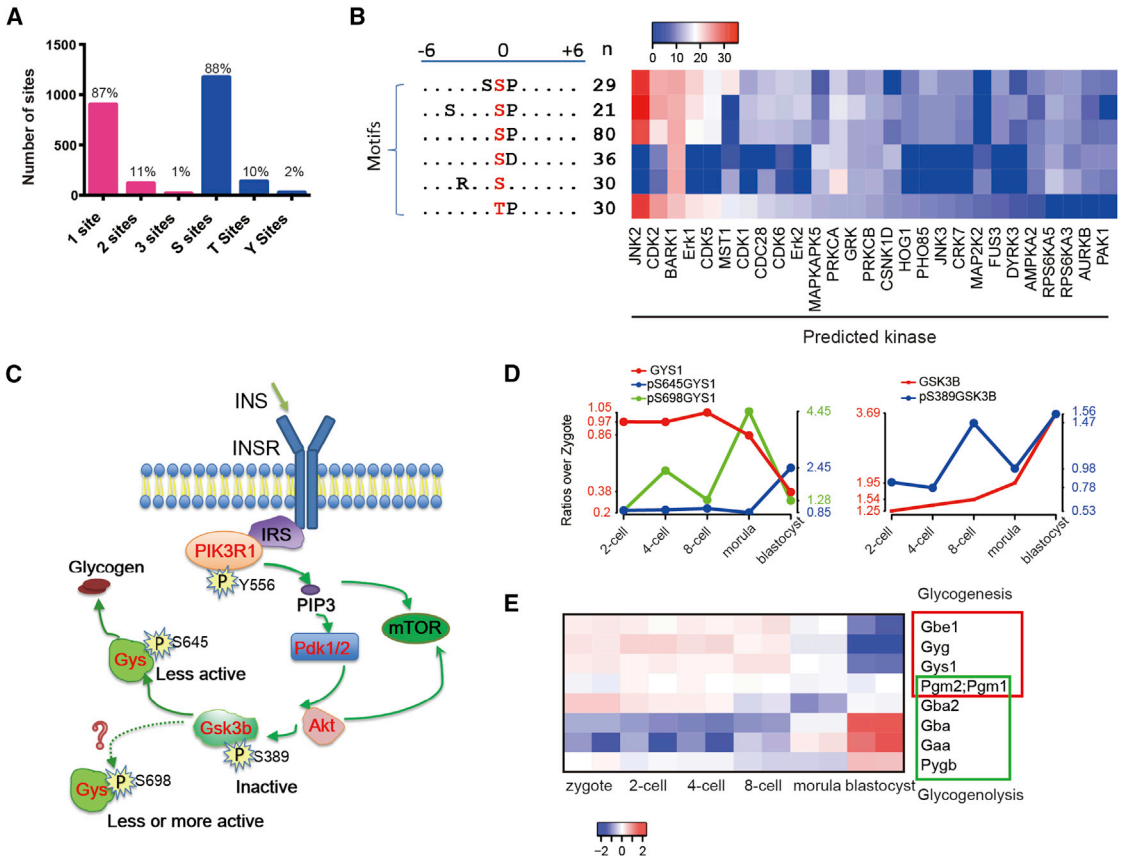


Figure 3. Temporal Regulation of Protein Phosphorylation in Early Embryos

(A) Confidently localized phosphorylation sites are classified into 1, 2, and 3 sites per peptide and into serine, threonine, and tyrosine sites. The proportion of each class is displayed.

(B) Changed phosphorylation sites are extracted to motifs and predicted kinases are displayed. The number (n) represents the number of phosphorylation sites that share a common motif, and the scores represent the likelihood of the predicted kinases.

(C) A model of regulation between glycogenesis downstream of insulin signaling, with the phosphorylation sites identified in this study labeled.

(D) Expression dynamics of Gys1 and two of its phosphorylation sites are displayed in the left panel, and that of its kinase GSK3b and one GSK3b phosphorylation site are displayed in the right panel. The expression ratios of five later stages are calculated based on the level of zygote stage.

(E) Temporal expression dynamics of key proteins in glycogenesis and glycogenolysis pathways.

See also Tables S2, S3, and S4.

identified that show a diverse trend of protein expression (Figures 4A and S3A), and we constructed a protein-protein interaction network for each module with a number of core factors that have the most connectivity (Figures 4B, 4C, and S3C–S3H). In addition, GO enrichment analysis of the protein members in each module revealed distinct biological functions (Figures S3B–S3H). These results not only make it possible to understand the dynamic events during early embryo development, but can also help identify key factors. Intriguingly, proteins in the pink module were highly expressed in the zygote, but rapidly degraded after 2-cell stage. Proteins in this module are enriched in redox proteins and DNA-binding proteins, consistent with DNA demethylation and reprogramming during fertilization (Figures 4B, S3A, and S3B). Some important proteins essential for germ-cell development and early development, including Gdf9, Pmm2, Chmp4b, and Larp7, were defined as core factors in this module (Figure 4B). In contrast, proteins in the blue module maintained their expression levels until morula, and sharply increase in blastocyst (Figure 4C and Figure S3A). Proteins in this module were enriched in pathways related to RNA splicing, ribosome function, and ribonucleoprotein complex formation, all related to RNA metabolism (Figure S3B). Other pathways in the blue module include oxidative phosphorylation, citrate cycle, and cell cycle, which are necessary to produce energy in support of embryonic growth and cell division.

We further hypothesize that proteins highly expressed in the blastocyst and with high connectivity in the blue module might be indispensable during early embryo development. To test this, we selected several core factors, Rrp9, Cript and Zcchc8 (Figure 4C), in the center of blue module, all of which are poorly studied in early embryogenesis. We first tested the mRNA expression of these genes in different tissues and found that they were all highly expressed in the testes and placenta (Figures

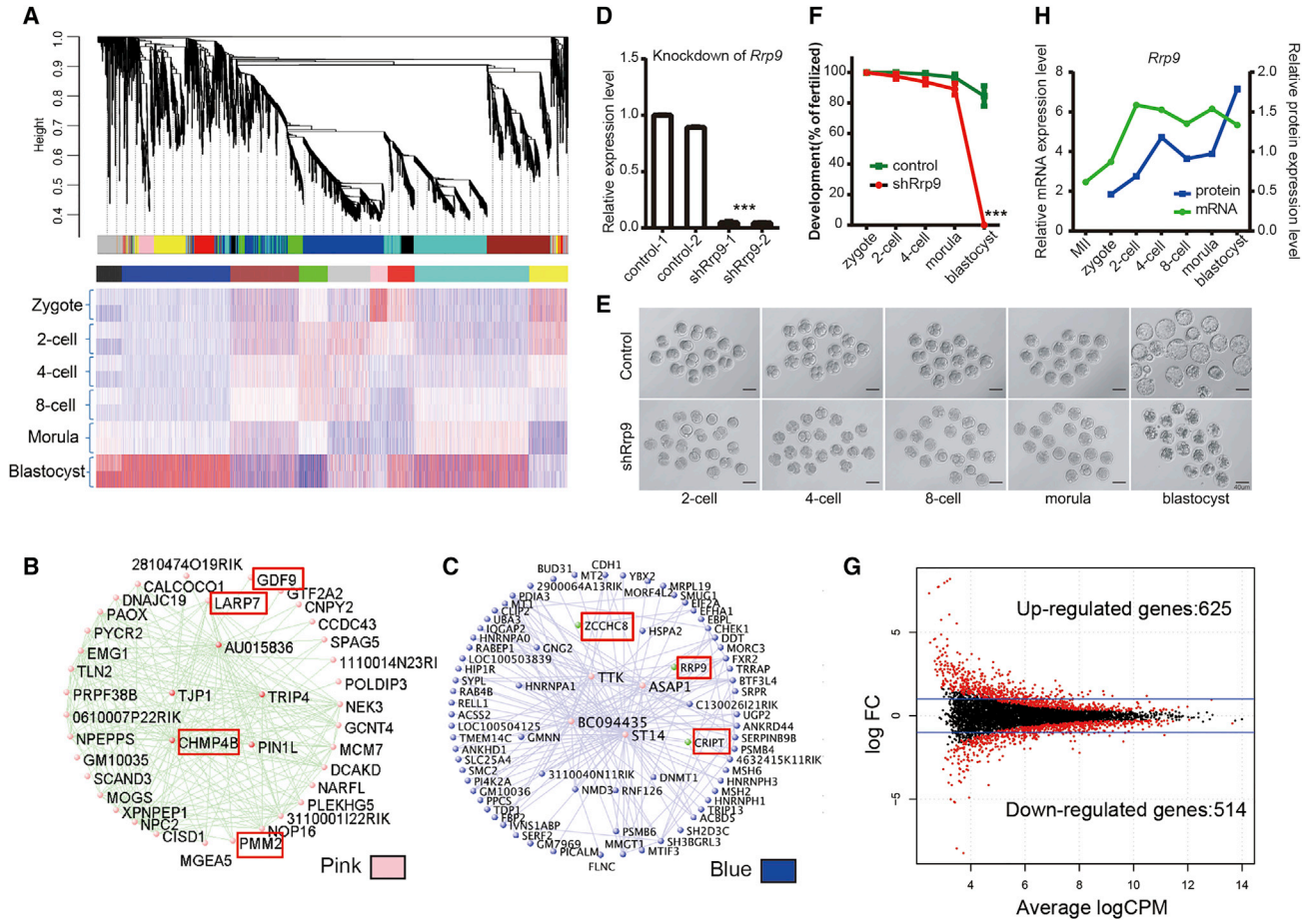


Figure 4. Network Analysis of Protein Expression in the Early Embryo Identifies Critical Factors Dictating Embryonic Development

(A) Hierarchical clustering dendrogram shows co-expression modules of proteins that are color coded. The co-expressed proteins are rearranged, and their expression levels are displayed in the heat map. The intensity of the color represents log2 transformed relative expression levels of each protein.

(B and C) Network connections of proteins in pink (B) and blue (C) modules are displayed. Proteins in the pink module are enriched in mitochondrial components, whereas proteins in the blue module are enriched in maternal proteins and imprinted gene products. Each protein discussed in the text is highlighted with a red box.

(D) Efficiency of siRNA-mediated knockdown of Rrp9. Error bars show SD.

(E) Knockdown of Rrp9 results in developmental arrest at blastocyst stage. Scale bars represent 40 μm.

(F) Percent of successfully developed embryos at each stage after knocking down Rrp9. The developmental rate is calculated related to the number of zygotes. Our data show the results of three repeated experiments. Error bars show SD.

(G) Dot plot shows differentially expressed genes in morula stage between wild-type (WT) and knockdown of *Rrp9* conducted by edgeR (version 3.10.2). The red plots represent the genes that were called differentially expressed genes with a false discovery rate (FDR) < 0.05 and a mean fold change > 1.5. The blue lines show the fold change = 2 and −2.

(H) Correlation of temporal expression between Rrp9 mRNA and protein.

S4A–S4C). We injected small-interfering RNAs (siRNAs) into MII oocytes followed by intracytoplasmic sperm injection (ICSI) to knock down these genes in embryogenesis (Figures 4D and S4D), and further monitored the development efficiency of injected embryos. Compared with the control, the development efficiency of gene-knockdown (KD) embryos decreased specifically in the formation of the blastocyst (Figures 4E, 4F, and S4E). Strikingly, the KD of Rrp9, a component of the nucleolar small nuclear ribonucleoprotein particle, completely blocked the morula-to-blastocyst transition (Figures 4E and 4F). We measured the global gene expression changes after *Rrp9* KD embryos in the morula stage by RNA sequencing (RNA-seq) and identified 625 and 514 genes up- and downregulated after Rrp9 KD (Figure 4G). GO analysis revealed that upregulated genes were enriched in response to stress, regulation of hydrolase activity, and apoptosis. Downregulated genes were enriched in protein transport, cell division, and non-coding RNA (ncRNA) processing (Figure S4F). Apoptosis-related genes were upregulated in Rrp9 knock-down embryos. Additional immunostaining results confirmed that *Rrp9* KD embryos showed notable apoptosis after the morula stage (Figure S4G). These data suggested that the developmental arrest of Rrp9 KD embryos might correlate with the transcriptome defects and cell apoptosis after the morula stage. We further compared the

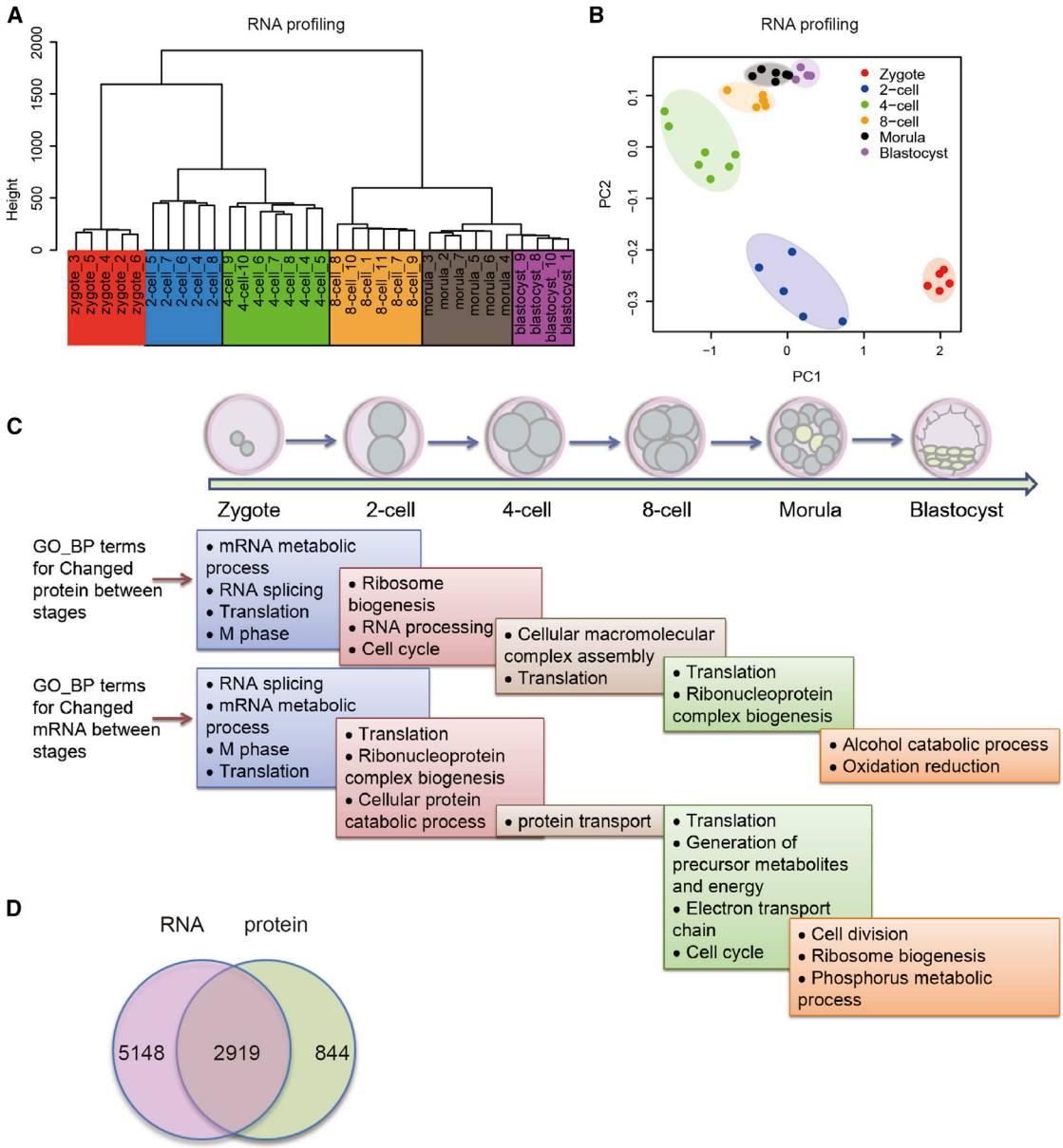


Figure 5. Comparative Analyses of Transcriptomic and Proteomic Data

(A) Unsupervised clustering of mRNA expression in six embryonic stages analyzed by RNA-seq of a single embryo.

(B) Principal component analysis of mRNA expression. The same color represents the same embryonic stage in replicates.

(C) GO terms of between-stage, significantly changed mRNA and proteins.

(D) Overlap between mRNA and protein measurements.

See also Table S5.

mRNA and protein expression dynamics of these three genes and found that Rrp9 and Cript were both activated after fertilization, but Zcchc8 presented at a high level in the oocyte and blastocyst stages (Figures 4H and S4H). We also found that the dynamics of the corresponding RNA and protein of these three genes were different in some stages, suggesting that different regulatory mechanisms may exist in transcription and translation.

Integrated Analysis of Transcriptome and Proteome Data Reveals Coordinated Control of Transcription and Translation and Identifies Previously Undefined Exon-Junction-Derived Peptides

To delineate the correlations between RNA and protein expression at a systematic level, we performed single-embryo RNA-seq experiments from each of the six developmental stages. Hierarchical clustering and PCA (Figures 5A and 5B)

of the RNA-seq data showed a clear difference from that of the proteomic data (Figures 1A and 1B), nevertheless, our RNA-seq results are comparable with published single-cell RNA-seq results (Xue et al., 2013). We grouped the differentially expressed genes (comparing mRNA expression side by side with protein expression) between each of two consecutive stages into GO terms and found that the mRNA metabolic process, mRNA processing, and RNA splicing are among the most enriched terms in early stages (Figure 5C). Consistent with the WGCNA results, RNA metabolism likely represents the most active biological process in early embryos. Overall, 2,919 genes overlapped between the detected transcriptome (8,067 RNAs RPKM > 0.1 in at least one sample for each stage) and the proteome (3,763 quantified proteins) in the whole process (Figure 5D). Among these genes, there were poorer correlations comparing mRNA and protein expression between each of two consecutive stages, but better correlation in protein expression per se between stages (Figure 6A). Our data thus suggest that changes in protein expression appear to be more gradual and the expression level of mRNA tend to have fluctuations at certain stages.

We then defined the up- and downregulated mRNA and protein between each of two consecutive stages. Transition from the zygote to the 2-cell embryo results in thousands of genes downregulated at the mRNA level (Figure 6B), however, massive downregulation at the protein level was not observed (Figure 6C). Moreover, in the overlapped genes, more than 600 declined at the mRNA level, but less than 3% declined at the protein level (Figure 6D). These data suggest that maternal mRNA degradation from the zygote to the 2-cell stage, the first step in the maternal-to-zygote transition, does not necessarily lead to changes in proteins. In later developmental stages, a drastic increase of mRNA occurred during the 8-cell to morula transition, which indicates a dramatic onset of a new wave of transcription (Figure 6B). However, these changes were not observed at the protein level until the next stage (Figures 6C and 6E), suggesting a regulation gap between mRNA and protein expression in the ZGA process.

Because epigenetic modifiers and transcription factors play essential roles in initiating embryonic development programs and establishing DNA and histone methylation patterns for genomic imprinting and chromatin assembly, we assembled mRNA and protein expression of 29 epigenetic modifiers and 8 chromatin remodelers in a heat map (Figures S5A–S5C; Table S5). Interestingly, writers and erasers for DNA methylation correlate extremely well between mRNA and protein expression across all six stages, whereas readers, writers, and erasers for histone modifications and chromatin remodelers are less well correlated. The *de novo* DNA methyltransferase DNMT3a is highly expressed in zygotes, but decreased in the next two stages in a step-wise fashion, whereas the maintenance DNA methyltransferase DNMT1 (DNMT1o) maintains a high expression level until the morula stage and is extensively degraded in the blastocyst stage. The DNA dioxidase TET3 has been proven to play important roles in demethylation of the paternal genome (Gu et al., 2011) and is highly expressed in zygotes and rapidly removed afterward. Moreover, single-strand-selective monofunctional uracil DNA glycosylase (SMUG1) (Masaoka et al., 2003) is also highly expressed in zygotes and might play an important role in demethylation of paternal genomic DNA through a different mechanism.

To explore previously undefined gene products, we took advantage of RNA-seq data and adopted a proteogenomic approach (Kim et al., 2014; Wilhelm et al., 2014) to extract exon junctions and constructed a customized peptide sequence database to search the tandem MS (MS/MS) spectra. After removing known peptides, 45 and 47 previously undefined peptides from the two MS/MS datasets emerged from the analysis with 20 overlaps (Figure 6F; Table S6). Genomic locations of these peptides include intergenic regions, 3′- and 5′-UTRs, introns, as well as other open-reading frames (ORFs). Pairwise, between-stage expression of these 20 peptides displayed a variety of temporal patterns and is presented in a heat map (Figure S6A). For a junction derived from the gene *Paip1* (Figure S6B), we identified with high confidence a peptide from several of the unmapped MS/MS spectra together with quantitative information shown by the reporter ion intensities (Figure 6G). The genomic representation, annotated MS/MS spectrum, as well as the reporter ion information of another junction are displayed in Figures S6C–S6E.

DISCUSSION

Development is a complex process mediated by gene transcription, translation, and metabolism. Technical advances of next-generation sequencing provide detailed information on gene transcription, but how gene transcription correlates with and is regulated by translation and post-translational modifications during development is still poorly understood. In this work, we described the expression changes of more than 4,000 proteins during six stages of mouse pre-implantation embryo development. These provide a comprehensive landscape of proteome and phosphoproteome dynamics of mouse embryo during pre-implantation development and give direct insight into the molecular details on the regulation of the developmental process.

Pre-implantation development is a process that is precisely controlled by a series of genes and transcription factors (Edwards, 2003). We identified a number of core factors by WGCNA analysis, and most of them are previously uncharacterized proteins in embryos. Further, we demonstrated the fundamental importance of three core factors using loss-of-function studies. This raises the exciting potential to identify more critical factors from our proteomic data and deepen our understanding of developmental regulation.

Comparative analysis between transcriptomic and proteomic data disclosed abundant post-transcriptional RNA regulation. Consistent with the findings in *Xenopus* (Peshkin et al., 2015) and zebrafish (Alli Shaik et al., 2014), the correlation between RNA expression and protein expression in the same stages is poor. This further highlights the importance and the value of proteomic studies.

Overall, our data not only provide a valuable resource for further mechanistic studies of individual proteins in pre-implantation embryos, but also suggest additional players and regulatory mechanisms governing early embryo development.

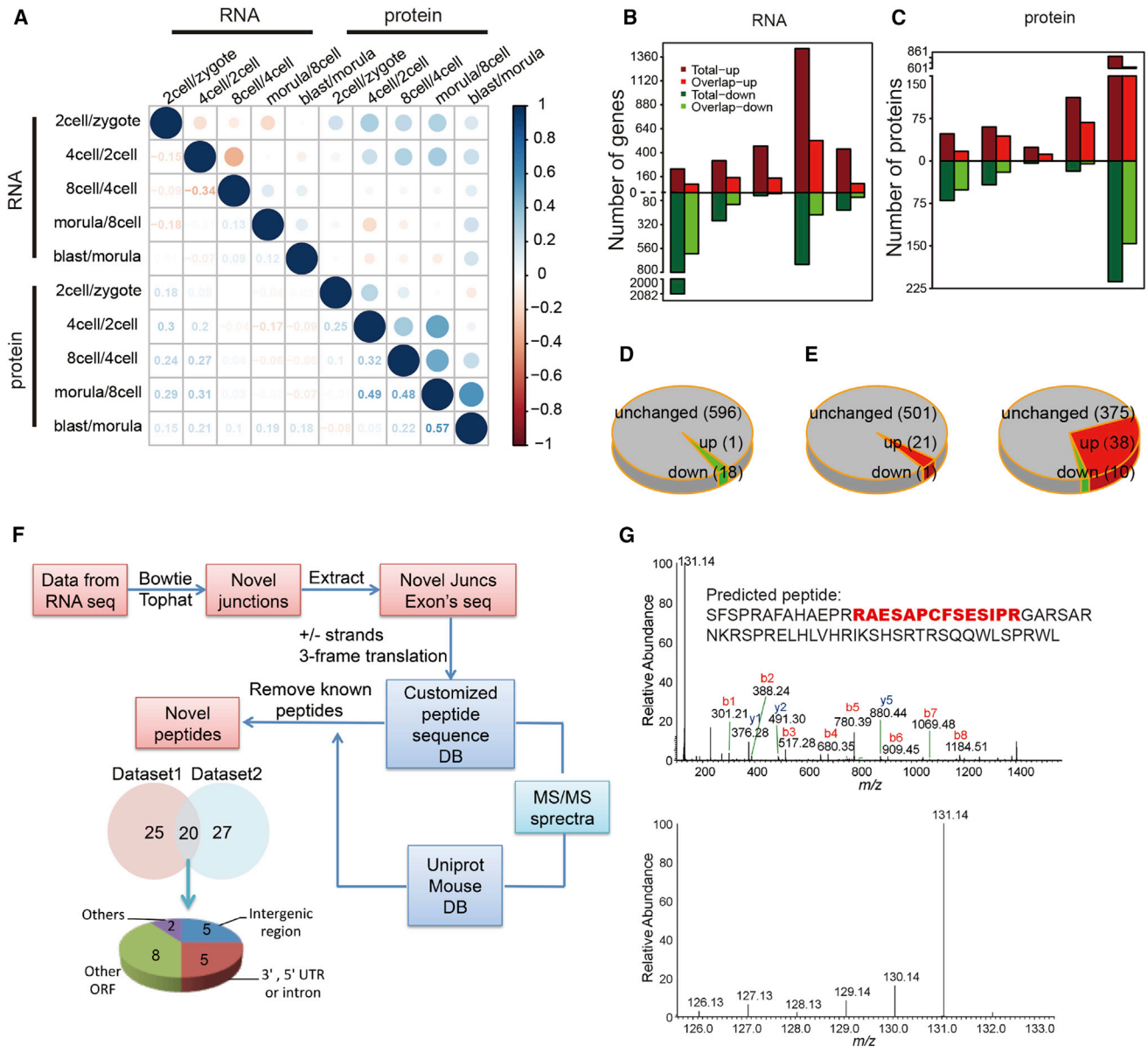


Figure 6. Proteogenomic Analysis Correlates RNA and Protein Expression and Identifies Novel Exon Junction Products

(A) Correlation of RNA and protein expression between two consecutive developmental stages. Pearson's correlation coefficient is coded in the form of the size and color. Abbreviation: blast-blastocyst.

(B and C) Significantly changed mRNAs (B) and proteins (C) between stages are calculated and displayed in the bar graphs. The between-stage order of the comparison is the same as in (A). Red: upregulation; green: downregulation. Dark red and dark green represent all mRNAs and proteins, whereas light red and light green represent overlapped mRNAs and proteins.

(D) Among the large number of downregulated mRNAs between the 2-cell and zygote stage, only a small portion (2.93%) of proteins are downregulated.

(E) Among the large number of upregulated mRNAs between the morula and 8-cell stage, only 0.19% of proteins are upregulated (left); however, between the blastocyst and morula stage, the upregulated proteins increased to 13.8%.

(F) Flow chart of the proteogenomic approach to identifying novel exon-exon junctions from RNA-seq data and to annotating unmapped MS/MS spectra.

(G) Based on the predicted peptide sequence derived from the novel junction, an unmatched MS/MS spectrum is annotated as a tryptic peptide from this sequence (upper panel), and the reporter ion intensity representing relative peptide expression is displayed in the lower panel.

See also Table S6.

EXPERIMENTAL PROCEDURES

Contact for Reagent and Resource Sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Shaorong Gao (gaoshaorong@tongji.edu.cn). Further details and an outline of resources used in this work can be found in the Supplemental Experimental Procedures.

Animals and Collection of Mouse Embryos

The specific-pathogen-free (SPF) grade B6D2F1 mice were bred by crossing female C57BL/6 and male DBA/2 mice using standard procedures and housed in the animal facility of Tongji University (Shanghai, China).

B6D2F1 female mice (8–10 weeks old) were super-ovulated by injection with 7 IU each of pregnant mare serum gonadotropin (PMSG), followed by injection of 5 IU of human chorionic gonadotropin (hCG) (Sansheng Pharmaceutical Industrial,) 48 hr later. The super-ovulated female mice were mated with male mice. Then, the zygote- (24 hr post-hCG injection [phCG]), 2-cell- (48 hr phCG), 4-cell- (58 hr phCG), 8-cell- (68 hr phCG), and morula-stage (78 hr phCG) embryos were collected from the oviducts of female B6D2F1 mice. Blastocyst-stage embryos were collected from the uterus of the females 92 hr after hCG injection.

Quantitative MS

For the TMT-based quantitative MS assay, around 8,000 embryos (zygote-, 2-cell-, 4-cell-, 8-cell-, morula-, and blastocyst-stage embryos) for each sample were harvested and lysed with radioimmunoprecipitation assay (RIPA) buffer. The MS assay was performed using the TMTsixplex Isobaric Label Reagent Set (Thermo Fisher Scientific, catalog number 90061). The detailed protocols and reagents are described in the Supplemental Experimental Procedures.

MS Data Analysis

The MS/MS spectra were searched with the ProLucid algorithm (Xu et al., 2015) against the International Protein Index (IPI) mouse protein database. The search parameters were set as follows: 50-ppm mass tolerance for precursor ions; 0.02-Da mass tolerance for product ions; three missed cleavage sites were allowed for trypsin digestion; enzyme was chosen as semi-tryptic, which allows non-specific cleavage at the peptide N terminus; and three variable modifications were included: protein N-terminal acetylation, methionine oxidation, and cysteine carbamidomethylation. Two static modifications were included: protein N-terminal TMT labeling and lysine TMT labeling. The search results were further processed by the DTASelect program (Tabb et al., 2002) to control the false discovery rate to be under 1%. The assembled database matching file was used to obtain the TMT reporter ion intensities for each peptide using Census software (Park et al., 2008). The report ion mass tolerance was set to 10 ppm.

To identify protein post-translational modifications, we used the SEQUEST algorithm in the Proteome Discoverer software suite (Thermo Fisher Scientific). The search parameters included a differential modification on serine, threonine, and tyrosine residues of 79.9663 amu, indicating the addition of phosphorous group(s). For modification on lysine residues, differential modification of 114.0429 (ubiquitination), 42.0367 (acetylation), 14.0156 (mono-methylation), 28.0313 (di-methylation), and 42.0464 (tri-methylation) were applied. The database search results were assembled and filtered. Proteome Discoverer further calculated the quantitative information of the TMT-tagged reporter ions at the modified peptide level. For phosphorylation, the quantitative information and the localized phosphorylation sites were assembled to derive quantified phosphorylation sites. The Motif-X algorithm (Schwartz and Gygi, 2005) was then applied to extract the amino acid sequence patterns flanking the phosphorylation sites from site-localized phosphopeptides. To be considered as a motif, we allowed the occurrence of the same pattern to be more than 20 times, and the significance value was less than 0.000001. The background reference database was the IPI mouse protein database. The Group-based Prediction System (GPS, version 2.0) algorithm (Xue et al., 2008) was then used to predict the kinases responsible for the phosphorylation sites matching the motifs.

High-Throughput RNA-Seq

Single-cell RNA-seq was performed following previously published studies (Tang et al., 2010). In brief, the harvested single blastomeres were washed several times in 0.5% BSA-PBS (Sigma) solution and subsequently picked and transferred into lysate buffer by a mouth pipette. Diluted External RNA Controls Consortium (ERCC) mix (Life Technologies, catalog number 4456740) was added to the lysis buffer as a spike-in for each sample. Reverse transcription was performed directly on the cytoplasmic lysate. Terminal deoxynucleotidyl transferase was then used to add a poly(A) tail to the 3′ end of the first-strand cDNAs. The total cDNA library of the single cell was then amplified by 18–20 cycles for library construction. The amplified cDNA was fragmented using a Covaris sonicator (Covaris S220). To generate the sequence libraries, the TruSeq Library Prep Pooling kit (Illumina, number 15042173), the NEBNext Ultra DNA Library Prep Kit (New England BioLabs, number E7370), or the KAPA Hyper Prep Kit was used following the manufacturer's instructions. All adapters with index barcode used in this protect were diluted (or not) from the adapters offered by the TruSeq Library Prep Pooling kit. Paired-end 100-bp or 125-bp sequencing was performed on a HiSeq 2500 or 2000 sequencer (Illumina) at the National Institute of Biological Sciences (NIBS), Peking University and Berry Genomics Corporation.

Knocking Down of New Players in Early Embryos

siRNAs against mouse Rrp9, Zcchc8, Cript, and negative control were diluted in nuclease-free water to 20 μM. The siRNAs for each gene were mixed and diluted to a final concentration of 5 μM. Oocytes were injected with approximately 10 pL of 5-μM siRNAs using a Piezo-driven micromanipulator. The injected oocytes were then incubated for at least 3 hr for ICSI. To inject fresh spermatozoa, approximately 1 μL of the incubated sperm suspension was mixed with a drop of HEPES-buffered Chatot-Ziomek-Bavister (HCZB) containing 10% (w/v) polyvinylpyrrolidone (PVP; Irvine Scientific, Santa Ana, CA, USA). The sperm head was separated from the tail by the application of several Piezo pulses, and the head was then injected into the oocyte according to the method described by Kimura and Yanagimachi (1995).

siRNA sequences were designed based on published paper or using the GE Dharmacon siDesign Center Algorithm (http://dharmacon.gelifesciences.com/design-center/) and synthesized by the Biological Resource Center at NIBS. All siRNA-targeting sequences are shown in Supplemental Experimental Procedures.

HCZB medium was used for gamete handling and ICSI. Chatot-Ziomek-Bavister (CZB) medium was used for embryo culture in an atmosphere of 5% CO_2. For embryo incubation, CZB was overlaid with sterile mineral oil (Sigma).

Measuring mRNA Levels Using Real-Time PCR

To analyze the KD efficiency of siRNA, total RNA of 20 4-cell-stage embryos were purified using the RNeasy Mini Kit (QIAGEN, catalog number 74104) according to the manufacturer's instructions. The cDNA was synthesized by a reverse transcription system (Promega). Quantitative real-time PCR was performed using a SYBR Premix Ex Taq (Takara), and signals were detected with an ABI7500 Real-Time PCR System (Applied BioSystems). H2A.Z was used as the internal control. Primer sequences are shown in Supplemental Experimental Procedures.

Fuzzy C-means Clustering

Proteins from six embryonic stages were grouped into different clusters using Mfuzz package in R with fuzzy c-means algorithm (Kumar and E. Futschik, 2007).

WGCNA

A WGCNA was conducted for 12 mouse embryonic samples using the WGCNA package (Langfelder and Horvath, 2008) in R. Based on the topological overlap matrix measured from a pairwise correlation-based adjacency matrix, the neighborhood similarity among proteins were estimated and the protein co-expression modules, which are distinguished with different colors, were then identified by average linkage hierarchical clustering. Using the Dynamic Hybrid Tree Cut algorithm and a minimum module size of 75 proteins,

a total of eight modules were identified. The network of each module was visualized with VisANT version 4.0 (Hu et al., 2013).

Processing of RNA-Seq Data

A total of 85 Gb of RNA-seq raw data for 33 single-cell samples were generated and trimmed with Trimmomatic version 0.32 software to obtain clean reads, which were mapped to the UCSC hg38mm10 reference database. TopHat version 2.0.11 (Kim et al., 2013) alignment was used to identify novel junctions, and Cufflinks version 2.1.1 (Trapnell et al., 2010) was used to create gene expression profile with the reads per kilobase per million mapped reads (RPKM) method. All of the genes with an RPKM $\geq$0.1 are considered as expressed genes in the following analysis. Genes with a p value $\leq$0.01 and a fold change (FC) $\geq$2 in the transcript level between stages were considered to be differentially expressed. GO enrichment analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Huang et al., 2009), and p values for the representative GO terms shown in the present study were adjusted with the Benjamini-Hochberg procedure. Heat maps of the protein and gene expression profiles were generated using R packages.

Correlation between mRNA and Protein Expression

To correlate mRNA and protein expression between two consecutive stages, we compared the 3,763 genes identified in both proteomic datasets with the RNA-seq results and found 2,919 genes in common. Correlation analysis of mRNA and protein expression between two consecutive stages was carried out in these 2,919 genes using Pearson's correlation (Figure 4A).

Proteogenomic Analysis

Bowtie 2 version 2.2.1 (Langmead and Salzberg, 2012) and TopHat version 2.0.11 software packages were used to detect previously undefined junctions. The following steps were implemented as Python scripts to identify confident novel peptides. The novel junction sequences were extracted from UCSC hg38mm10 reference database, were then translated into peptide sequences with three-frame translation strategy on both sense strand and antisense strand to create a customized peptide sequence database. By matching our MS/MS spectra with the customized database, we extracted the matched peptide sequences, combined with the customized peptide sequence database made from the novel junctions, and performed another database search with unmatched spectra. Additionally, we determined the annotated spectra from our MS/MS spectra using the UniProt mouse database. By discarding the candidates of the confidant peptides included in the annotated spectra, we identified 20 high-confident novel peptides that mapped to the peptide sequences predicted by the novel junctions. The integrative genomics viewer (igv) (Robinson et al., 2011) was used as visualization tool for the novel junctions from RNA-seq.

Statistical Analysis

Data are presented as mean values ± SEM. Comparisons between the two groups was determined by two-tailed Student's *t* test unless otherwise stated. Statistical analyses were performed with GraphPad Prism software or R as described in Experimental Procedures for individual analysis, and statistical significance is displayed as not significant (NS), $p > 0.05$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

DATA AND SOFTWARE AVAILABILITY

The accession number for the mass spectrometry data reported in this paper is ProteomeXchange Consortium via the PRIDE partner repository: PXD003315. The accession numbers for the chromatin immunoprecipitation sequencing and RNA-seq data reported in this paper are GEO: GSE70605 and GSE106612.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and six tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2017.11.111.

ACKNOWLEDGMENTS

This study was supported by the Ministry of Science and Technology of China (2016YFA0100400), the National Natural Science Foundation of China (31721003, 31430056, 91319306, 31401266, 31322031, and 31371288), the Science and Technology Commission of Shanghai Municipality (15XD1503500 and 14CG16), and a Shanghai Pujiang Talent Project (14PJ1402900) and East China Normal University (ECNU) National "985" Project grant.

AUTHOR CONTRIBUTIONS

Y.G., L.L., and S.G. conceived the project; Y.G., X.L., C.L., Z.K., L.L., W.L., Y.W., X.K., J.L., Y.Z., J.Y., H.W., and S.C. performed all the experiments. Y.G., B.T., and L.L. analyzed the data. Y.G., B.T., L.L., and S.G. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 6, 2017
Revised: November 8, 2017
Accepted: November 29, 2017
Published: December 26, 2017

REFERENCES

Alli Shaik, A., Wee, S., Li, R.H., Li, Z., Carney, T.J., Mathavan, S., and Gunaratne, J. (2014). Functional mapping of the zebrafish early embryo proteome and transcriptome. J. Proteome Res. *13*, 5536–5550.

Beausoleil, S.A., Jedrychowski, M., Schwartz, D., Elias, J.E., Villén, J., Li, J., Cohn, M.A., Cantley, L.C., and Gygi, S.P. (2004). Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc. Natl. Acad. Sci. USA *101*, 12130–12135.

Bebbere, D., Masala, L., Albertini, D.F., and Ledda, S. (2016). The subcortical maternal complex: multiple functions for one biological structure? J. Assist. Reprod. Genet. *33*, 1431–1438.

Brill, L.M., Xiong, W., Lee, K.B., Ficarro, S.B., Crain, A., Xu, Y., Terskikh, A., Snyder, E.Y., and Ding, S. (2009). Phosphoproteomic analysis of human embryonic stem cells. Cell Stem Cell *5*, 204–213.

Dahl, J.A., Jung, I., Aanes, H., Greggains, G.D., Manaf, A., Lerdrup, M., Li, G., Kuan, S., Li, B., Lee, A.Y., et al. (2016). Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. Nature *537*, 548–552.

Deutsch, D.R., Fröhlich, T., Otte, K.A., Beck, A., Habermann, F.A., Wolf, E., and Arnold, G.J. (2014). Stage-specific proteome signatures in early bovine embryo development. J. Proteome Res. *13*, 4363–4376.

Edwards, R.G. (2003). Aspects of the molecular regulation of early mammalian development. Reprod. Biomed. Online *6*, 97–113.

Flyamer, I.M., Gassler, J., Imakaev, M., Brandão, H.B., Ulianov, S.V., Abdennur, N., Razin, S.V., Mirny, L.A., and Tachibana-Konwalski, K. (2017). Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. Nature *544*, 110–114.

Futschik, M.E., and Carlisle, B. (2005). Noise-robust soft clustering of gene expression time-course data. J. Bioinform. Comput. Biol. *3*, 965–988.

Gu, T.P., Guo, F., Yang, H., Wu, H.P., Xu, G.F., Liu, W., Xie, Z.G., Shi, L., He, X., Jin, S.G., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature *477*, 606–610.

Guo, H., Zhu, P., Wu, X., Li, X., Wen, L., and Tang, F. (2013). Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. Genome Res. *23*, 2126–2135.

Hu, Z., Chang, Y.C., Wang, Y., Huang, C.L., Liu, Y., Tian, F., Granger, B., and Delisi, C. (2013). VisANT 4.0: Integrative network platform to connect genes, drugs, diseases and therapies. Nucleic Acids Res. *41*, W225-31.

Huang, da W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

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

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

Kim, M.S., Pinto, S.M., Getnet, D., Nirujogi, R.S., Manda, S.S., Chaerkady, R., Madugundu, A.K., Kelkar, D.S., Isserlin, R., Jain, S., et al. (2014). A draft map of the human proteome. Nature *509*, 575–581.

Kimura, Y., and Yanagimachi, R. (1995). Intracytoplasmic sperm injection in the mouse. Biol Reprod *52*, 709–720.

Krauchunas, A.R., Horner, V.L., and Wolfner, M.F. (2012). Protein phosphorylation changes reveal new candidates in the regulation of egg activation and early embryogenesis in D. melanogaster. Dev. Biol. *370*, 125–134.

Kumar, L., and E. Futschik, M. (2007). Mfuzz: a software package for soft clustering of microarray data. Bioinformation *2*, 5–7.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Leung, C.Y., and Zernicka-Goetz, M. (2015). Mapping the journey from totipotency to lineage specification in the mouse embryo. Curr. Opin. Genet. Dev. *34*, 71–76.

Li, L., Baibakov, B., and Dean, J. (2008). A subcortical maternal complex essential for preimplantation mouse embryogenesis. Dev. Cell *15*, 416–425.

Liu, X., Wang, C., Liu, W., Li, J., Li, C., Kou, X., Chen, J., Zhao, Y., Gao, H., Wang, H., et al. (2016). Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. Nature *537*, 558–562.

Lu, F., Liu, Y., Inoue, A., Suzuki, T., Zhao, K., and Zhang, Y. (2016). Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development. Cell *165*, 1375–1388.

Lucitt, M.B., Price, T.S., Pizarro, A., Wu, W., Yocum, A.K., Seiler, C., Pack, M.A., Blair, I.A., Fitzgerald, G.A., and Grosser, T. (2008). Analysis of the zebrafish proteome during embryonic development. Mol. Cell. Proteomics *7*, 981–994.

Maekawa, M., Yamamoto, T., Tanoue, T., Yuasa, Y., Chisaka, O., and Nishida, E. (2005). Requirement of the MAP kinase signaling pathways for mouse preimplantation development. Development *132*, 1773–1783.

Maekawa, M., Yamamoto, T., Kohno, M., Takeichi, M., and Nishida, E. (2007). Requirement for ERK MAP kinase in mouse preimplantation development. Development *134*, 2751–2759.

Masaoka, A., Matsubara, M., Hasegawa, R., Tanaka, T., Kurisu, S., Terato, H., Ohyama, Y., Karino, N., Matsuda, A., and Ide, H. (2003). Mammalian 5-formyluracil-DNA glycosylase. 2. Role of SMUG1 uracil-DNA glycosylase in repair of 5-formyluracil and other oxidized and deaminated base lesions. Biochemistry *42*, 5003–5012.

Ong, S.E., and Mann, M. (2005). Mass spectrometry-based proteomics turns quantitative. Nat. Chem. Biol. *1*, 252–262.

Ozias, C.B., and Stern, S. (1973). GLycogen levels of preimplantation mouse embryos developing in vitro. Biol. Reprod. *8*, 467–472.

Paranjpe, S.S., and Veenstra, G.J. (2015). Establishing pluripotency in early development. Biochim. Biophys. Acta *1849*, 626–636.

Park, S.K., Venable, J.D., Xu, T., and Yates, J.R., 3rd. (2008). A quantitative analysis software tool for mass spectrometry-based proteomics. Nat. Methods *5*, 319–322.

Peshkin, L., Wühr, M., Pearl, E., Haas, W., Freeman, R.M., Jr., Gerhart, J.C., Klein, A.M., Horb, M., Gygi, S.P., and Kirschner, M.W. (2015). On the relationship of protein and mRNA dynamics in vertebrate embryonic development. Dev. Cell *35*, 383–394.

Riethmacher, D., Brinkmann, V., and Birchmeier, C. (1995). A targeted mutation in the mouse E-cadherin gene results in defective preimplantation development. Proc. Natl. Acad. Sci. USA *92*, 855–859.

Rigbolt, K.T., Prokhorova, T.A., Akimov, V., Henningsen, J., Johansen, P.T., Kratchmarova, I., Kassem, M., Mann, M., Olsen, J.V., and Blagoev, B. (2011). System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. Sci. Signal. *4*, rs3.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2013). Corrigendum: global quantification of mammalian gene expression control. Nature *495*, 126–127.

Schwartz, D., and Gygi, S.P. (2005). An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. Nat. Biotechnol. *23*, 1391–1398.

Snider, N.T., and Omary, M.B. (2014). Post-translational modifications of intermediate filament proteins: mechanisms and functions. Nat. Rev. Mol. Cell Biol. *15*, 163–177.

Tabb, D.L., McDonald, W.H., and Yates, J.R., 3rd. (2002). DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. J. Proteome Res. *1*, 21–26.

Tang, F., Barbacioru, C., Nordman, E., Li, B., Xu, N., Bashkirov, V.I., Lao, K., and Surani, M.A. (2010). RNA-Seq analysis to capture the transcriptome landscape of a single cell. Nat Protoc *5*, 516–535.

Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A.K., and Hamon, C. (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal. Chem. *75*, 1895–1904.

Thomson, J.L., and Brinster, R.L. (1966). Glycogen content of preimplantation mouse embryos. Anat. Rec. *155*, 97–102.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Van Hoof, D., Muñoz, J., Braam, S.R., Pinkse, M.W., Linding, R., Heck, A.J., Mummery, C.L., and Krijgsveld, J. (2009). Phosphorylation dynamics during early differentiation of human embryonic stem cells. Cell Stem Cell *5*, 214–226.

Wang, S., Kou, Z., Jing, Z., Zhang, Y., Guo, X., Dong, M., Wilmut, I., and Gao, S. (2010). Proteome of mouse oocytes at different developmental stages. Proc. Natl. Acad. Sci. USA *107*, 17639–17644.

Wassarman, P.M., and Kinloch, R.A. (1992). Gene expression during oogenesis in mice. Mutat. Res. *296*, 3–15.

Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A.M., Lieberenz, M., Savitski, M.M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., et al. (2014). Mass-spectrometry-based draft of the human proteome. Nature *509*, 582–587.

Wu, J., Huang, B., Chen, H., Yin, Q., Liu, Y., Xiang, Y., Zhang, B., Liu, B., Wang, Q., Xia, W., et al. (2016). The landscape of accessible chromatin in mammalian preimplantation embryos. Nature *534*, 652–657.

Wühr, M., Freeman, R.M., Jr., Presler, M., Horb, M.E., Peshkin, L., Gygi, S., and Kirschner, M.W. (2014). Deep proteomics of the Xenopus laevis egg using an mRNA-derived reference database. Curr. Biol. *24*, 1467–1475.

Xu, T., Park, S.K., Venable, J.D., Wohlschlegel, J.A., Diedrich, J.K., Cociorva, D., Lu, B., Liao, L., Hewel, J., Han, X., et al. (2015). ProLuCID: An improved SEQUEST-like algorithm with enhanced sensitivity and specificity. J. Proteomics *129*, 16–24.

Xue, Y., Ren, J., Gao, X., Jin, C., Wen, L., and Yao, X. (2008). GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. Mol. Cell. Proteomics *7*, 1598–1608.

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

Zhang, B., Zheng, H., Huang, B., Li, W., Xiang, Y., Peng, X., Ming, J., Wu, X., Zhang, Y., Xu, Q., et al. (2016). Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. Nature *537*, 553–557.

Zhou, F., Lu, Y., Ficarro, S.B., Adelmant, G., Jiang, W., Luckey, C.J., and Marto, J.A. (2013). Genome-scale proteome quantification by DEEP SEQ mass spectrometry. Nat. Commun. *4*, 2171.