A critical review of single-cell RNA sequencing-based genetic profiling for human embryos

Motivation

Embryonic development is a highly complex process that requires precise spatial and temporal regulations. Through the preimplantation stages, single-cell zygotes undergo drastic gene expression changes to form multicellular blastocysts. Proper blastocyst formation is crucial to the success of fetus development, as poor blastocyst quality often lead to failure in post-*in vitro* fertilization (IVF) implantation[1]. The process of blastocyst outgrowth to derive human embryonic stem cells (hESCs) also requires unique transition at the transcriptome level[2]. The ability of hESCs to differentiate into different cell lineages enables them to serve as promising tools for organ reconstruction and tissue engineering[3]. Thus, characterizing the genetic signatures of hESCs and cells across different preimplantation stages is not only critical to better understand the embryonic development but also impactful for the current IVF processes and regenerative medicine as a whole.

Genetic profiling was first performed in hESCs as an attempt to identify groups of genes responsible for proliferation and pluripotency[4]. Although these studies provide valuable insights which later led to the revolutionizing discovery of induced pluripotent stem cells[5], cultured hESCs for some time were assumed to be equivalent to their inner cell mass counterpart. In 2009, studies by Reijo Pera *et al*. indicated that hESCs and inner cell mass form two distinct clusters under global gene expression analysis[1]. The group further proposed that hESCs may be at a later developmental time point compared to the inner cell mass cells, or hESCs may simply be a distinct population that has no *in vivo* correspondence. This finding motivated further studies to investigate and better understand the temporal gene expression changes across different preimplantation stages.

Previous global genetic profiling methods include DNA microarray, expressed sequenced tag enumeration, massively parallel signature sequencing and serial analysis of gene expression[6]. These techniques, however, require a large sample size for sufficient mRNA collection, which is difficult for human embryo studies. Moreover, cells are inherently heterogeneous. Pulling together transcripts from individual cells in the embryo may mask the intrinsic heterogeneity relevant to the transcriptome signatures of each stage. These critical bottlenecks call for a profiling method that can detect specific gene expressions at single-cell sensitivity. With the advance in technology, single-cell RNA sequencing (RNA-seq) greatly facilitated a more accurate genetic profiling for human preimplantation development. This review will focus on the techniques and findings by Xue *et al*.[7] and Yan *et al*.[8].

Human Preimplantation Development and Previous Genetic Profiling Findings

Preimplantaion development is the process by which an oocyte becomes fertilized and develops into a blastocyst before attaching onto the uterine wall. When sperm first enters a mature oocyte, the nuclei from the sperm and the egg remain separated, a stage known as the pronucleus stage. Subsequently, the nuclear membrane dissolves to allow mixing of both paternal and maternal chromosomes, giving rise to a diploid zygote[9]. The zygote undergoes three cleavages to make an 8-cell embryo.

During the cleavage process, the embryo volume remains relatively constant, suggesting a low demand for cell growth and a high demand for DNA synthesis[10]. Followed by cleavages, the embryo undergoes compaction to form the morula. Individual cells in the 2-cell stage to morula are referred to as blastomeres. Compact morula undergoes first cell-fate decision to give rise to the trophectoderm and the inner cell mass in blastocyst. Cells from the inner cell mass are considered pluripotent, as they have the ability to differentiate into any cell type in a functional organism. In contrast, trophectodermal cells are multipotent because it only forms the placental tissue for fetal support. Human embryonic stem cells are derived from the inner cell mass and can be cultured *in vitro* without losing their pluripotency[11].

Previous genomic analyses revealed that the transcriptome changes during human preimplantation development are highly dynamic. At the onset of first cleavage, maternal mRNAs degrade monotonically. In contrast, the zygotic genes are activated in a wave-like manner, and the activation is most prominent between the 4-cell to 8-cell stage transition[12]. For instance, multiple transcription factors and translation initiators such as *GTF2A1, PTMA, RUNX2, YY1, EIF1AX* and *EIF4A3* are highly expressed. Genes associated with morula compaction and blastocyst formation are subsequently upregulated at the corresponding stages. Some examples are pluripotency genes *NANOG* and *SOX2*, proliferation genes *SERTAD1, H2AFZ* and *JARID1B* and differentiation gene *EOMES*. **Figure 1** summarizes the preimplantation process and the corresponding transcriptome dynamics.

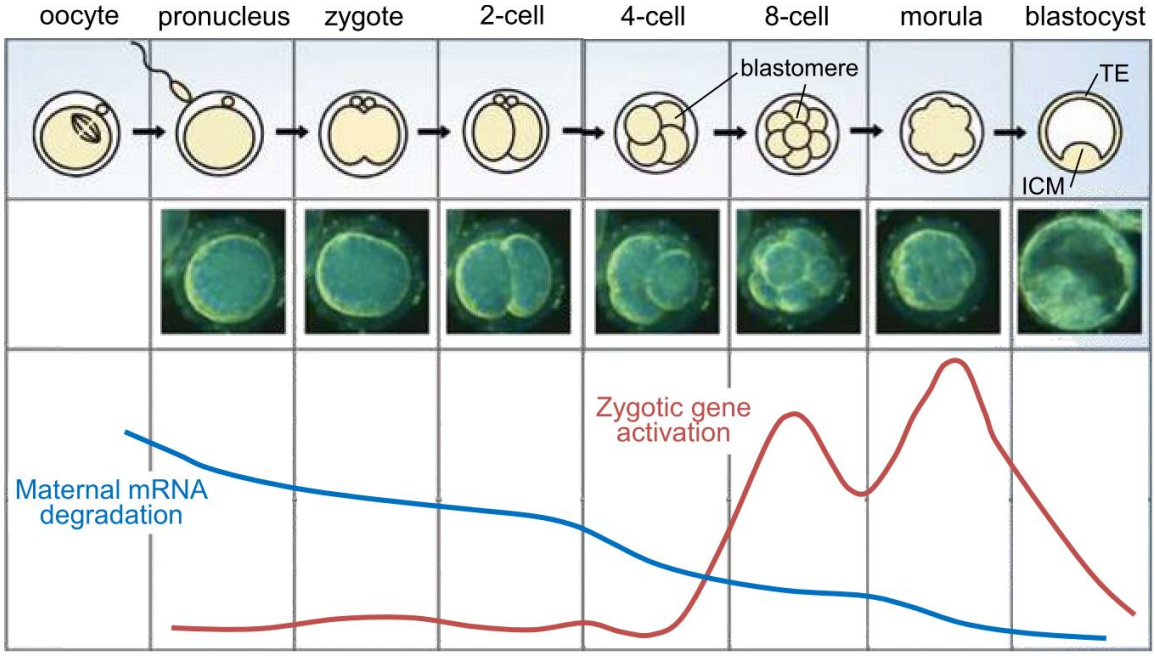

Figure 1. Human preimplantation development. Individual cells in the developing embryo are termed blastomeres. The blastocyst consists of two cell types: the pluripotent inner cell mass (ICM) which develops into the fetus and the trophectoderm (TE) which gives rise to the placenta. The lower panel depicts the major gene expression changes across different stages. Maternal mRNA gradually degrade from the onset of embryo development, while zygotic genes experience waves of activation at the 4-cell to 8-cell transition followed by morula compaction and blastocyst cavitation. Figure modified from ref. 12.

Single-cell RNA Sequencing Technique

Standard microarrays require micrograms of input mRNA for transcriptome analysis. However, single cell contains mRNA on the order of 0.1 picogram, which is 10^7-fold lower than microarray analysis threshold. Due to this low abundance in mRNA, amplification is crucial for single-cell RNA sequencing[13]. First, complimentary DNAs (cDNAs) are made from whole-cell lysate via reverse transcription. To distinguish mRNAs transcripts from other functional RNAs, polyadenylated RNA strands are reverse transcribed with bias using oligo(dT) primers. The resulting cDNAs are further capped with poly(A) tail at the 3' end, followed by multiple rounds of amplification. Both polymerase chain reaction (PCR)-based and *in vitro* transcription (IVT)-based amplification are commonly used. PCR amplifies cDNAs exponentially to shorten the waiting time. However, primer dimers and nonspecific byproducts can accumulate, reducing the amplification fidelity. IVT, on the other hand, is more stringent and produces fewer byproducts. The drawback is that IVT is time consuming as it amplifies cDNA linearly, and it is limited to generate cDNAs less than 1kb in length. Once amplified, the cDNA library can be used for microarray analysis to quantify known gene expressions or profiled using any next-generation sequencing techniques[14]. Both Xue *et al*. and Yan *et al*. constructed their cDNA libraries using PCR amplification and obtained transcriptome data by Illumina-based sequencing.

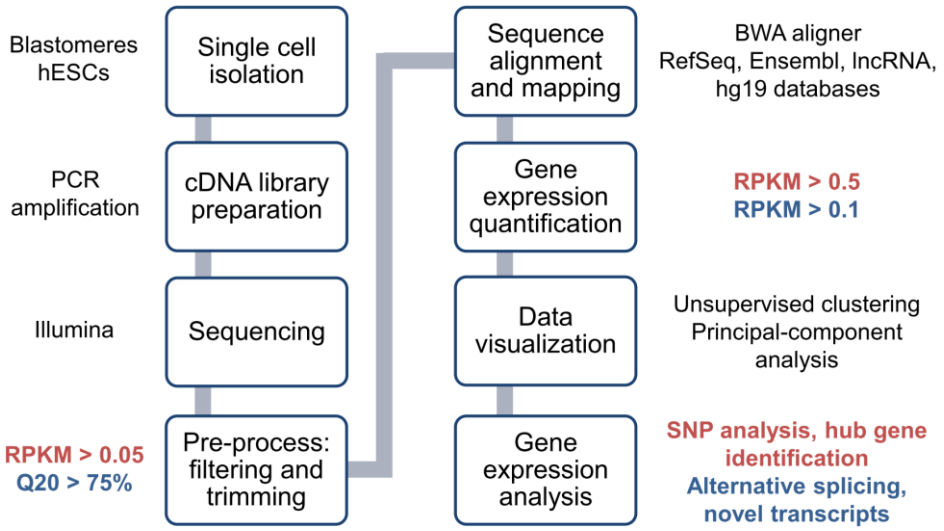

Figure 2. Flow chart of the single-cell RNA sequencing process. Red and blue captions indicate method choices in Xue *et al*. and Yan *et al*.'s studies, respectively. Reads per kilobase per million reads (RPKM) reflects mRNA abundance and is often used to gate analysis threshold. Q20 corresponds to the quality control factor assigned by Yan *et al*.[8] Burrows-Wheeler Aligner (BWA) tool was used by both groups for sequence alignment and mapping to the human reference genome. Both groups clustered and visualized their data using unsupervised hierarchical approach and principal-component analysis. Xue *et al*. then focuses on single-nucleotide polymorphism studies and hub gene identification, while Yan *et al*. further analyzed the transcriptome to identify novel transcripts.

Single-cell RNA Sequencing Data Analysis

Subsequent to cDNA library construction, amplification and sequencing, raw data need to be trimmed, aligned and mapped against reference genome, and quantified

for gene expression. These processes are often facilitated by packages such as FASTX and Trimmomatic for trimming, BWA, SOAP and NovoAlign for alignment, and BWA, Cufflinks, casper and EdgeR for quantification[15].

A. Data Visualization: Gene Clustering

Hundreds to thousands of genes can be expressed in a single cell at a given time. To gain insights from genomic profiling, individual genes are often clustered into groups to reveal expression patterns. Most clustering algorithms utilize one of the 7 models: connectivity, centroid, distribution, density, subspace, group and graph-based[16]. This review will focus on one prominent algorithm, the connectivity-based hierarchical clustering, for embryonic gene expression analysis. Prior to applying clustering algorithms, gene expression data are first transformed such that each gene is represented by a vector and each vector component corresponds to the –log of its expression in a particular experiment[17].

	oocyte	pronucleus	zygote	2-cell	4-cell	8-cell	morula	blastocyst
Gene A	[0.2	0.5	0.3	0.7	0.3	0.15	0.9	0.8]
Gene A'	[0.70	0.30	0.52	0.15	0.52	0.82	0.05	0.10]

The diagram above depicts a hypothetical gene expression profile (A) at various stages through preimplantation development and its transformed vector (A'). For a data set with *n* number of genes and *m* number of cells, the input matrix size will be *n* x *m* for clustering analysis.

Hierarchical Clustering

Hierarchical clustering is a proximity-based agglomerative method. The distances between each pair of genes are first calculated; then the genes with shortest distance are grouped to form clusters. Gene clusters are subsequently treated as a single unit for distance calculation, and the grouping process iterates until all genes are agglomerated under a single cluster. Two ways to determine the pairwise distance are the Minkowski measurements (Euclidean, Manhattan, Chebyshev) which detect the spatial proximity between two genes, and correlation distances (Pearson, Spearman, Kendall) which measure the similarity in gene expression without regard to the absolute scale. Specifically, Euclidean distance and Pearson correlation are commonly calculated, and their formulae are as follows. A gene pair is represented by indices *i* and *j*, and the corresponding transformed gene expression is denoted by *G.*

$$\text{Euclidean: } D_{ij} = \sum_{k=1}^{m} (G_{ik} - G_{jk})^2$$

$$\text{Pearson: } p_{ij} = \frac{\sum_{k=1}^{m} (G_{ik} - \bar{G}_i)(G_{jk} - \bar{G}_j)}{\sqrt{\sum_{k=1}^{m} (G_{ik} - \bar{G}_i)^2} \sqrt{\sum_{k=1}^{m} (G_{jk} - \bar{G}_j)^2}}$$

Methods to compare distances between clusters include single linkage (nearest neighbor joining), complete linkage (furthest neighbor joining), average linkage (average distance between all genes in two cluster pairs) and centroid linkage

(distance between two cluster centroids). Each calculation methods have associated pros and cons. While single and complete linkage methods are computationally simpler to implement, average and centroid linkage methods are less sensitive to noise and outliers[16]. Both Xue *et al.* and Yan *et al*. used hierarchical clustering with Euclidean distance calculation for their embryonic gene expression analysis. Yan *et al.*, however, used complete linkage clustering, while Xue *et al.* used average linkage clustering (**Figure 3**).

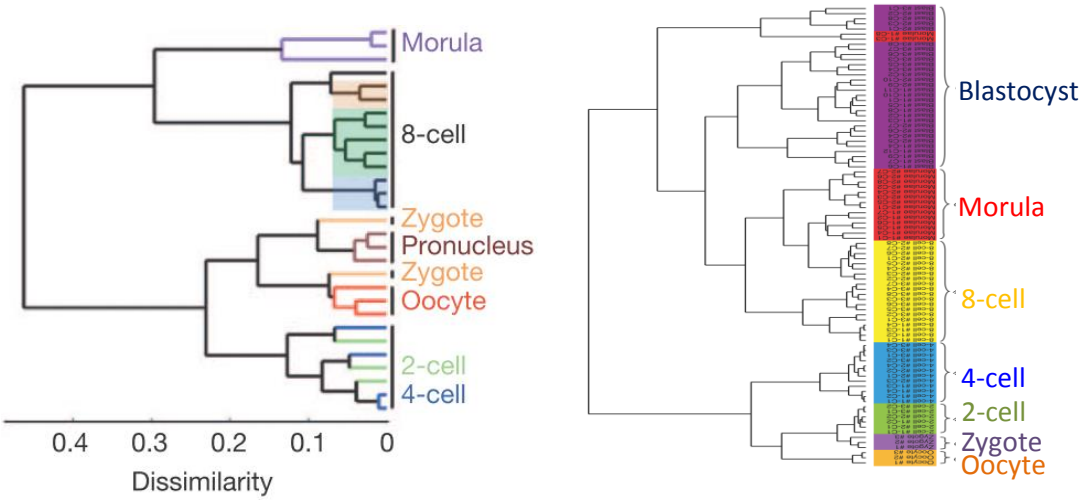

Figure 3. Hierarchical clustering of genes expressed during human preimplantation development. The left panel shows the result from Xue *et al*. and the right panel is a modified tree from Yan *et al.* for ease of comparison.

Both findings suggest that genes activated at the 8-cell stage are closest related to those activated in morula. However, due to different choices in linkage methods, Xue *et al.* classified oocyte and zygote to be more similar, while Yan *et al*. found zygote and 2-cell blastomeres to cluster more closely. To resolve the discrepancy between the two, one can take the data from both groups and cluster them using the alternative methods. If Xue *et al.*'s data show the same clustering pattern as Yan *et al*.'s when complete linkage method is applied, one can conclude that the difference arise solely from data clustering. However, if the same result is obtained with the alternative clustering method, the difference resides in data collection and processing. Since both studies used similar cell collection and sequence alignment methods, it will be interesting to combine both datasets and perform a uniform clustering algorithm to test the robustness of cluster assignment to different experimental conditions.

Alternative Clustering Methods and Validation

Another commonly used clustering algorithm is centroid-based K-means clustering. This method first generate *x* partitions and assign each partition a random partition vector of size *m* (number of cells in our case of interest). Each gene vector is subsequently binned into different partitions based on their similarity to the partition vector. Once the genes are partitioned, each partition vector is recalculated based on the gene vectors it contains, and all genes are reassigned. The partition vector serves as an effective centroid for each cluster, and the process iterates until a solution converges. The K-means clustering

method is sensitive to noise and outliers because all genes are forced into partitions each round and outliers can greatly affect their corresponding partition vectors. Moreover, results can fluctuate greatly based on the initial choice of partition number, making this method less desirable.

Various efforts have been made to improve the K-means clustering method, and one such modification is the Fuzzy C-Means (FCM) method[16]. Rather than force assigning each gene into a partition, FCM allows for weighted partial assignments. The partition vector is then calculated based on a weighted contribution of its gene members. Bandyopadhyay *et al.* took this idea a step further and proposed an improved two-step clustering algorithm[18]. They proposed that within datasets, there exist genes that can fall under multiple clusters, or have significant multi-class membership (SiMM). Their algorithm first identifies these SiMM points using a VGA-based method[19], excludes them from the dataset, performs FCM algorithm on the remaining points, and reassign SiMM points back to the FCM-determined partitions. This approach avoids the confusion regarding initial SiMM point assignments and greatly increases the clustering robustness to noise and outliers. Evidence suggests that a few groups of genes remain constitutively expressed throughout the embryonic preimplantation development[12]. This can make cluster assignments for these genes more difficult and the two-step SiMM clustering algorithm can offer an ideal solution to this problem.

In both studies, the authors validated their clustering results using gene ontology (GO) terms. Statistics are calculated to quantify the probably of associating a specific GO term to a cluster. Although this validation method gives biological relevance to the data, it relies on existing gene annotation in the databases[20]. An alternative method proposed by Yeung *et al*. bypasses this shortcoming[21]. This method determines a scalar quantity called the Figure of Merit (FOM) which measures the predictive power of a given clustering algorithm. First, one data point is removed from the data set, and the remaining set is clustered by any method of choice. Second, the excluded data point is placed back into a group based on its expression similarities to the clustered genes. One then calculates the root mean square deviation, or the FOM, of the excluded gene with respect to its cluster mean. This process iterates until all data points are excluded once, and the sum of FOMs is the score for the clustering algorithm of choice. The smaller the cumulative FOM is, the higher the predictive power a clustering algorithm has. Despite the lack of biological information, the FOM calculation enables direct and quantitative comparison between different clustering methods. This can serve as a powerful complementary validation method to the GO term assignments.

B. Data Visualization: Principal-Component Analysis

Clustering techniques can introduce artifacts and biases to data organization depending on the method of choice. To reveal the underlying structures in a multidimensional dataset without bias, the method of principal-component analysis (PCA) is often used[22]. In an experiment where m cells are sequenced and n total genes are expressed, the dataset has m points, each with n dimensions. For $n > 3$, the data become difficult to visualize. The goal of PCA is to

reduce the dataset down to only two or three dimensions and observe the underlying patterns. Briefly, each data point is represented by an eigenvalue (magnitude) and an eigenvector (direction). Then, the data points are projected onto an axis to give the largest variance along the line. This axis is known as the first principal component. To determine the subsequent principal components, the data are projected onto a different axis that gives the next largest variance. Each principal component axes must be orthogonal to each other to maximize sampling across the data space. For visualization purposes, data points are often represented by 2D or 3D plots as shown in **Fig. 4**.

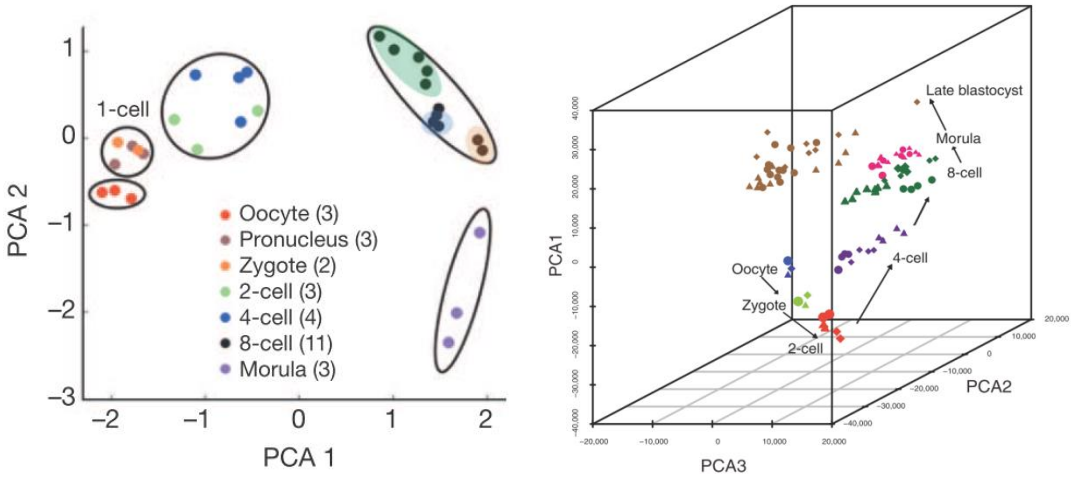

Figure 4. Principal-component analysis of transcriptomes from single blastomeres. Left panel shows results from Xue *et al*. and right panel shows results from Yan *et al*.

Through both hierarchical clustering and principal-component analysis, the authors observed that individual cells from different stages of preimplantation clustered separately by genetic profiling. This suggests that each developmental stage can be characterized by its unique set of transcriptome.

Findings using Single-Cell RNA-seq

In previous human embryo whole genome analyses, two prominent transcriptome dynamics were observed: the gradual degradation of maternal mRNA and the wave-like activation of zygotic genes. Specifically, the first wave of zygotic gene activation (ZGA) occurs at the 4-cell to 8-cell transition, and the associated gene functions are transcription and translation. A second wave of ZGA peaks at the morula stage as the embryo compacts. Pluripotency genes are highly upregulated at this time. Via microarray analysis, Vassena *et al*. identified 255 genes associated with ZGA[23]. Single-cell sequencing analyses also capture these features as shown by the heat maps (**Fig. 5**). Using a more stringent RPKM threshold, Xue *et al*. identified 149 differentially expressed genes between oocytes and zygotes, 70 of which were downregulated after fertilization and 79 were upregulated. Approximately half of the upregulated genes were highly expressed in 2-cell and 4-cell stages but not in 8-cell stage, while the other half were further upregulated after the 8-cell stage. They further noted that the first set of upregulated genes associate with transcription regulation and the second with RNA processing and splicing. Similarly, Yan *et al*. reported an approximate 1:1 ratio for upregulated versus downregulated genes between the 4-cell and 8-cell stages (2495 up vs. 2675 down). The genes highly expressed at 8-cell

stage are enriched for RNA metabolism and translation functions. These findings demonstrated that ZGA phenomena observed via traditional genomic profiling techniques can be reliably detected using single-cell RNA sequencing.

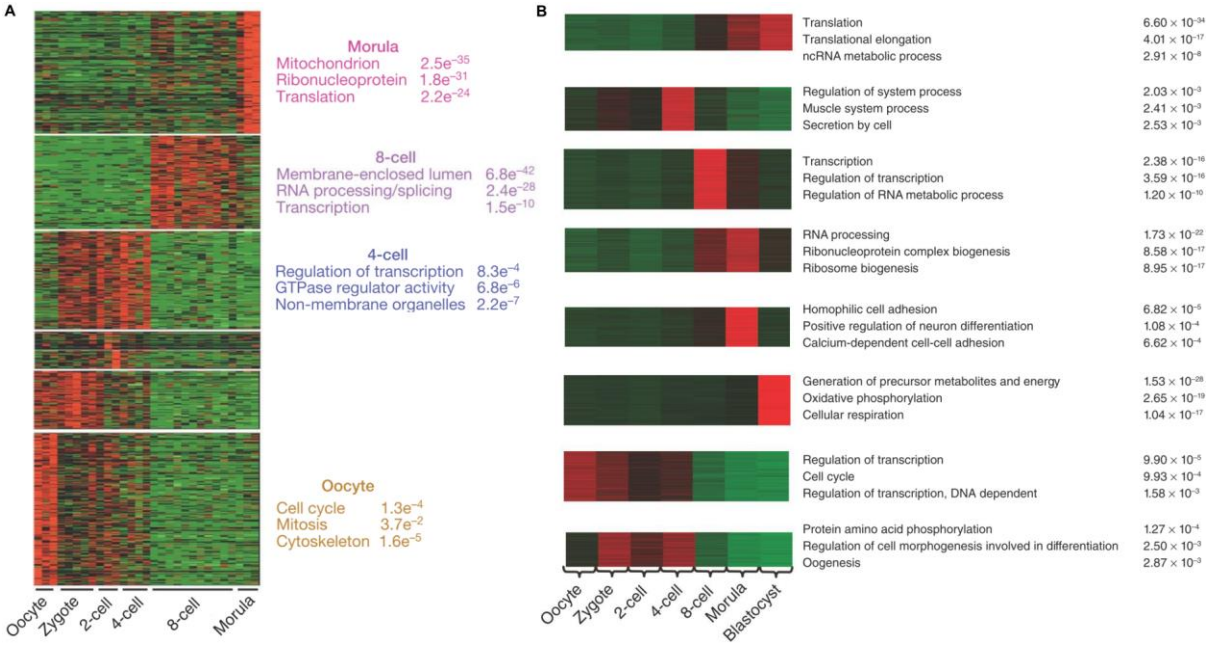

Figure 5. Heat maps indicating gene expressions across different preimplantation stages. A Xue *et al.* identified 149 differentially expressed genes between oocyte and zygote, and 7313 total genes across 7 stages of preimplantation development. Figure modified to ease comparison. **B** Yan *et al*. detected a total of 11,066 genes in sampled oocytes and blastomeres, 2495 of which were highly upregulated during ZGA. Genes activated at early preimplantation development are associated with cell cycle and cell division. After the 8-cell stage, blastomeres express RNA processing and metabolic genes.

Xue *et al.* further performed a weighted gene co-expression network analysis (WGCNA) to closer examine the links between each gene clusters. Their data consist of 25 coexpression modules and 9 of which were highly stage specific. The group then identified 491 intramodular hub genes across all stage-specific modules. When compared with the mouse dataset, Xue *et al*. found that both the stage-specific modules and the hub genes are highly conserved. They speculated that these modules act as the core gene networks governing each developmental stage, and the hub genes are key players in preimplantation stage transition. **Fig. 6A** highlighted some of the key hub genes found through WGCNA.

The advantage of single-cell sequencing over traditional genomic analysis lies in its ability to discern heterogeneity within a population at single-base resolution. Xue *et al*. took advantage of this feature and analyzed the parental genome contributions via single-nucleotide polymorphism analysis (scheme showed in **Fig. 6B**). They discovered that despite the gradual maternal mRNA degradation throughout the embryonic development, more than 50% of the genes activated at the 8-cell stage still exhibited monoallelic maternal expression patterns (**Fig. 6C**). One such example was the *ASB6* locus. In contrast, maternally activated cell-cycle regulator *CDCA2* during 2-cell and 4-cell stages transitioned to a transient paternal activation at the 8-cell

stage. The authors also identified several SNPs in *CDCA2* that can result in missense variants of the gene. They argued that single-cell sequencing combined with SNP analysis can be a powerful tool to screen for deleterious variants in embryos.

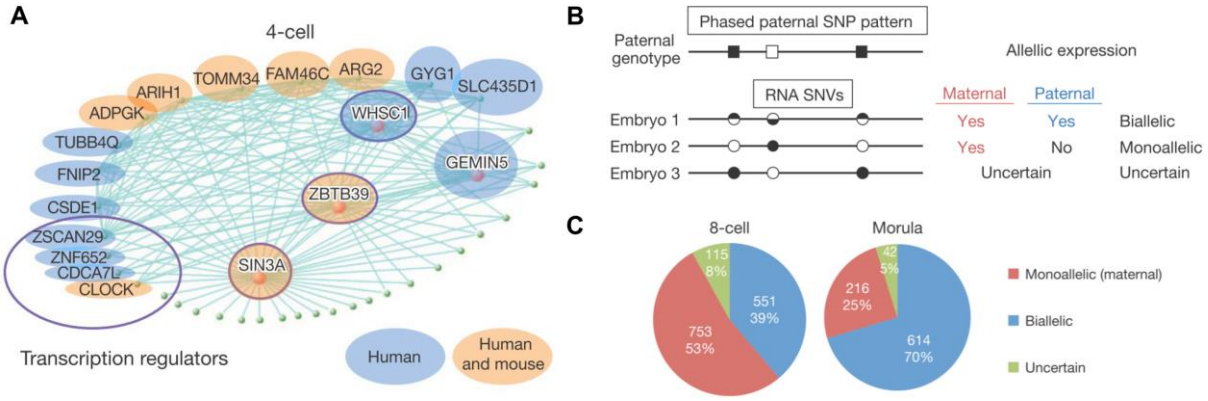

Figure 6. WGCNA and SNP analyses on single-cell RNA sequenced human embryos[7]. **A** A Network connecting the coexpression modules and the hub genes. Highly connective hub genes are highlighted by red dots. Transcription regulators are circled in purple. Genes shaded in blue and orange were independently validated. **B** Schematics of allelic expression assignment based on SNP patterns. **C** Pie charts summarizing the distribution of assigned genes exhibiting single-nucleotide variants.

RNA modification is a key signature of the 8-cell stage as identified previously by gene annotation. Yan *et al*. followed this lead and analyzed the transcriptome for alterative splice variants, long noncoding RNAs (lncRNA) and novel RNA transcripts. They isolated 2372 distinct genes with more than 2 isoforms, and 206 of which were present in all stages of preimplantation development. They found that in 34% of the alternatively spliced genes, two or more isoforms were equally expressed; while in the remaining genes, one major isoform dominated. Importantly, the isoforms can be differentially expressed at different stages. For example, the authors found that *FOXP1* exon 18b was enriched in the blastocyst stage but not in the 2-cell or 4-cell stages. This isoform targets a different DNA sequence and is required for pluripotency. For the lncRNA analysis, Yan *et al*. detected 18,383 total unique lncRNAs. On average, each blastomere contained ~3000 lncRNAs, which accounted for 5% of the transcriptome. When the authors analyzed these noncoding RNAs more closely, they found that in certain populations, lncRNA had copy numbers up to 40% of those for the coding transcript copy numbers. They also found that lncRNA expression was stage-specific (**Fig. 7A**), suggesting a regulatory role for lncRNA in preimplantation development[24]. Lastly, through this transcriptome analysis, the group identified 253 possible novel protein coding genes and 2733 lncRNAs. 40% of these novel lncRNAs were conserved amongst all blastomeres from the same stage, again emphasizing the potential regulatory function of lncRNAs.

Microarray and whole-genome sequencing analyses on human embryos were first aimed towards the discovery of pluripotency-specifying genes. Yan *et al*. further sequenced two different stages of human embryonic stem cells (hESCs, passage 0 and passage 10) and compared them with the blastocyst inner cell mass (ICM), where hESCs are derived from. Consistent with previous findings, ICM cells clustered

independently from hESCs (**Fig. 7B**). 975 genes were upregulated and 523 genes were downregulated during the ICM to hESC transition. Notably, many of the pluripotency associated genes including *SOX2*, *TDGF1*, *NODAL* and *LEFTY1/2* were upregulated by more than 5 folds in comparison to the expressions in ICM, while others like *OCT4, NANOG, KLF4* and *STELLA* were significantly downregulated. The authors detected no significant differences in gene expressions between P0 and P10 hESCs. 138 known and 37 novel long noncoding RNAs were identified unique to ICM and 2286 known and 194 novel lncRNAs were found unique to hESCs. Further comparison between hESCs and cells from the primitive endoderm (first stage after blastocyst implantation) showed a large degree of dissimilarity. Together, Yan *et al*.'s results support the notion that ICM cells and their derivative hESCs are two distinct populations of pluripotent cells. The marked difference in lncRNA level provides a new avenue to probe its regulatory role, which may be a crucial control for the differential gene expressions in the two populations.

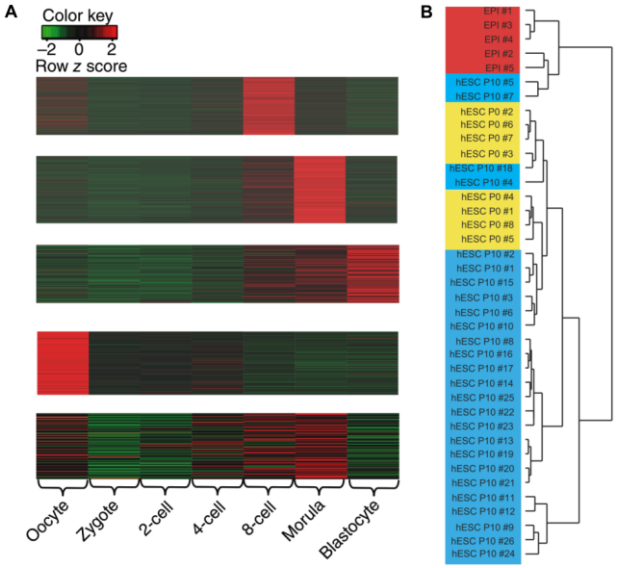

Figure 7. **Selected results from Yan *et al*.'s study on embryo and hESC genomic sequencing. A** Hierarchically clustered lncRNA expression during human preimplantation development. Many lncRNA clusters were expressed in a stage-specific manner. **B** Hierarchical clustering of genes expressed in late blastocyst inner cell mass (epiblast, EPI) and in human embryonic stem cells (hESC). Epiblasts are shaded in red, newly generated hESC (passage 0, P0) in yellow, and later stage hESC (passage 10, P10) in blue.

Summary and Future Prospects

In summary, single-cell RNA sequencing serves as a powerful tool to study the human preimplantation development. Studies by Xue *et al.* and Yan *et al.* independently demonstrated that each developmental stage can be characterized by a unique set of transcriptomes. Both groups contributed a comprehensive list of genes associated with major events in development such as the maternal mRNA degradation and the zygotic gene activation. Xue *et al*. took advantage of the single-base resolution in single-cell RNA-seq and investigated the SNP in each blastomere. Yan *et al.* leveraged the knowledge from gene annotation and further explored RNA alternative splicing and lncRNA expression throughout the preimplantation process. In broader context, Xue *et al*. constructed a network linking all the stage-specific gene modules and identified several hub genes that may be central to the transition between stages. Yan *et al.* compared gene expressions of cells from the ICM, hESCs and primitive endoderm and found that three cell types are distinct populations of its own. The stage-specific and population-specific lncRNA expression may be the key to regulate differential gene expression across the human preimplantation stages.

From a technique standpoint, both studies utilized similar sequencing and analysis methods. cDNA libraries are constructed using PCR amplification, followed by Illumina sequencing and BWA alignment and genome mapping. Hierarchical clustering algorithm and principal-component analysis were used to visualize the data, and gene annotations were used to validate the clustering results. Although hierarchical clustering is easy to implement and has revealed important biological insights, this algorithm may not be the best fit to describe embryonic preimplantation gene expressions. Genes constitutively expressed throughout the developmental process can be difficult to cluster into any particular stage. With a two-step significant multi-class membership (SiMM) algorithm, genes that have the characteristics of multiple clusters can be more accurately placed. In addition to GO term validation, the Figure of Merit index can be used to estimate and compare the predictive power of various clustering methods. The review by Ning *et al*. provides additional improvement suggestions for single-cell sequencing technique and analysis[25].

The insights gained from Xue *et al*. and Yan *et al.*'s analyses provide multiple directions forward. One can use the SNP analysis to trace the paternal gene contribution to an embryo, or to construct a list of deleterious variants to incorporate into future embryo diagnosis. The list of hub genes provide good leads for proteomic studies in human embryos. These genes may translate into key regulators that gate the transition from one stage to another. Systematic knockout experiments will reveal their roles in early development. The stage-specific and population-specific expression of long noncoding RNAs is yet another finding that stands in need for further investigation. lncRNAs can exert transcriptional, post-transcriptional, as well as epigenetic regulation on gene expressions. Uncovering their function in blastomeres and hESCs will be a significant step forward for developmental and stem cell biology.

In conclusion, single-cell RNA sequencing technique is well-suited for human preimplantation development studies. It accommodates the scarcity of human embryo samples while producing reliable results in agreement with traditional microarray and whole-genome studies. Moreover, single-cell RNA-seq detects heterogeneity within a cell population at a single-nucleotide level. This degree of sensitivity has allowed many novel studies that have previously been difficult, if at all possible to proceed. Efforts have been put to combine single-cell RNA-seq with other techniques including live cell imaging and fluorescent protein reporters[13]. These hybrid methods, together with rapidly improving sequencing analysis algorithms, will serve as a promising next-generation analytical platform to study the dynamic gene expression network in any complex biological systems.

References

[1] Gardner D.K. *et al*. (2000) “Blastocyst score affects implantation and pregnancy outcome: towards a single blastocyst transfer”. *Fertil Steril*. **73**(6), 1155-8.

[2] Reijo Pera, R.A. *et al.* (2009) “Gene expression profiles of human inner cell mass cells and embryonic stem cells”. *Differentiation* **78**, 18-23.

[3] Passier, R. (2003) “Potential of human embryonic stem cells in regenerative medicine”. *Horm Res.* **60**, 11-14.

[4] Richards, M. *et al*. (2004) “The transcriptome profile of human embryonic stem cells as defined by SAGE” *Stem Cells* **22**(1), 51-64.

[5] Takahashi, K. & S. Yamanaka (2006) “Induction of pluripotent stem cells from mouse embryonic and adult fibroblast” *Cell* **126**(4), 663-676.

[6] Bhattacharya, B. *et al*. (2009) “A review of gene expression profiling of human embryonic stem cell lines and their differentiated progeny” *Curr Stem Cell Res Ther.* **4**(2), 98-106.

[7] Xue, Z. *et al*. (2013) “Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing” *Nature* **500**, 593-597.

[8] Yan, L. *et al*. (2013) “Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells” *Nature Structural & Molecular Biology* **20**(9), 1131-1139.

[9] Gilbert SF. (2000) Developmental Biology. 6th ed. Sunderland (MA): Sinauer Associates. Fusion of the Genetic Material. http://www.ncbi.nlm.nih.gov/books/NBK10055/

[10] Forgács, G. & Newman, Stuart A. (2005). "Cleavage and blastula formation". *Biological physics of the developing embryo*. Cambridge University Press. p. 27.

[11] Thomson, J.A. *et al*. (1998) “Embryonic stem cell lines derived from human blastocysts” *Science* **282**(5391), 1145-1147.

[12] Wong, C.C. *et al*. (2010) “Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst stage” *Nature Biotechnology* **28**, 1115-1121.

[13] Tang F. *et al*. (2011) “Development and applications of single-cell transcriptome analysis” *Nature Methods* **8**(4).

[14] Metzker, M.L. (2010) “Sequencing technologies – the next generation” *Nature Reviews Genetics* **11**, 31-46.

[15] "List of RNA-Seq bioinformatics tools" *Wikipedia: The Free Encyclopedia*. Wikimedia Foundation, Inc. Web. 12 Mar. 2014. < http://en.wikipedia.org/wiki/List_of_RNA-Seq_ bioinformatics_tools>

[16] Kerr, G. *et al*. (2007) “Techniques for clustering gene expression data” *Comput Biol Med.* **38**(3), 283-293.

[17] Sherlock, G. (2010) “Clustering and functional analysis of coordinately regulated genes” PowerPoint presentation. Web. 12 Mar. 2014 <http://biochem218.stanford.edu/>

[18] Bandyopadhyay, S. *et al*. (2007) “An improved algorithm for clustering gene expression data” *Bioinformatics* **23**(21), 2859-2865.

[19] Maulik, U. & S. Bandyopadhyay (2003) “Fuzzy partitioning using a real-coded variable-length genetic algorithm for pixel classification” *IEEE Trans. Geosci. Remote Sens.* **41**, 1075-1081.

[20] Gibbons, F.D. & F.P. Roth (2002). “Judging the quality of gene expression-based clustering methods using gene annotation” *Genome Res.* **12**(10), 1574-1581.

[21] Yeung, K.Y. *et al*. (2000) “Validating clustering for gene expression data” *Bioinformatics* **17**(4), 309-318.

[22] Smith, L.I. (2002) “A tutorial on principal components analysis” Web. 13 Mar. 2014 <http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf>

[23] Vassena, R. *et al*. (2011) “Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development” *Development* **138**, 3699-3709.

[24] Ponting, C.P. *et al*. (2009) “Evolution and functions of long noncoding RNAs” *Cell*, **136**(4), 629-641.

[25] Ning, L. *et al*. (2014) “Current challenges in the bioinformatics of single cell genomics” *Frontiers in Oncology* **4**, 7.