The Transcriptome of a Human Polar Body Accurately Reflects Its Sibling Oocyte

Improved methods are needed to reliably and accurately evaluate oocyte quality prior to fertilization and transfer into the woman of human embryos created through in vitro fertilization (IVF). All oocytes that are retrieved and matured in culture are exposed to sperm with little in the way of evaluating the oocyte quality. Furthermore, embryos created through IVF are currently evaluated for developmental potential by morphology, a criterion lacking in quantitation and accuracy. With the recent successes in oocyte vitrification and storage, clear metrics are needed to determine oocyte quality prior to fertilizing. The first polar body (PB) is extruded from the oocyte before fertilization and can be biopsied without damaging the oocyte. Here, we tested the hypothesis that the PB transcriptome is representative of that of the oocyte. Polar body biopsy was performed on metaphase II (MII) oocytes followed by single-cell transcriptome analysis of the oocyte and its sibling PB. Over 12,700 unique mRNAs and miRNAs from the oocyte samples were compared with the 5,431 mRNAs recovered from the sibling PBs (5,256 shared mRNAs or 97%, including miRNAs). The results show that human PBs reflect the oocyte transcript profile and suggest that mRNA detection and quantification through high-throughput quantitative PCR could result in the first molecular diagnostic for gene expression in MII oocytes. This could allow for both oocyte ranking and embryo preferences in IVF applications.

The clinical importance of healthy oocyte development is evidenced by the impressive pregnancy rates seen with infertile women using assisted reproductive technology with oocytes from young, fertile donors. Oocytes from young women have lower rates of meiotic errors and aneuploidy, and although aneuploidy is the most common cause of developmental arrest, screening embryos for aneuploidy does not exclude all embryos of poor prognosis. Earlier studies have demonstrated that as the primary oocyte develops, it transcribes thousands of genes whose products are necessary for fertilization and early embryonic development. Prior to meiosis I, the germinal vesicle breaks down and transcriptional factors disengage from chromatin, rendering the cell transcriptionally silent (1). This is particularly relevant as the human zygotic genome is not activated and does not transcribe its DNA for 2–3 days following fertilization (2, 3), the exact period during which an IVF clinician decides which embryo to transfer into the woman (4, see especially Fig. 37). Therefore, mRNAs needed for fertilization and early embryonic development must be present in the oocyte in sufficient quantity or ratio before the first polar body is extruded and guide the majority of embryonic processes to day 3. Thus, the transcriptome of the oocyte may predict both oocyte quality and the early developmental potential of the embryo.

The ability to measure oocyte gene expression without harming the oocyte may prove helpful to clinicians caring for patients using assisted reproductive technology. Biopsying the polar body would allow embryologists to test for functional control of gene expression in the oocyte resulting from mRNA transcription, turnover, and from epigenetic processes that depend on more complex determinants than having an appropriate number of chromosomes (5–7), all without compromising the oocyte.

The transcriptome of a polar body has never been reported, and a polar body biopsy involves its careful removal through microdissection. This procedure can be performed without damaging the sibling oocyte or developing embryo (8) and with advances in oocyte vitrification (9), this could be helpful for patients with ethical objections to fertilizing multiple oocytes and creating supernumerary embryos, or it can be applied to the growing practice of oocyte vitrification for donor egg banking. One also can imagine using gene expression information...
**Polar Body Transcriptome**

from a polar body to prioritize embryos for transfer in an IVF cycle. Here, we are the first to report the analysis of polar body transcriptomes from any organism and analyze its miRNA population with that of its sibling oocyte.

**MATERIALS AND METHODS**

*Human Oocyte Collection and Polar Body Biopsy*—Human oocytes were collected from infertility patients undergoing controlled ovarian hyperstimulation for IVF under standard clinical protocol. Germlinal vesicle and MI staged oocytes that were not mature for a clinically indicated intracytoplasmic sperm injection procedure underwent *in vitro* maturation for 24 h and were used in the study if they extruded a polar body. Written consent was obtained from all patients to use discarded tissue and oocytes for research, and the study was approved by the institutional review board at Women & Infants Hospital. Briefly, patients underwent controlled ovarian hyperstimulation, with either luteal down-regulation using a GnRH agonist, pituitary suppression using a GnRH antagonist, or a microdose lutein “flare protocol” consisting of daily lutein injections initiated in the follicular phase with gonadotropins. Oocytes were aspirated by ultrasound guided transvaginal oocyte retrieval 36 h after injection with recombinant human chorionic gonadotropin. Four hours after retrieval, all oocytes were mechanically stripped of cumulus cells. Intracytoplasmic sperm injection was performed in all oocytes with visible polar bodies. After injection of all MII oocytes, any remaining immature oocytes were cultured for 20–24 h in IVM media. Immature oocytes were examined the next day, and oocytes that extruded a polar body were used for our study. A total of 22 oocytes and sibling polar bodies were collected and individually processed in this study.

**Biopsy and WTA Amplification**—All biopsies were performed at 200X magnification after mechanical zona drilling with a polar body biopsy needle (Cook Medical, Bloomington, IN). Polar bodies were aspirated into a glass micropipette with a polar body biopsy needle (Cook Medical, Bloomington, IN). The polar body was then processed in this study.

In an attempt to maximize cDNA yield, 10 μl of the amplified cDNA was added to 65 μl of amplification mix, and the contents underwent an additional 15 rounds of amplification specific for cDNA containing the WTA products. The second amplification was combined with the remainder of the first amplification step. The final concentration for all libraries was between 30 and 40 ng/μl as measured by QuBit (Invitrogen). Each individual oocyte and polar body were processed in separate reactions and for those samples that were pooled, 10 oocytes or polar bodies were pooled together in a common tube after the two rounds of WTA amplification. The 22 oocyte and sibling polar body pairs (44 cells) were split into two replicates of 10 pooled cells and two replicates of a single cell, for a total of eight samples, four oocyte samples, and four polar body samples.

**Illumina Library Preparation and Sequencing**—The cDNA resulting from two rounds of the WTA amplification yielded fragments between 100 and 300 bp in length. The cDNA was not subjected to any additional shearing, and libraries were prepared using the NEBNext DNA Sample Prep Kit (New England Biolabs) with adapters and PCR primers from Integrated DNA Technologies (Coralville, IA). The standard protocol was used with starting material of no less than 1.5 μg of total cDNA with one slight modification. After the ligation of the adapters to the ends of the cDNA molecules, PCR amplification was performed without an intervening gel purification. After the PCR step, gel purification of the completed library was performed and a wide band of 200–450 bp was cut from the gel. The library concentration was determined by quantitative PCR (Kapa Biosystems, Woburn, MA) and size by Bioanalyzer (Agilent Technologies, Santa Clara, CA). The samples were sequenced for 42 bp on a GAIIx (Illumina, Inc., San Diego, CA) using a custom sequencing primer consisting of the 6 bp most 3’ of the Illumina sequencing primer fused to the WTA2 primer sequence (Fig. 1b).

**Mapping and Statistical Analysis**—The raw sequences were mapped against the human genome (UCSC hg18) using Illumina software, Casava (version 1.7) using 32 bp of the read and allowing only two mismatches. The raw gene counts were then loaded into edgeR (21), which normalized the counts using the TMM method (trimmed mean of M values) (22), and these counts were used for all further analyses. The geometric means of the TMM counts of every gene for all oocyte or polar body samples was used to generate a list of the expression levels of all genes. The R package ranked list comparison (23) was used to analyze the expression level lists.

**RESULTS AND DISCUSSION**

**Analysis of Detected Genes and Gene Expression Levels**—We analyzed the transcriptomes of 22 oocytes alongside their 22 polar body siblings by high-throughput DNA sequencing. The samples were grouped into two biological replicates of 10 oocytes and their sibling polar bodies and two biological replicates of single oocytes and their single polar bodies. We developed a method for quantitative cDNA construction from both a single oocyte and its sibling polar body, and we detected a total of 12,883 genes through mapping of >27 million reads from these oocytes and polar bodies (Table 1). From this result, we
estimate that between 14,000–15,000 genes are expressed in the human oocyte (supplemental Fig. 1). The genes that were expressed in each oocyte highly correlated with those that were expressed in other oocytes. Of the 7,523 genes detected in the smallest oocyte sample, 84.5% of the genes were detected in all four oocyte samples and 98% were detected in at least two samples (Fig. 2a). Furthermore, of the four oocyte/polar body pairs in this study, 90% of all the genes detected in a polar body sample were also detected in the sibling oocyte sample (Fig. 2b). This result might be expected because the polar body and oocyte shared a common ooplasm no more than 24 h prior to polar body biopsy, but no less surprising because of the diversity of the transcripts detected in the polar body samples. Comparing the overlap of each sibling oocyte/polar body pair with every other pair reveals that 279 genes (28.0% of genes detected in the smallest overlap pool) are expressed in all four paired samples (Fig. 2c and supplemental Table 1); 962 genes were detected in both the oocyte and sibling polar body samples in at least three of the four pairs (Fig. 2c).

Although the sample to sample overlap of genes was very high in all examined comparisons, a critical component of the analysis was testing the abundance of those genes products. We performed a pairwise comparison of each oocyte sample and tested the levels of all gene products shared between each oocyte pair. The expression level of any gene in any oocyte sample correlated very strongly (Pearson correlation > 0.88) with the expression level in any other oocyte (supplemental Fig. 2). One hypothesis is that the polar body is a depository of the oocyte, that transcripts or cellular contents no longer needed or undesirable in the oocyte are transported to the polar body. We tested this hypothesis by doing a differential gene enrichment analysis of all oocyte samples against all polar body samples and found no genes that were differentially enriched between the two populations at any levels of significance (supplemental Fig. 3). Multiple normalization methods were used in this testing, and all results arrived at the same conclusion: the transcriptome of the polar body accurately reflects its sibling oocyte. The observation that transcript abundance is very similar between oocytes and polar bodies for all detected genes strongly argues against the interpretation that messages are selectively transported into or out of the polar body. A more parsimonious explanation is that as the polar body is extruded, it captures a representative portion of the ooplasm and therefore a representative transcriptome. Furthermore, no genes were sampled in all four polar body samples without being detected in a single

TABLE 1

| Sequecing reactions | No. of mappabel reads | No. of genes found |
|---------------------|-----------------------|--------------------|
| Pool of 10 oocytes 1 | 5,363,802 | 9,765 |
| Pool of 10 oocytes 2 | 1,731,513 | 7,525 |
| Pool of 10 polar bodies 1 | 326,176 | 1,053 |
| Pool of 10 polar bodies 2 | 207,303 | 2,224 |
| Single oocyte 1 | 3,692,168 | 8,681 |
| Single oocyte 2 | 14,573,567 | 11,881 |
| Single polar body 1 | 904,103 | 2,293 |
| Single polar body 2 | 376,441 | 3,515 |
| Total of all oocyte samples | 25,361,050 | 12,708 |
| Total of all polar body samples | 1,814,023 | 5,431 |
| Total of all samples | 27,175,073 | 12,883 |
oocyte, further evidence that specific transcripts are not localized to the site of the budding polar body and that the polar body inherits the general transcriptionally silenced state of the oocyte (1). These results support the interpretation that the transcriptome of one oocyte is very similar to that of another oocyte in both the genes expressed and the expression levels and that the transcriptome of a polar body is directly representative of the transcriptome of its sibling oocyte.

Examination of Gene Expression Profiles—We took particular care to examine the data using the WTA procedure coupled with Illumina sequencing. A potential concern was that of amplification bias, but a formal test using the Ambion Cells-to-Ct kit and quantitative PCR, which has been previously tested to linearly amplify transcripts (10), would have been prohibitively costly. However, the relative rank abundance of half a dozen previously reported transcripts (10) was recapitulated in this study using the WTA method. A second concern was the introduction of contaminants or the amplification of PCR mutations due especially to the two rounds of amplification in this study, potentially leading to mapping error. To control for this variable, we examined the number of reads that mapped to genes on the Y chromosome; because this study was conducted on oocytes and polar bodies that had never been exposed to sperm, there is no biological template for Y chromosome genes. Of the >27 million reads in this study, only 128 reads ($<5 \times 10^{-4}$% of the total reads) mapped to two genes found on the Y chromosome (USP9Y and DDX3Y), both of which have paralogs on the X chromosome (USP9X and DDX3X). The percent identity for the two sets of paralogs are very high (USP9 = 88.3% and DDX3 = 88.8%); therefore, Y chromosome mapping could be due to sequencing errors or statistical mapping error. The third and final concern was the recovery of sufficient material from a single oocyte, much less from a single polar body. To address this, we used two different types of samples: pooled and single specimens. We felt that the pooled specimens would decrease the oocyte to oocyte variability in gene expression and ensure that the cDNA would not be limiting. The single specimens would serve as a test of the clinical feasibility of reliably detecting transcripts in single cells as well as testing the variability of oocyte to oocyte differences in transcript abundance.

We used the bioinformatics package DAVID (11) to test whether the genes that were detected in all oocyte samples completed specific enzymatic pathways. Of all of the detected genes in the oocyte samples, several dozen KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways were significantly complete, including “oocyte meiosis” ($p$ value = $4.5 \times 10^{-7}$) and “Progesterone-mediated oocyte maturation” ($p$ value = $3.6 \times 10^{-4}$) (supplemental Fig. 4), providing additional evidence that the transcriptome we are detecting are representative of oocytes.

Intriguingly, a number of miRNAs in oocytes and polar bodies were detected, some of which are very abundant. Previous reports suggest that the miRNA pathway is non-functional in mouse oocytes (12, 13), and because our sequencing technique would have missed mature miRNAs, we hypothesize that the high abundance of miRNAs is due to the presence of pre-processed miRNAs. The down-regulation of the miRNA pathway may be due to the repression of Drosha/DGCR8, which would allow preprocessed hairpins to accumulate. This hypothesis is supported in part by the observation that the transcriptome of the mouse oocyte is minimally impacted by the loss of zygotic DGCR8 during early development (13). An abundance of preprocessed miRNA-containing transcripts in the nucleus could explain the slightly higher abundance (statistically non-significant) of miRNA genes in the polar body compared with the oocyte. The tran-

FIGURE 2. a, all four oocytes samples show a high degree of overlap with each other. The total number of genes for each sample is shown in parentheses, and the total number of all genes in all four samples is equal to 12,708 genes. 66.7% of all genes were detected in at least three of the four oocyte samples, and 50.0% of all genes were detected in all four oocyte samples. b, the larger circle represents the total number of genes in each of the four oocyte samples, and the smaller circle shows the overlap of the four sibling polar bodies. The percentage was calculated by taking the total number of genes shared between the sibling oocyte and polar body and dividing by the total number of genes in the polar body. c, the overlaps of genes transcribed in sibling oocyte and PB samples among the four independent comparisons are represented as described in a. In total, there were 4,973 genes found between all the overlap data sets and 279 genes that were sampled in all eight samples. Of the 4,973 overlap gene set, ~46% were detected in at least two of the overlaps.
scriptome of the polar body may be enriched for mRNAs that associate with the meiotic spindle. Because the polar body is significantly smaller than the oocyte, it may be unduly enriched for such transcripts.

We also report that the abundance of some transcripts may vary significantly between single cells. The most notable example was the transcript level of the Anthrax receptor, ANTXR2, which varied in expression between the two single oocytes by 4 orders of magnitude. Remarkably, this difference in abundance was also reflected in the sister polar body by the same 4 orders of magnitude difference. Previous literature has shown ANTXR2 to be detected in oocytes (14) and significantly downregulated (15) in human oocytes compared with a reference. Such variation may reflect atypical regulation resulting from differences in the oocyte genome, differences in nutritional status of the oocyte, of donor age, and/or environmental influences on the oocyte.

**Clinical Feasibility Test and Microarray Comparison**—For successful transition into a clinical application using this approach, it is important to identify a cohort of transcripts that are reliably detected, whose expression levels are predictive of the developmental competence of a given oocyte and its resulting embryo. To develop a list of candidate transcripts for future studies, we generated a separate rank order list of the transcript abundance for oocytes and for polar bodies by taking the geometric means of the TMM-normalized gene counts in all four samples. The two separate lists were then compared with each other in discrete subsets to test for significant overlap between the oocyte and polar body samples (combined p value = 6.16 e^{-250}) and within the 700 genes, 460 genes are shared between the oocyte and polar body lists (Fig. 3). Nearly half of the 460 shared genes (215 genes) were detected in all oocyte and polar body samples (Fig. 2c and supplemental Table 1).

Analysis of our data set strongly suggests that the polar body captures a representative transcriptome of the oocyte and that the transcriptomes of both single oocytes and single polar bodies can be assessed quantitatively. Although several microarray studies of human oocytes have been reported (14–19), this is also the first study that uses deep sequencing technology on human oocytes. We compared our sequencing data set to that of three microarray studies to validate our findings (14–16). Those genes that were most highly enriched in oocytes compared with the reference in microarray studies were more likely to be the same genes that were most abundant in our dataset (Fig. 4). Additionally, the genes reported to be significantly enriched in the reference compared with the oocytes were more likely to be genes detected at very low abundance in our oocyte data set. The total number of genes detected in polar bodies was highly variable between each sibling pair of oocytes and polar bodies, but the genes most abundant in oocytes were much more likely to be detected in one or more polar body samples. Furthermore, genes that were previously reported to be detected in oocytes and polar bodies by quantitative PCR (10) were present in the same rank abundance in this study. One potential concern we had for biopsying the polar body was the fear that instead of sequencing the transcriptome of the polar body, we would contaminate the sample with a cumulus cell or other accessory cell. Comparing our data to that of a microarray study of human oocytes and cumulus cells, the
polar body transcriptome is distinct from cumulus cells and instead very closely aligns with oocyte samples (Fig. 4).

CONCLUSIONS

We demonstrated that detection and quantification of mRNA in polar bodies is possible and reflects the transcript profile of the MII oocyte. The quantification of mRNAs is of particular importance, and we also report that some transcripts can have highly variable expression in different oocytes and that this variance can be reflected in the polar body. This variance may reflect many factors seen by the oocyte during its development including nutritional status and environmental insults to the oocyte. Genes with higher levels of expression in oocytes are more reliably detectable in the sibling polar body, suggesting that a failure to identify a particular mRNA in the polar body relates to transcript levels within the oocyte that fall below a critical threshold. This finding is consistent with our previous results in both human and sea star oocytes (10, 20). Our results suggest that the detection and analysis of polar body mRNA may provide insight into oocyte quality, a critical metric needed by the clinical before fertilization and transfer of the resulting embryo back into the women.

FIGURE 4. The 6,355 genes that were sampled in all four oocyte samples were compared with the polar body samples as well as previously published microarray studies of oocytes. Comparing the natural log of the geometric mean of the TMM normalized counts of the oocyte samples (lane 1) and the polar body samples (lane 2); genes that are most abundant in the oocytes samples (dark blue) are more likely to be sampled in the polar body samples and are also found at similar expression levels. Select genes are shown on the left, including genes that have been shown to be highly expressed in oocytes (black text) (10) as well as genes significantly up-regulated in cumulus cells (green text) that were found (16). There is a high variation between the individual polar bodies (lane 3), but genes expressed at a high level in oocytes are more likely to be detected in polar bodies (black) than not (white). There is a very strong correlation between genes that are significantly up-regulated (lane 4) or detected (lane 5) in oocytes with microarray studies and the genes that were most abundant in oocytes and polar bodies in this study. The inverse correlation is found with genes that were significantly up-regulated in the reference compared with this data set.
Acknowledgments—We especially thank Linda Sousa and Dr. Jean-nine Witmyer for assistance in performing the polar body biopsies, Christoph Schorl for microarray analysis and sequencing support, Lingsheng Dong for bioinformatics software help, the Center for Computing and Visualization at Brown for hardware support, Nicola Neretti and Michael Antosh for help with the ranked list R package, Brian Ward for synthesizing and helping design the custom sequencing primer, Casey Dunn for insightful discussions, and Domenic Fenoglio for the gift of the WTA2 kits.

REFERENCES

1. Sun, F., Fang, H., Li, R., Gao, T., Zheng, J., Chen, X., Ying, W., and Sheng, H. Z. (2007) Cell Res. 17, 117–134
2. Braude, P., Bolton, V., and Moore, S. (1988) Nature 332, 459–461
3. Vassena, R., Boué, S., González-Roca, E., Aran, B., Auer, H., Veiga, A., and Belmonte, J. C. (2011) Development 138, 3699–3709
4. Centers for Disease Control and Prevention, American Society for Reproductive Medicine, Society for Assisted Reproductive Technology (2010) 2008 Assisted Reproductive Technology Success Rates: National Summary and Fertility Clinic Reports, U.S. Department of Health and Human Services, Atlanta, GA
5. Biddle, A., Simeoni, L., and Gurdon, J. B. (2009) Development 136, 2695–2703
6. Evsikov, A. V., and Marin de Evsikova, C. (2009) Mol. Reprod. Dev. 76, 805–818
7. Seli, E, Lalioti, M. D., Flaherty, S. M., Sakkas, D., Terzi, N., and Steitz, J. A. (2005) Proc. Natl. Acad. Sci. U.S.A. 102, 367–372
8. Verlinsky, Y., Ginsberg, N., Lifchez, A., Valle, J., Moise, J., and Strom, C. M. (1990) Hum. Reprod. 5, 826–829
9. Rienzi, L., Romano, S., Albricci, L., Maggiulli, R., Capalbo, A., Baroni, E., Colamaria, S., Sapienza, F., and Ubaldi, F. (2010) Hum. Reprod. 25, 66–73
10. Klatsky, P. C., Wessel, G. M., and Carson, S. A. (2010) Mol. Hum. Reprod. 16, 938–943
11. Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009) Nat. Protoc. 4, 44–57
12. Ma, J., Flemr, M., Stein, P., Berninger, P., Malik, R., Zavolan, M., Svoboda, P., and Schultz, R. M. (2010) Curr. Biol. 20, 265–270
13. Suh, N., Baehner, L., Moltzahn, F., Melton, C., Shenoy, A., Chen, J., and Blelloch, R. (2010) Curr. Biol. 20, 271–277
14. Grondahl, M. L., Yding Andersen, C., Bogstad, J., Nielsen, F. C., Meinezt, H., and Borup, R. (2010) Hum. Reprod. 25, 957–968
15. Kocabas, A. M., Crosby, J., Ross, P. I., Otu, H. H., Beyhan, Z., Can, H., Tam, W. L., Rosa, G. J., Halgren, R. G., Lim, B., Fernandez, E., and Cibelli, J. B. (2006) Proc. Natl. Acad. Sci. U.S.A. 103, 14027–14032
16. Assou, S., Anahory, T., Pantesco, V., Le Carrou, T., Pellestor, F., Klein, B., Reytmann, L., Dechaud, H., De Vos, J., and Hamamah, S. (2006) Hum. Reprod. 21, 1705–1719
17. Bermúdez, M. G., Wells, D., Malter, H., Munné, S., Cohen, J., and Steuerwald, N. M. (2004) Reprod. Biomed. Online 8, 325–337
18. Gasca, S., Pellestor, F., Assou, S., Loup, V., Anahory, T., Dechaud, H., De Vos, J., and Hamamah, S. (2007) Reprod. Biomed. Online 14, 175–183
19. Jones, G. M., Cram, D. S., Song, B., Magli, M. C., Gianaroli, L., Lacham-Kaplan, O., Findlay, J. K., Jenkin, G., and Trounson, A. O. (2008) Hum. Reprod. 23, 1138–1144
20. Klatsky, P. C., Carson, S. A., and Wessel, G. M. (2010) Mol. Reprod. Dev. 77, 386
21. Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010) Bioinformatics 26, 139–140
22. Robinson, M. D., and Oshlack, A. (2010) Genome Biol. 11, R25
23. Antosh, M., Fox, D., Helfand, S. L., Cooper, L. N., and Neretti, N. (2011) Aging 3, 576–583
24. Edgar, R., Domrachev, M., and Lash, A. E. (2002) Nucleic Acids Res 30, 207–210