Original Article

# Primary differentiation in the human blastocyst: Comparative molecular portraits of inner cell mass and trophectoderm cells.

James Adjaye[1*], John Huntriss[2], Ralf Herwig[1], Alia BenKahla[1], Thore C. Brink[1], Christoph Wierling[1], Claus Hultschig[1], Detlef Groth[1], Marie-Laure Yaspo[1], Helen M. Picton[2], Roger G. Gosden[3], Hans Lehrach[1]

[1] Max-Planck Institute for Molecular Genetics (Department of Vertebrate Genomics), Ihnestrasse 73, D-14195 Berlin, Germany.

[2] Reproduction and Early Development Research Group, Department of Obstetrics and Gynaecology, University of Leeds, Belmont Grove, Leeds, W.Yorkshire, LS2 9NS, UK.

[3] Center for Reproductive Medicine & Infertility, Weill Medical College, Cornell University, 505 East 70th Street, New York, NY 10021.

* To whom correspondence should be addressed

Email. adjaye@molgen.mpg.de

Tel. 0049-30-8413-1216

Fax. 0049-30-8413-1128

1

**KEY WORDS**

Preimplantation development, Blastocyst, Inner Cell Mass, Trophectoderm, Pluripotency, Embryonic stem cells, Trophoblastic stem cells, *OCT4*, *CDX2*, Differentiation, Microarrays.

**ABSTRACT**

The primary differentiation event during mammalian development occurs at the blastocyst stage and leads to the delineation of the inner cell mass and the trophectoderm. We provide the first global mRNA expression data from immunosurgically dissected inner cell mass (ICM) cells, trophectoderm (TE) cells and intact human blastocysts. Using a cDNA microarray composed of 15,529 cDNAs from known and novel genes, we identify marker transcripts specific to the ICM (e.g. *OCT4*/*POU5F1*, *NANOG*, *HMGB1*and *DPPA5*) and TE (e.g. *CDX2, ATP1B3, SFN* and *IPL*), in addition to novel ICM - and TE-specific ESTs. The expression patterns suggest that the emergence of pluripotent ICM and TE cell lineages from the morula is controlled by metabolic and signalling pathways which include, *inter alia*, WNT, MAPK, TGF-beta, NOTCH, integrin-mediated cell adhesion, Phosphatidylinositol 3-Kinase and apoptosis. These data enhance our understanding of the first step in human cellular differentiation and, hence, the derivation of both embryonic stem (ES) cells and trophoblastic stem (TS) cells from these lineages.

**INTRODUCTION**

The earliest stages of human cellular differentiation occur during the transition from the morula to blastocyst stage of preimplantation development, and give rise to divergent lineages. These are: 1) the trophectoderm (TE), which gives rise to the cytotrophoblast and syncytiotrophoblast of the placenta and 2) the inner cell mass (ICM), which generates the embryo proper and extraembryonic tissues [1]. The immediate outcomes of blastocyst morphogenesis, namely epithelial differentiation and segregation of the ICM, are mainly regulated autonomously and they involve temporally regulated gene expression, cell polarisation and cell–cell interactions leading to the generation of differentiated progeny. The ICM and TE are the sources of embryonic stem (ES) cells and trophoblastic stem (TS) cells, respectively. ES cells are pluripotent, which implies that they are capable of differentiating into cells representing the three primary germ layers - endoderm, ectoderm and mesoderm [2]. In contrast, TS cells have a more restricted developmental potential and are only capable of forming trophectodermal cell lineage in vivo or giant cells in vitro as demonstrated in the mouse [3].

A number of "stemness" genes have been identified in murine [4-7] and human [8-12] embryonic stem cell lines under culture conditions. However, the weak overlap between these genes (~25%) may be a consequence of the heterogeneous genetic backgrounds of the various cell lines and/ or epigenetic and chromosomal instability following differentiation in vitro [13,14]. Hence, genes for "stemness" may be better identified 'at source' within the pluripotent inner cell mass of the mammalian blastocyst.

Whilst specific markers for the TE and ICM, such as *CDX2* and *OCT4* [15-17], have been identified, these are limited and under-explored. Here, we provide the first global analysis of differential gene expression in intact human blastocysts and immunosurgically isolated ICM and TE cells for identifying novel marker genes specific to these lineages. We have uncovered gene expression patterns which suggest that the delineation of the ICM and TE from the morula may be controlled by signalling pathways which are also crucial for determining cell fate later in

3

embryonic development, in the ontogenesis of some cancers and, crucial for advances in technology, the maintenance of human ES and TS cells in vitro.

## MATERIALS AND METHODS

**Clone selection and microarray fabrication.**

All transcripts in the human ENSEMBL catalogue v.3.26.1 were blasted against the 5-prime ends of previously sequenced clones from the Human UNIGENE set RZPD2.1. Sequences sharing a best hit with any of these transcripts were assigned to the corresponding ENSEMBL gene. All matches had to fulfil two criteria: >96% identity over more than 50 nucleotides, which corresponds to a score of 100 and an e-value of 1e-20. To avoid redundancy, a single clone was selected for each ENSEMBL gene. For the generation of microarray slides, cDNA inserts from a collection of 15,529 annotated human cDNAs were amplified, purified, spotted and processed as previously described [18].

**Preimplantation embryos, generation of T7-linked double-stranded cDNA and amplified RNA**.

Preimplantation embryos surplus to requirement for IVF treatment were donated by patients for research attending the Jones Institute at Eastern Virginia Medical School (EVMS) and at Leeds General Infirmary (LGI). The IVF-derived blastocysts were similar to those used for deriving ES cells, i.e. fully expanded with a well defined ICM and TE, lacking signs of degeneration and at stage 3 according to the grading system described [19].

All samples were obtained after informed consent under protocols approved by the Institutional Review Board at EVMS and Research Ethics Committee of the LGI and licensed by the Human Fertilisation and Embryology Authority (UK). Permission was also granted by the ethics commission of the Free University Berlin, allowing the use of the generated RNA samples in Germany. Samples were washed in PBS, and lysed and stored at -80°C in 50 µl Dynal lysis buffer (Dynabeads, Dynal, UK) supplemented with 10% RNA Later (Ambion Inc., Austin, TX).

cDNA samples were derived from duplicate intact blastocysts, the trophectoderm (TE) and the inner cell mass (ICM) of the human blastocyst were generated as follows. Ten pronucleate

embryos were cryopreserved and, after thawing, cultured for up to six days in sequential medium (Irvine Scientific, Santa Anna, CA). At the blastocyst stage, the embryos underwent immunosurgery to separate the TE and ICM [20]. Unhatched blastocysts were exposed briefly to acidified Tyrode's medium (Irvine Scientific, Santa Ana, CA) for removing the zona pellucida. They were exposed to a rabbit antiserum raised against the BeWo trophoblastic cell line (Atlantic Antibodies, Windham, ME) and subsequently incubated with guinea pig complement (Gibco BRL, Grand Island, NY). Degenerated TE cells were separated from the ICM by pipetting through a finely-drawn Pasteur pipette. The isolated ICM cells were washed three times in fresh medium and checked microscopically for the absence of adhering TE cells before being added to a tube containing 50 µl of RNA Later. The lysed TE cells, together with the microdroplet of medium, were also collected, washed and transferred to separate tubes for analysis. Samples were stored for one month at -80$^{\circ}$C before shipping to the UK.

**cDNA amplification and in vitro transcription**.

T7 promoter-linked double-stranded cDNA samples derived from ICM, TE and intact blastocysts were generated according to published protocols [21]. Briefly, mRNA was extracted from thawed lysed cells using Oligo-dT magnetic beads (Dynabeads, Dynal, UK). cDNA was generated employing T7 promoter-linked oligo-dT primers for the reverse transcription step and whole-transcriptome amplification was executed using a modified SMART amplification protocol (BD Biosciences, San Jose, CA). Amplified RNA was generated using the MegaScript T7 High Yield Transcription kit (Ambion). RNA purity, integrity and concentrations were evaluated on the Agilent 2100 Bioanalyzer.

**Direct labelling of RNA and hybridisations**.

MIAME (Minimum Information About Microarray Experiments) guidelines were adhered to in our experimental design [22].

Four independent labelling (including dye-swaps, i.e Cy3 and Cy5) reactions per aRNA sample (ICM, TE and intact blastocyst) were carried out using 3 µg aRNA per reaction. Direct incorporation of Cy3 and Cy5 during reverse transcription was carried out in a 20 µl reaction volume using 1 µg of a random hexamer primer. Purification of labelled cDNAs, hybridisations, slide washes and capturing of fluorescence images described in detail in supplementary materials and methods which is published online as supporting information.

**Data analysis**.

Data normalization was carried out as described previously [23]. Gene expression was judged according to the comparison to a negative control sample (full details in supplementary materials and methods which is published online as supporting information).

For each cDNA, we performed statistical tests based on the replicate signals in experiments with ICM and TE samples. This was done for the two blastocysts separately. Three standard tests were used in parallel, Student's T-test, the Welch test and Wilcoxon's rank-sum test [24]. To evaluate differential expression of the genes, P-values of Wilcoxon's rank-sum test were preferred as a reference since this test does not depend for its validity on a specific distribution (e.g. Gaussian). A recursive function was implemented for calculating the exact P-values.

**Pathway analysis**.

Array data were used to test whether specific pathways showed differential expression. Pathways were taken from the KEGG database. Consider for each pathway $i$, the set of related genes $(x_{i1}, y_{i1}),...,(x_{in_i}, y_{in_i})$. Here, $x_{ij}, y_{ij}$ denote the expression level of the $j$-th gene in TE and ICM, respectively. Wilcoxon's matched pairs signed rank test were used to calculate a Z-score for the

differences $d_{ij} = x_{ij} - y_{ij}$ for each pathway $i$. These differences were ranked, and the ranks of differences with negative signs, $R_{neg}$, and those with positive signs, $R_{pos}$, were summed. The test statistic is the smaller of the two numbers, $R = \min\{R_{pos}, R_{neg}\}$. If the pathway is not affected by the treatments $R_{neg}$ and $R_{pos}$ will be fairly equal, but, if there is a trend of under- or over-expression, the test statistic will be small. The Z-score is defined as $z = \dfrac{|R - E(R)|}{\sqrt{Var(R)}}$ where E is the expectation and Var is the variance of R. These were calculated as $E(R) = \dfrac{n_i(n_i + 1)}{4}$ and

$$Var(R) = \dfrac{n_i(n_i + 1)(2n_i + 1)}{24}$$ respectively [25].

**Real-Time RT-PCR analysis.**

This was carried out for a set of genes on the ABI PRISM 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA). Full details of the experimental set up and list of primers, annealing temperatures, and genes under investigation are shown in Supplementary Table 10, and supplementary materials and methods which are published online as supporting information.).

**Functional annotation of expressed genes.**

We used the GO-terminology related to Ensembl version 3.0 to characterise the following subsets of our expression data:

All) All genes on the array.

TE) All genes highly expressed in the TE ($> 0.9$)

ICM) All genes highly expressed in the ICM ($> 0.9$)

BL) All genes highly expressed in the blastocyst ($> 0.9$)

The GO vocabulary was taken from the Gene Ontology website in May 2004 (http://www.geneontology.org). This was imported into the sqlite-database (http://www.sqlite.org). Data analysis was carried out using the R-Statistics software

(http://www.r-project.org). We have developed a web-application on our Goblet webserver (http://goblet.molgen.mpg.de/blastocyst), which allows users to view genes based on their expression pattern or on their Gene Ontology annotation.

**RESULTS**

**Validation of messenger RNA isolation and amplification.** Using a modification of previously described protocols [21], we have generated microgram amounts ($10^6$-fold amplification) of aRNA (amplified RNA) from isolated mRNAs from duplicated sets of ICM, TE and intact blastocysts from separate individuals. Similar protocols have been used to generate microgram amounts of aRNA needed for expression profiling [26-28]. To verify the success of immunosurgery, mRNA isolation and subsequent in vitro transcription to generate enough aRNA for hybridisations, gene-specific RT-PCR was performed for previously established markers of both ICM and TE (Fig.1 A, B, C). The transcription factors, *OCT4, NANOG, REX1 /ZFP42* and *SOX2,* and the gene encoding the signal transduction adaptor protein, *DAB2,* are over-expressed in the ICM. During mouse preimplantation development *Oct4*, *Nanog* and *Dab2* are required *in vivo* and *in vitro* for the establishment and maintenance of the pluripotent ICM in blastocysts and in cultured ES cells [16,28-31]. All the ICM enriched transcripts shown in Fig.1C have been designated as stemness genes [2-10]. Trophoblastic determining genes, *HCG/CGB5*, *KRT18, HAND1, PSG3, CDX2* and *TBX1* [33], were enriched in the TE samples. Among these genes, the transcription factor, *Cdx2*, has been shown to be instructive for TE differentiation during mouse preimplantation development [34]. Having satisfactorily established that the pools of the TE and ICM aRNA samples do indeed reflect the transcriptomes of these cells, we proceeded to carry out whole genome expression analysis. Transcription profiles were generated by employing a cDNA microarray (Ensembl Chip) consisting of 15,529 re-sequenced and annotated clones.

The reproducibility of amplification and subsequent hybridisations between replicates were assessed by calculating Pearson correlation coefficients (Supplementary Fig. 1, which is published online as supporting information) The values indicate a high degree of reproducibility in mRNA isolation, amplification and array hybridisation.

**Expression profiling distinguishes ICM from TE**

ICM and TE cells were isolated from two blastocysts from different individuals. Four independent hybridisation experiments were performed for each biological replicate, including Cy-dye swaps. Additionally, RNA from two blastocysts were pooled to generate a reference sample. An overview of comparative gene expression in ICM and TE cells is shown in Fig. 2A and reveals as expected a high overall correlation between the data (0.90), which helps to validate them. To judge whether a given gene is expressed in these cell types, we compared its signal against a negative control sample and computed a numerical value to judge gene expression (BG-tag, see methods). This number reflects the proportion of background noise in relation to the actual signal [23]. Typically, a BG-tag of 0.9 indicates a detectable signal for the probe (Fig. 2B). Using this criterion, we found that 7,481 (48%) of genes represented on the chip (probes) were detected in one of the three cell types (ICM, TE, blastocyst (BL)). As demonstrated in Fig. 2B, most of these genes are either specific to the intact blastocyst (2,880) or common to all three cell types (2,031). The number of genes that are expressed exclusively in the immunosurgically isolated ICM or TE is rather low (292 and 345, respectively).

Although these global statistics might point to trends in over- and under-expression of genes, it should be noted that, globally, the differences between ICM and TE are not large enough to account for the variance in biological replicates. For example, if we apply standard statistical procedures, such as principal component analysis (PCA) and hierarchical clustering for grouping the biological replicates using a large unfiltered set of genes, we observe that the clustering result reflects technical reproducibility rather than biological characteristics (Fig. 2C). To specifically address biological significance, the two blastocysts were analysed separately. This was accomplished by comparing the TE and ICM replicates pertinent to each blastocyst using statistical tests for differential expression (see materials and methods). Three distinct tests in parallel  (Student's t-test, Welch test, Wilcoxon's rank-sum test) were employed to help overcome individual bias [24]. By adopting this approach, a subset of 78 candidate marker genes were identified, consisting of genes which are differentially expressed in ICM and TE (within one of

the blastocysts) at the 0.05-level of significance. Of these 78 genes, 23 (29%) were completely novel, and 17 were common to both blastocysts. By repeating the PCA and clustering analysis using these genes, we observed a clear biological separation of ICM and TE across the biological replicates (Fig. 2D).

These newly identified genes may complement existing markers for ICM and TE, such as *OCT4* and *CDX2* (Figs. 1B and Figs. 3A,B) and represent diverse biological functions. For example, the TE marker, *SFN* (stratifin), is an epithelial cell antigen, which is exclusively expressed in keratinocytes. Its role in cell proliferation and apoptosis suggests that the protein could be relevant to the regulation of growth and differentiation of multiple cell types through the protein kinase C signalling pathway [35]. Among the putative ICM markers, *HMGB1* and *GLTSCR2* have been identified as potential stemness genes based on expression studies in the human ES lines, HSF-1, HSF-6 and H9 [10]. *HMGB1* is a member of the high mobility group of transcription factor encoding proteins that act primarily as architectural facilitators in the assembly of nucleoprotein complexes, as in the initiation of transcription of target genes. Murine Hmgb1 has been shown to be a co-activator of Oct4 [36]. *GLTSCR2* is a gene of unknown function residing in the glioma tumour suppressor region of chromosome 19q. Of the proposed list of stemness genes [4,5,12] only *ITGA6* (integrin alpha 6 chain) appeared in all the stem cell lines analysed, but by comparing this list with our ICM markers we have also identified *ITGA5* as another candidate for stemness. These findings highlight the potential importance of integrins in cell adhesion as well as in cell-surface mediated signalling for establishing and maintaining pluripotency.

The expression levels obtained from microarrays were verified independently for selected markers using Real-Time PCR (Fig. 3B). For example, *ATP1B3,* a $Na^+$/$K^+$-ATPase, is over-expressed approximately 3-fold in TE, thus reflecting the roles played by these ATPases in driving transepithelial $Na^+$ and fluid transport for blastocoel formation [37]. Another ICM marker, *HMGB1,* is also over-expressed to the same degree. Unexpectedly, the ribosomal proteins, *RPL14, RPL7A, RPL19* and *RPL32,* were identified as ICM-specific markers, corroborating

previous findings implying that some large subunits of ribosomal proteins are stem cell markers [2-10] and that these proteins might bind and inhibit the translation of specific mRNAs.

The full sets of data for global gene expression in the duplicate ICMs, TEs and pooled blastocysts are presented (Supplementary Table 1., which is published online as supporting information). In general, the magnitude of expression recorded after RT-PCR was greater than from microarrays, an observation which is consistent with previous findings [4,26].

## Functional annotation of expressed genes

The Gene Ontology vocabulary provides a unified terminology for the description of genes and their products [38]. It is divided into three main categories, "Molecular Function", "Biological Process" and "Cellular Component". A comparative analysis of these three categories at a global level within the ICM, TE and intact blastocyst, did not reveal a bias towards a particular category in these cell types (data not shown). In contrast, by repeating this analysis on "Molecular Function" using the 78 marker genes of which 51 have GO annotations we observed a slight bias towards specific molecular functions within ICM and TE cells (Fig. 4). For example, the ribosomal proteins, *RPL14, RPL7A, RPL19* and *RPL32,* under structural molecular activity (GO:0005198) are all expressed in the ICM (Fig. 4A). A more detailed description of these markers with respect to their chromosomal localisation and ontology is presented (Supplementary Table 2, which is published online). For a global overview, we combined the expression data from the ICM, TE and intact blastocysts to create a database for searching for expression levels and related Gene Ontologies (http://goblet.molgen.mpg.de/blastocyst).

## Developmentally conserved signalling pathways.

During embryogenesis the specification and proper arrangement of new cell types require the coordinated regulation of gene expression and precise interactions between adjacent cells. These morphogenetic changes depend on the interaction of extracellular ligands with their receptors.

Delineation of signalling pathways will be fundamental for understanding the mechanisms regulating pluripotency and self-renewal in cultured ES cell lines. We searched the ICM and TE data for components of these pathways by assigning P-values using Wilcoxon's matched pair signed rank test, as described in Methods. This strategy is distinct from the commonly employed strategy for identifying differentially expressed genes using repeated measurements with a two-sample location test, such as Student's t test or Wilcoxon's rank-sum test as we directly involve groups of genes associated with particular pathways instead of conventional gene-wise analysis. The data indicate the involvement of WNT, MAPK, TGF-ß /BMP, NOTCH, integrin-mediated cell adhesion, apoptosis signalling pathways, and metabolic processes such as glycolysis, sterol biosynthesis, androgen and estrogen metabolism. The full list of signalling and metabolic pathways identified by these analyses is in Table 1, and Supplementary Table 3, which is published online as supporting information. Pathway annotations were adopted from the KEGG (Kyoto Encyclopedia of Genes and Genomes; http://www.genome.jp/kegg) database.

Such explorative approaches have also been employed using expressed sequence tag and array data generated from undifferentiated and differentiated human ES cells [8, 39] and array data relating to mouse preimplantation development [27, 28].

*Apoptosis in the mammalian blastocyst.* Regulation of cell population size and lineage determination is mediated by cell cycle control, differentiation and programmed cell death (PCD) or apoptosis. The latter is characterized by chromatin condensation, nuclear membrane blebbing, and fragmentation in the cytoplasm and nucleus [40].

Apoptosis is evident at the blastocyst stage, if not earlier. It is mainly restricted to the ICM to regulate the size of the cell mass and perhaps to eliminate cells retaining the potential to form TE ectopically [41]. A list of expressed genes involved in the apoptosis signalling pathway is shown (Supplementary Table 4, which is published online as supporting information).

*NOTCH signalling in the blastocyst.* Cell-cell signalling mediated by the Notch receptor determines cell fate and regulates pattern formation in many phyla. The Notch signalling pathway is operational in both the ICM and TE and is known to function in the mouse blastocyst [27]. We observed expression of several ligands and receptors in this pathway (Supplementary Table 5, which is published online as supporting information) consistent with a mouse model in which Wnt and Notch act sequentially to set up the initial asymmetry in the zygote and influencing polarity of the developing embryo [27].

*Transforming growth factor-*β *(TGF-*β*) signalling.* The TGF-β family consists of multi-functional growth and differentiation factors regulating many cellular processes through complex signal-transduction pathways. The family members include TGF-β isoforms, activins and bone morphogenic proteins (BMPs). Expression of the signalling type I and type II receptors for TGF-β in mouse and human fertilised oocytes and blastocysts suggested a role for TGF-β in early preimplantation development, potentially in the outgrowth of parietal endoderm [42]. Differential expression of TGF-ß isoforms, activins, *BMP*s, *MADH*s/*SMAD*s was also evident. In particular *BMP4*, previously shown to induce the differentiation to trophoblast when over expressed in human ES cells [43] is 2.28 fold enriched in the TE. Other components of the TGF-beta signalling cascade are shown (Supplementary Table 6, which is published online as supporting information).

*Integrin and cadherin-mediated cell adhesion.* The ICM and TE originate from the division of polar blastomeres when their cleavage furrows parallel their apical surfaces. These blastomeres polarize in response to asymmetric cell-cell contact. Pathway analysis identified signalling pathways related to integrin-mediated cell adhesion. In addition, several $Na^+/K^+$-ATPases (e.g. *ATP1B3*- Fig. 3B) were over-expressed in TE, reflecting their presumptive roles in driving fluid transport across this epithelium. Other cell adhesion-related genes were also detected, as expected where intercellular junctions are important for controlling blastocyst permeability [44]. However, there was over-expression in TE of a subset of these genes, e.g., Desmocollin 2 (*DSC2* x1.55),

Protocadherins (*PCDH7* x1.67, *PCDH11* x1.57, *PCDHB7* x1.62), E-cadherins (*CDH19* x1.9, *CDH24* x1.54, *CDH22* x1.82), tight junction proteins (*TJP1* x1.4, *TJP2* x1.8), Claudins (*CLDN2* x1.4, *CLDN16* x1.79, *CLDN10* x2.25) and seven-pass transmembrane receptor of the cadherin superfamily (*CELSR2* x1.46). For the tight junction constituents, *OCLN* (occludins), *JAM-2*, (Junction adhesion molecule 2) and *CGN* (Cingulin) a lack of over-expression in the TE may be due to the fact that translational rather than transcriptional control is operative due to cell contact symmetry [44]. A comprehensive listing of these genes and their expression ratios is given (Supplementary Tables 1 and 7, which are published online as supporting information).

*WNT signalling in the blastocyst.* The *WNT* gene family consists of numerous conserved glycoproteins that regulate pattern formation during embryogenesis in a wide variety of tissues, including the nervous system. It has recently been shown that activation of the canonical WNT/Wnt pathway is sufficient to maintain self-renewal of both human and mouse ES cells [45], and also that this pathway is operative during human and mouse preimplantation development [21, 27]. We detected differential expression of transcripts encoding WNT ligands (*WNT*s), receptors of the *Frizzled* gene family (*FZD*), *Frizzled*-related protein family (*SFRP*) and intracellular signal transducers and modifiers (*DVL1, AXIN*). The genes encoding Casein kinase 1 alpha (*CSNK1A*), dishevelled activator of morphogenesis 1 (*DAAM1*), which are agonists of the WNT pathway are both over-expressed in the TE (Supplementary Table 8, which is published online as supporting information). These agonists were up regulated in differentiated ES cells [39]. In addition, we find that *GSK-3B* (glycogen synthase 3 kinase) expression is down regulated in the ICM, thus corroborating the reported inactivation of *GSK-3B*, which leads to the activation of the WNT pathway in maintaining the undifferentiated state of ES cells [45].

*MAPK signalling pathway.* The mitogen-activated protein kinase (MAPK) pathways regulate cell growth, differentiation, proliferation and death. All members of the MAPK pathway have been shown to be expressed during mouse preimplantation development [46]. Differential expression

of the various genes involved is shown (Supplementary Table 9, which is published online as supporting information).

**Epigenetic regulation of lineage-specific gene expression**.

De novo methylation of DNA by cytosine-5-methyltransferases is a well-characterised mechanism of epigenetic transcriptional control and it has been shown that this mode of transcriptional control may contribute to the differentiation of the ICM and TE at the blastocyst stage [47]. Dnmt3b protein is specifically localised in the ICM of mouse blastocysts [48]. In addition, expression of *DNMT3B* and *DNMT3A* is enriched in undifferentiated human embryonic stem cells [8, 9, 10, 49] as well as in the ICM cell lineage (Fig.1C). This expression pattern suggests an important role in ICM-specific methylation in the blastocyst [47]. In contrast, transcripts of *DNMT3L* were expressed in both ICM and TE (Fig.1C). We also detected differential expression of several methyltransferases (Supplementary Table 1, which is published online as supporting information).

Other epigenetic regulators of X-inactivation, imprinting, maintenance of pluripotency and the establishment of the trophectoderm lineages including *EZH2* (enhancer of zeste homolog 2), *EED* (embryonic ectoderm development) and *CTCF* (CCCTC-binding zinc finger protein) are expressed at high levels in the blastocyst and all ICM and TE samples [50-52].

**Imprinting**

Several imprinted genes (*H19, GRB10, SNURF, MEST, NAP1, UBE3A, DLX5, MAGEL2, OSBPL5/OBPH1* and *ATP10A*) were expressed at medium to high levels (BG-tag 70% to >90%) in the ICM, TE and blastocyst. Our strict criteria for determining statistically significant differential expression between ICM and TE based upon the microarray data may occasionally obscure more subtle differential expression patterns that are revealed when assayed by alternative methods. Real-time PCR identified a clear TE-biased (30-fold higher) expression for the imprinted gene, *IPL* (imprinted in placenta and liver) (Fig. 3B). Significantly, *IPL/PHLDA2* is a

marker of human cytotrophoblast and in the mouse *Ipl* restricts placental growth [53]. Thus, such imprinted genes that can act as regulators of nutrient supply at the feto-maternal interface may also influence growth and development of the early embryo.

**DISCUSSION**

This is the first report of gene expression in the inner cell mass compared with trophectodermal cells from human blastocysts obtained after immunosurgery. To obtain molecular profiles at the primary stage of differentiation during human embryogenesis has required extensive internal and external validation because only minute amounts of RNA from a few score cells were available in each specimen. Considering the potential sources of biological and technical variation, the results were surprisingly consistent and the immunosurgery had evidently not caused significant degradation of RNAs from either compartment of the blastocyst. The results were also compared with global reference data obtained from intact blastocysts. There were minor differences in expression profiles between intact blastocysts and the ICM or TE cells which might be accounted for by loss of primitive endoderm or even blastocoelic fluid, although studies in the mouse have indicated that total loss of endoderm is unlikely [54]. Major factors that could have contributed to the variabilities in the number of genes detected as expressed in the intact blastocysts and isolated ICM and TE cells include the cryopreservation and immunosurgical procedures adopted in this study. Although these procedures have been shown to affect the normal pattern of gene expression in preimplantation embryos [55, 56], we must stress that both studies were based on RT-PCR analysis on single genes and therefore not conclusive. Additionally, although blastocysts have high implantation potential there may well be considerable variability as a result of slight differences in maturity and competence [49].

The molecular profiles were consistent with expectations based on experimental mammalian embryology and were predictive of cellular physiology. For example, transcripts representing integrin- and cadherin-mediated cell adhesion, MAPK and other gene products involved in the cell cycle and in apoptosis were identified. These components were anticipated in

view of the rapid growth and dynamic morphogenesis of embryos at this transitional stage. Likewise, molecules representing pathways for glucose and sterol biosynthesis were also unsurprising. Absentees were presumably the less abundant transcripts from these pathways.

The most interesting, and potentially significant, findings were among the representatives of key signalling pathways. Members of the WNT, TGF-ß/BMP, NOTCH and phosphotidylinositol 3-kinase pathways were found. These genes are likely to be finely regulated to determine cell fate, adhesion and migration because, when expression goes awry, stem cells may undergo malignant transformation [57]. The expression of *OCT4* and *NANOG* are already established as markers of pluripotency, but there were other genes that were expressed in a similar temporal and spatial pattern. It will be important to investigate whether these genes are candidate markers of the ICM and their role in maintaining an undifferentiated state.

*Perspectives*. These data provide a reference for both normal development of the blastocyst in vivo and in vitro and for derivatives of the ICM and TE. Many human blastocysts generated in vitro fail to implant after transfer to a receptive uterus, and not all ICMs give rise to competent embryonic stem (ES) cell and trophoblastic stem cell lineages in culture. Markers of pluripotent stem cells are valuable tools for predicting cellular phenotype and are needed for advancing ES cell technology. Identification of other molecules that have roles in intracellular signalling is critical for the control of cellular differentiation and fate. This knowledge is critical for the key goal of directing the differentiation of stem cells or maintaining them as proliferating populations of pluripotent precursors, and preferably without feeder cell layers [45]. Some genes that are regulated epigenetically evidently have an unstable imprinting status and research to safeguard the health of children born after assisted reproductive technologies must carefully assess the impact of culture conditions on the expression of these genes in preimplantation embryos [58]. Thus, there are many reasons why the molecular profiles of the ICM and TE cells in the human blastocyst are valuable, and, in the long-term, they carry implications for fertility, reproductive health and the prospects of stem cell-based technologies for treating human diseases.

**Abbreviations**

ICM, inner cell mass; TE, trophectoderm; BL, blastocyst; BG-tag, background tag.

**References**

1. Gardner, RL, & Beddington, RS. Multi-lineage stem cells in the mammalian embryo. J. Cell Sci. Suppl. 1988; 10: 11-27.

2. Thomson, JA, Itskovitz-Eldor, J, Shapiro, SS, et al. Embryonic stem cell lines derived from human blastocysts. Science 1998; 282:1145-1147.

3. Yan, J, Tanaka, S, Oda, M, et al. Retinoic acid promotes differentiation of trophoblast stem cells to a giant cell fate. Dev. Biol. 2001; 235:422-432.

4. Ramalho-Santos, M, Yoon, S, Matsuzaki, Y, et al. "Stemness" transcriptional profiling of embryonic and adult stem cells. Science 2002; 298:597-600.

5. Ivanova, NB, Dimos, JT, Schaniel, C, et al. A stem cell molecular signature. Science 2002; 298: 601-604.

6. Fortunel, NO, Out, HH, Ng, HH, et al. Comments on "'Stemness': transcriptional profiling of embryonic and adult stem cells" and "A stem cell molecular signature". Science 2003; 302:393.

7. Sharov, AA, Piao, Y, Matoba, R, et al. Transcriptome analysis of mouse stem cells and early embryos. PloS Biol. 2003; 1:41-419 (E74).

8. Sato, N, Sanjuan, IM, Heke, M, et al. Molecular signature of human embryonic stem cells and its comparison with the mouse. Dev. Biol. 2003; 260:404-413.

9. Sperger, JM, Chen, X, Draper, JS, et al. Gene expression patterns in human embryonic stem cells and pluripotent germ cell tumors. Proc. Natl. Acad. Sci. 2003; 100:13350- 13355.

10. Richard, M, Tan, S-P, Tan, J-H, et al. The transcriptome profile of human embryonic stem cells as defined by SAGE. Stem Cells 2004; 22:51-64.

11. Bhattacharya, B, Miura, T, Brandenberger, R, et al. Gene expression in human embryonic stem cell lines: Unique molecular signature. Blood 2004; 103:2956-2964.

12. Abeyta, M.J, Clark, AT, Rodriguez, RT, et al. Unique gene expression signatures of independently-derived human embryonic stem cell lines. Hum. Mol.Genet. 2004; 13:601-608.

13. Humpherys, D, Eggan, K, Akutsu, H, et al. Epigenetic instability in ES cells and cloned mice. Science 2001; 293:95-97.

14. Draper, JS, Smith, K, Gokhale, P, et al. Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. Nat. Biotechnol. 2004; 22:53-54.

15. Tanaka, TS, Kunath, T, Kimber, WL, et al. Gene expression profiling of embryo-derived stem cells reveals candidate genes associated with pluripotency and lineage specificity. Genome Res. 2002; 12:1921-1928.

16. Chambers, I, Colby, D, Robertson, M, et al. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell 2003; 113:643-655.

17. Hansis, C, Grifo JA & Krey LC. Oct-4 expression in inner cell mass and trophectoderm of human blastocysts. Mol. Hum. Reprod. 2000; 6:999-1004.

18. Adjaye, J, Herwig, R, Herrmann, D, et al. Cross-species hybridisations of human and bovine orthologous genes on high density cDNA microarrays. BMC Genomics 2004; 5: 83

19. Veeck, LL, & Zaninovic N. An atlas of Human Blastocysts. Parthenon Press, New York and London (2003).

20. Lanzendorf, SE, Boyd, CA, Wright, DL, et al. Use of human gametes obtained from anonymous donors for the production of human embryonic stem cell lines. Fertil. Steril. 2001; 76: 132-137.

21. Adjaye, J, Bolton, V & Monk, M. Developmental expression of specific genes detected in high-quality cDNA libraries from single human preimplantation embryos. Gene 1999; 237:373-383.

22. Brazma, A, Hingamp, P, Quackenbush, J, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat. Genet. 2001; 29:365-371.

23. Kahlem, P, Sultan, M, Herwig, R et al. Transcript level alterations reflect gene dosage effects across several tissues in a mouse model of Down Syndrome. Genome Res. 2004; 14:1258-1267.

24. Herwig, R, Aanstad, P, Clark, M et al. Statistical evaluation of differential expression on cDNA nylon arrays with replicated experiments. Nucleic Acids Res. 2001; 29: e117.

25. Lehmann, E. In Nonparametrics: Statistical Methods based on Ranks. Holden-Day Inc., San Francisco. 1975

26. Dobson, AT, Raja. R, Abeyta, MJ, et al. The unique transcriptome through day 3 of human preimplantation development. Hum Mol Genet. 2004; 3:1461-1470.

27. Wang, QT, Piotrowska, K, Ciemerych, MA, et al. A genome-wide study of gene activity reveals developmental signaling pathways in the preimplantation mouse embryo. Dev. Cell 2004; 6:133-144.

28. Hamatani, T, Carter, MG, Sharov, AA et al. Dynamics of global gene expression changes during mouse preimplantation development. Dev. Cell 2004; 6:117-131.

29. Niwa, H, Miyazaki, J & Smith, AG. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat. Genet. 2000; 24:372-376.

30. Nichols, J, Zevnik, B, Anastassiadis, K, et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell 1998; 95: 379-391.

31. Yang, DH, Smith, ER, Roland, IH, et al. Disabled-2 is essential for endoderm cell position and structure formation during mouse embryogenesis. Dev. Biol. 2002; 251: 27-44.

32. Mitsui, K, Tokuzawa, Y, Itoh, H, et al. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell 2003; 113: 631- 642.

33. Kunath, T, Strumpf, D & Rossant, J. Early trophoblast determination and stem cell maintenance in the mouse. Placenta 2004; 25:S32-S38.

34. Chawengsaksophak, K, de Graaff, W, Rossant, J et al. Cdx2 is essential for axial elongation in mouse development. Proc. Natl Acad Sci U S A. 2004; 101:7641-7645.

35. Leffers, H, Madsen, P, Rasmussen, HH, et al (1993). Molecular cloning and expression of the transformation sensitive epithelial marker stratifin. A member of a protein family that has been involved in the protein kinase C signalling pathway.  J.  Mol. Biol. 1993; 231:982-998.

36. Butterroni, C, De Felici, M, Schöler, et al. Phage display screening reveals an association between germline-specific transcription factor Oct-4 and muliple cellular proteins. J. Mol. Biol. 2004; 304:529-540.

37. Kidder, GM.  Trophectoderm development and function: the roles of Na+/K(+)-ATPase subunit isoforms. Can. J. Physiol. Pharmacol. 2002; 80:110-115.

38. Gene Ontology Consortium. Tool for the unification of biology. Nat. Genet. 2000; 25:25-29.

39. Brandenberger, R, Wei, H, Zhang, S, et al. Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation. Nat Biotechnol. 2004; 22:707-716.

40. Juriscova, A & Acton, BM. Deadly decisions: the role of genes regulating programmed cell death in human preimplantation embryo development. Reprod. 2004; 128:281-291.

41. Hardy, K, Stark, J & Winston, RML. Maintenance of the inner cell mass in human blastocysts from fragmented embryos. Biol. Reprod. 2003; 68:1165-1169.

42. Roelen, BA & Mummery, CL. Transforming growth factor-betas in pre-gastrulation development of mammals. Mol. Reprod. Dev. 2000; 56:220-226.

43. Zu, R-H, Chen, X, Li, DS, et al. BMP4 initiates human embryonic stem cell differentiation to trophoblast. Nat. Biotechnol. 2002; 20(12): 1261-1264.

44. Ghassemifar, MR, Eckert, JJ, Houghton, FD et al. Gene expression regulating epithelial intercellular junction biogenesis during human blastocyst development in vitro. Mol. Hum. Reprod. 2003; 9:245-252.

45. Sato, N, Meijer, L, Skaltsounis, L, et al. Molecular signature of human embryonic stem cells and its comparison with the mouse. Nat. Med. 2004; 10:55-63.

46. Wang, Y,Wang, F, Sun, T et al. Entire mitogen activated protein kinase (MAPK) pathway is present in preimplantation mouse embryos. Dev Dyn. 2004; 231: 72-87.

47. Santos, F, Hendrich, B, Reik, W, et al. Dynamic reprogramming of DNA methylation in the early mouse embryo. Dev. Biol. 2002; 241:172-182.

48. Watanabe, D, Suetake, I, Tada, T, et al. Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. Mech. Dev. 2002; 118: 187-190.

49. Huntriss, J, Hinkins, M, Oliver, B, et al. Expression of mRNAs for DNA methyltransferases and methyl-CpG- binding proteins in the human female germ line, preimplantation embryos, and embryonic stem cells. Mol. Reprod. Dev. 2004; 67: 323-336.

50. Surani, MA. Reprogramming of genome function through epigenetic inheritance. Nature 2001; 414:122-128.

51. Erhardt, S, Su, IH, Schneider, R et al. Consequences of the depletion of zygotic and embryonic enhancer of zeste 2 during preimplantation mouse development. Development 2003; 130:4235-4248.

52. Fedoriw, AM, Stein, P, Svoboda, P. Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. Science 2004; 303:238-240.

53. Salas, M, John, R, Saxena, A. et al. Placental growth retardation due to loss of imprinting of Phlda2. Mech. Dev. 2004; 121: 119-1210.

54. Gardner, RL. Multi-lineage stem cells in the mammalian embryo. J. Embryol. Exp. Morph. 1985; 88: 303-326.

55. Tachataki, M, Winston, RML, & Taylor DM. Quantitative RT-PCR reveals tuberous sclerosis gene, TSC2, mRNA degradation following cryopreservation in the human preimplantation embryo. Mol. Hum. Reprod. 2003; 9: 593-601.

56. Cauffmann, G, Van de Velde, H, Liebaers, I, & Van Steirteghem, A. DAZL expression in human oocytes, preimplantation embryos and embryonic stem cells. Mol. Hum. Reprod. Doi:10.1093/molehr/gah 167.

57. Sell, S. (2004). Stem cell origin of cancer and differentiation therapy. Crit. Rev. in Oncol and Hematol. 2004; 51: 1-28.

58. Gosden, R, Trasler, J, Lucifero, D et al. Rare congenital disorders, imprinted genes, and assisted reproductive technology. Lancet 2003; 361:1975-1977.

**Figure Legends**

**Figure 1.** Isolation of ICM and TE cells, RNA amplification and expression of known ICM and TE markers. (**A**) Photograph of ICM cells isolated by immunosurgery. (**B**) RNA amplification of cells derived from duplicate sets of ICM, TE and intact blastocysts with RNA size standard shown to the left. (**C**) Confirmation of expression of known ICM/ES and TE markers. *β-ACTIN* and *GAPDH* are used as endogenous controls.

**Figure 2.** Global statistics. (**A**) Regression analysis revealed a global correlation (*r*) between mean TE- (X-axis) and ICM-signals (Y-axis) of 0.90, representing 4,601 probes corresponding to expressed genes, i.e. mean BG>0.9 in TE or ICM. There was no general trend in over- and under-expression. The box shows the regression parameters for the linear regression model $y = 1.01x + 0.19$. The lines show the 2-fold (blue) and 4-fold (magenta) boundaries and the signal range was approximately three orders of magnitude. (**B**) Venn diagram. Probes expressed in the different tissues (TE=trophectoderm, ICM=inner cell mass, BL=pool of two entire blastocysts). Expression was judged by signal detectability using a negative control sample present on each array. The average proportion of negative sample that is expressed below the probe's signal threshold across the replicate experiments indicates the detectability (BG value). An illustration with spots with different BG values shows that the level correlates with visual detectability. A BG level of 0.9 was used to judge expression of genes. (**C**) Global cell-type clustering. Principal component analysis (PCA) on ~600 pre-selected genes revealed no separation between biological replicates but rather between cell types, i.e. TE-1 and ICM-1 in blastocyst #1, and TE-2 and ICM-2 in blastocyst #2 and the pool of the two other blastocysts. The small image shows the same effect with a hierarchical clustering of the cells performed with ~8,000 genes using Pearson correlation as pairwise similarity measure and average-linkage as an update rule. (**D**) Cell-type clustering using ICM- and TE-specific markers. PCA and hierarchical clustering on the set of 78 markers

shows a clear biological separation across the individual blastocysts. Cluster analysis and PCA were generated with the J-Express Pro software (WWW.molmine.com)

**Figure 3.** Validation of selected markers.

(**A**) Hierarchical clustering of the 78 markers shows biological separation of TE and ICM. For each cDNA, the log-ratio (base 2) of the signal intensity and the mean intensity across all five different samples were calculated. Samples were ICM and TE from two blastocysts (ICM-1 and ICM-2, TE-1 and TE-2) and a pool of two intact blastocysts. Clustering shows a clear separation of the biological material. Two large groups are obtainable, ICM over-expressed genes (lower part of the dendrogram) and TE over-expressed genes (upper part).

(**B**) Real-Time PCR confirmation of array-derived expression ratios in duplicate TE and ICM samples derived from blastocysts #1 and #2. Array-derived ratios are denoted as BL1 (yellow) and BL2 (light blue), and Real-Time ratios as BL1 (dark blue) and BL2 (violet). The genes *IPL* and *OCT4* were not on the chip. Since ratios were presented as log-ratio (base 2), values above zero denote TE>ICM expression whereas values under zero indicate ICM>TE expression. The data represent averages from four independent hybridisation experiments and triplicate RT-PCR reactions.

(**C**) Box-plots. Four independent experiments, including a dye swap, were performed for each cell type in both of the blastocysts (TE-1, TE-2, ICM-1, ICM-2). Boxes show the range of the two inner replicates, the whiskers extend to the minimum and the maximum within each sample. The line displays the median value.

**Figure 4.** Functional annotation of ICM and TE marker genes.

(**A**) Distribution of the genes with respect to the Molecular function (GO:0003674) of the gene product. The terms under "obsolete function" are those that have been removed from the active function ontology by the GO curators. Full details of the obsolete lineages can be viewed in Supplementary Table 2 (which is published online as supporting information)
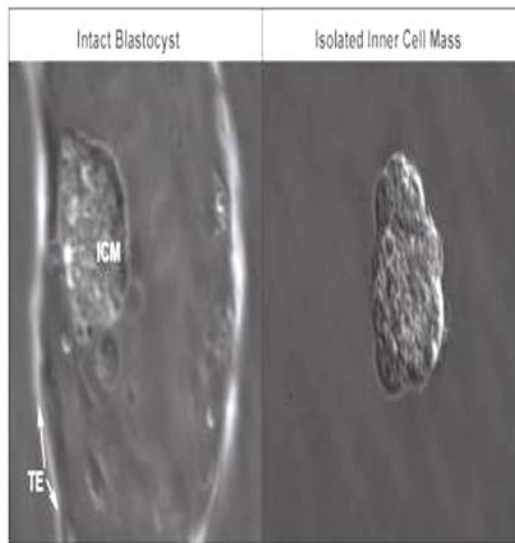
(**B**) Distribution of genes within the lineages of Molecular function (GO:0005488) defined as having binding activity. A high proportion of genes bind to nucleic acids, i.e. transcription factors, chromatin and RNA binding proteins.

(**C**) Distribution of genes within the lineages of Molecular function (GO:0003824) defined as having catalytic activity.
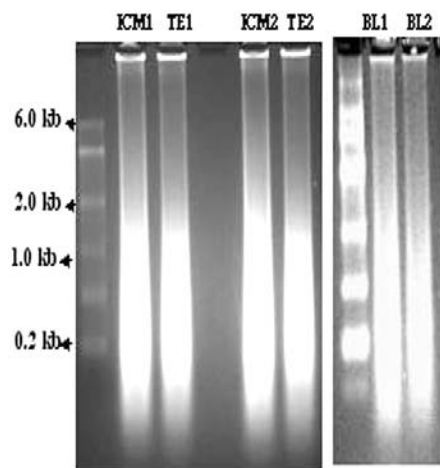
**Table 1.** Analysis of metabolic and signalling pathways operative in the blastocyst

Results of the analysis of highly significant pathways are shown (data from blastocyst 1, significance at the 0.01 level). Pathways were taken from the KEGG database except for the Notch signaling pathway, which lacks an annotation with human data. Gene expression was compared in the TE and ICM array results and the difference was used for computing Wilcoxon's paired signed rank test. Genes that were judged as non-detectable by the BG value criterion were excluded from analysis. Columns: ID = KEGG pathway identifier, PathwayDescription, #Genes = number of genes that were taken into account for computing the statistical test, Z-Score = standard normal approximation of the test statistic, P-Value = P-value of the standard normal distribution for the respective Z-Score, TE -up = number of genes that have higher expression in TE, ICM-up = number of genes that have higher expression in ICM. Tests were computed for all possible pathways, but it should be noted that the normal approximation is only valid for sample sizes >25.
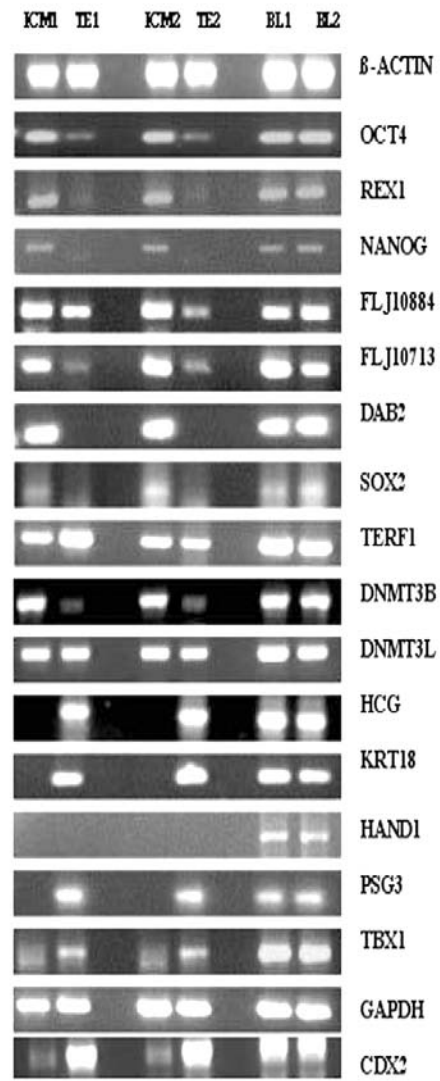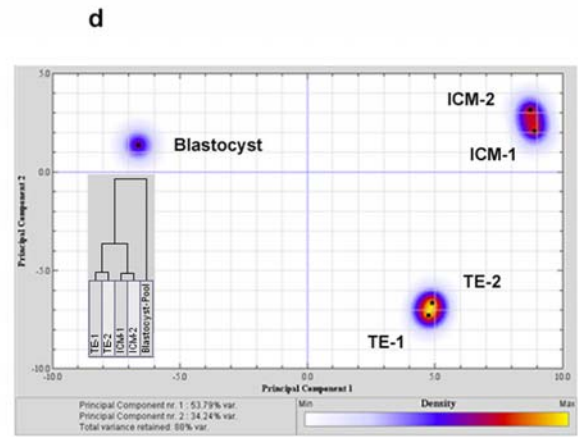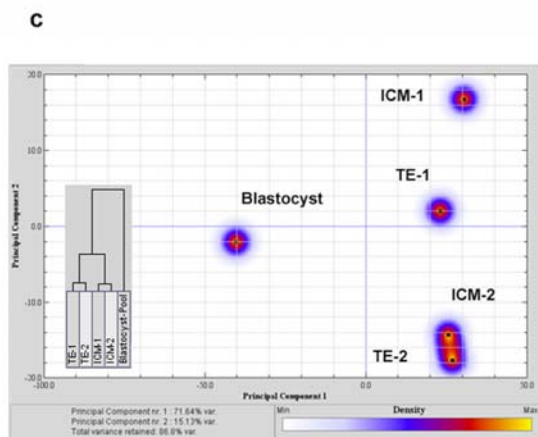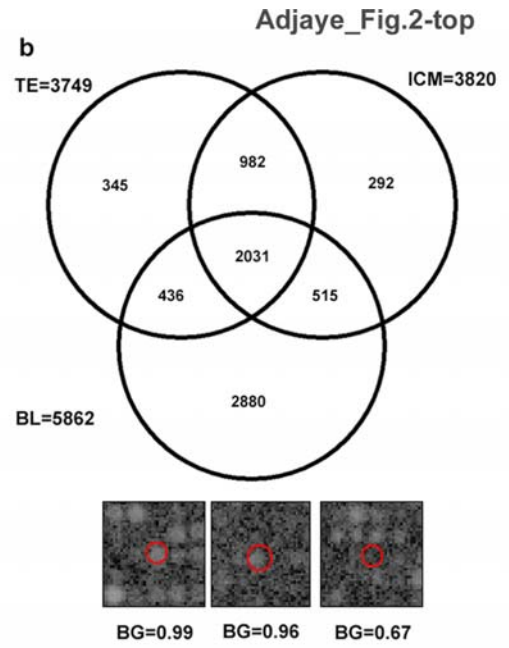
a



b



c
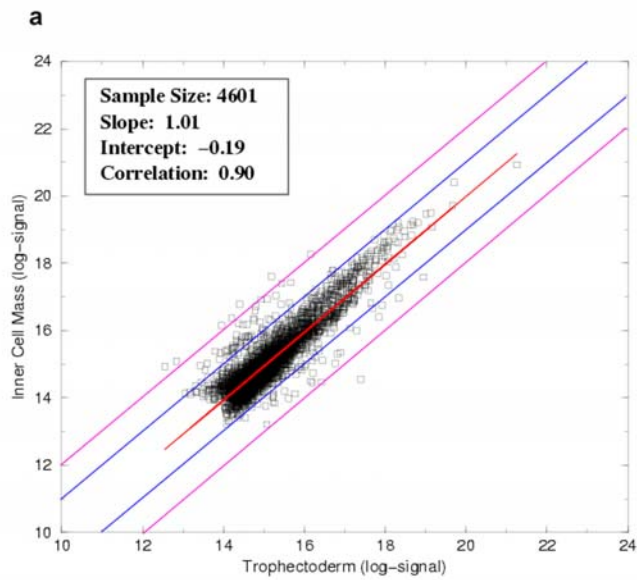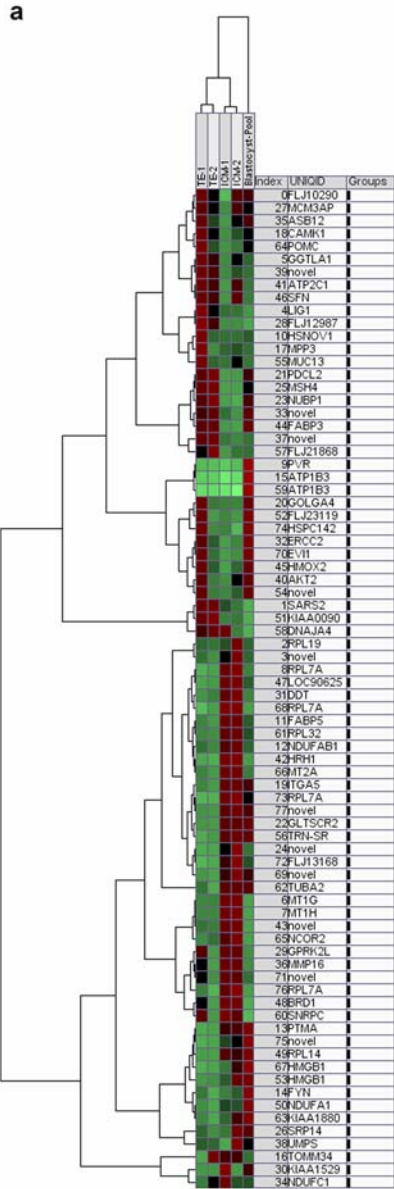


29

**a**



Sample Size: 4601
Slope: 1.01
Intercept: −0.19
Correlation: 0.90

**b**



TE=3749

ICM=3820

345    982    292

436    2031    515

BL=5862    2880

BG=0.99    BG=0.96    BG=0.67

**c**



ICM-1

TE-1

Blastocyst

ICM-2

TE-2

Principal Component nr. 1 : 71.64% var.
Principal Component nr. 2 : 15.13% var.
Total variance retained: 86.8% var.

**d**



ICM-2

ICM-1

Blastocyst

TE-2

TE-1

Principal Component nr. 1 : 53.79% var.
Principal Component nr. 2 : 34.24% var.
Total variance retained: 88% var.

30

a



b



c
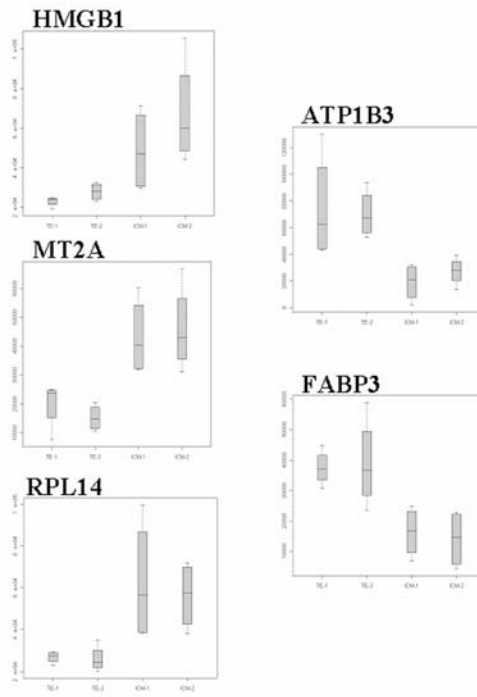
**Table 1.** Analysis of metabolic and signalling pathways operative in the blastocyst.

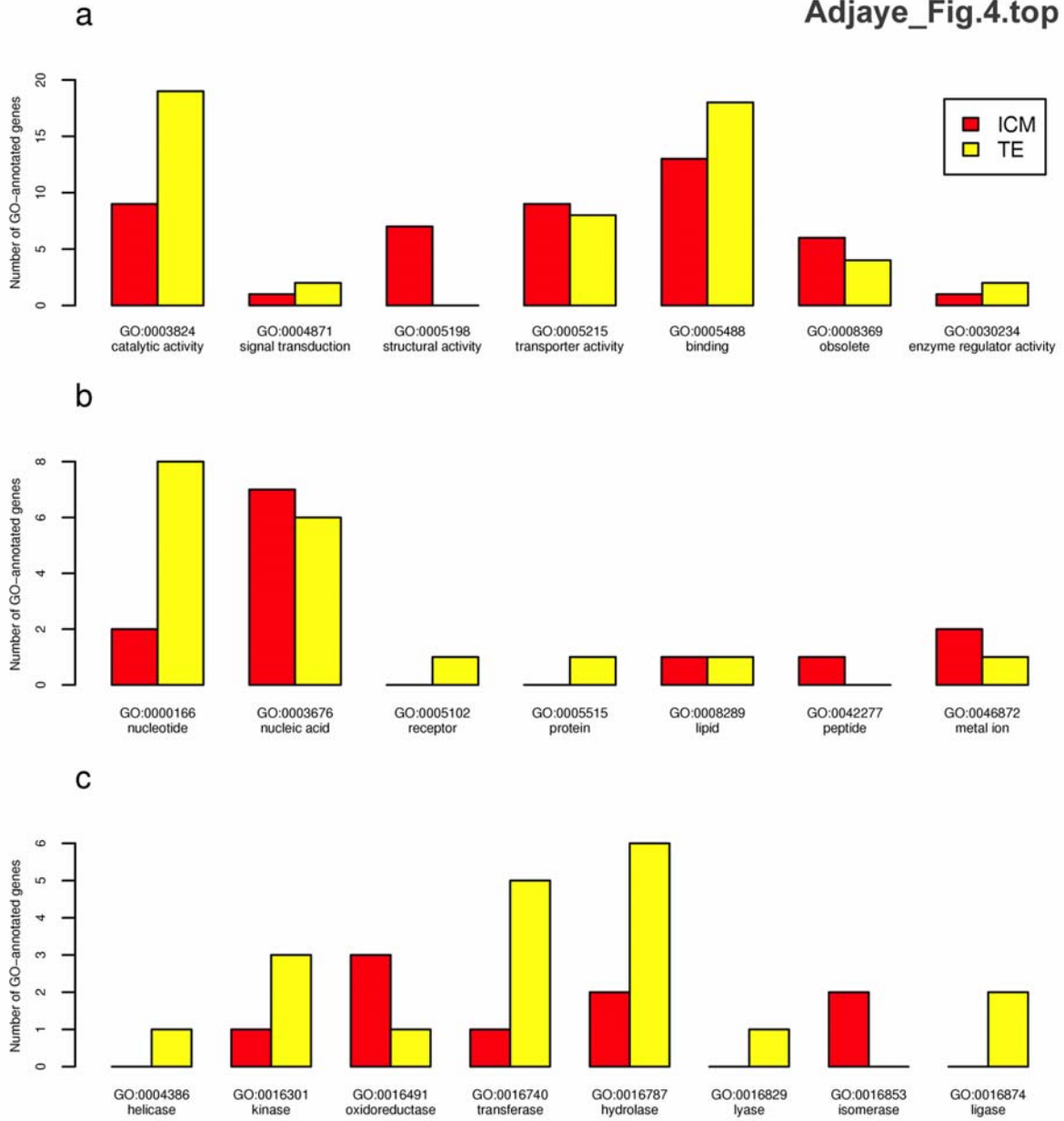| ID | PathwayDescription | #Genes | Z-Score | P-Value | TE-up | ICM-up |
|---|---|---|---|---|---|---|
| hsa04010 | MAPK signaling pathway | 131 | 5,694842 | 6,19205E-09 | 97 | 34 |
| hsa00500 | Starch and sucrose metabolism | 43 | 4,986955 | 3,07171E-07 | 35 | 8 |
| hsa04310 | Wnt signaling pathway | 72 | 4,629647 | 1,83341E-06 | 54 | 18 |
| hsa00632 | Benzoate degradation via CoA ligation | 36 | 3,927644 | 4,29089E-05 | 30 | 6 |
| hsa03010 | Ribosome | 44 | 3,792813 | 7,45E-05 | 9 | 35 |
| hsa00562 | Inositol phosphate metabolism | 31 | 3,782133 | 7,77705E-05 | 24 | 7 |
| hsa04620 | Toll-like receptor signaling pathway | 32 | 3,683691 | 0,000114972 | 24 | 8 |
| hsa00561 | Glycerolipid metabolism | 61 | 3,645259 | 0,000133595 | 46 | 15 |
| hsa00350 | Tyrosine metabolism | 29 | 3,600263 | 0,000158985 | 22 | 7 |
| hsa00280 | Valine, leucine and isoleucine degradation | 28 | 3,575113 | 0,000175077 | 24 | 4 |
| hsa04510 | Integrin-mediated cell adhesion | 37 | 3,537734 | 0,00020183 | 28 | 9 |
| hsa04610 | Complement and coagulation cascades | 38 | 3,487797 | 0,000243554 | 30 | 8 |
| hsa00340 | Histidine metabolism | 24 | 3,314286 | 0,000459445 | 21 | 3 |
| hsa00380 | Tryptophan metabolism | 47 | 3,259305 | 0,000558489 | 37 | 10 |
| hsa04070 | Phosphatidylinositol signaling system | 30 | 3,218945 | 0,000643379 | 22 | 8 |
| hsa00760 | Nicotinate and nicotinamide metabolism | 32 | 3,216218 | 0,000649524 | 22 | 10 |
| hsa00903 | Limonene and pinene degradation | 17 | 3,053308 | 0,001131736 | 16 | 1 |
| hsa00230 | Purine metabolism | 70 | 2,999244 | 0,001353321 | 49 | 21 |
| hsa00480 | Glutathione metabolism | 12 | 2,980965 | 0,001436777 | 11 | 1 |
| hsa00071 | Fatty acid metabolism | 31 | 2,939482 | 0,001643875 | 25 | 6 |
| hsa04110 | Cell cycle | 54 | 2,931779 | 0,0016852 | 39 | 15 |
| hsa00361 | gamma-Hexachlorocyclohexane degradation | 27 | 2,93105 | 0,00168916 | 23 | 4 |
| hsa00600 | Sphingoglycolipid metabolism | 36 | 2,922167 | 0,001738093 | 27 | 9 |
| hsa04350 | TGF-beta signaling pathway | 39 | 2,832865 | 0,002306706 | 28 | 11 |
| hsa00650 | Butanoate metabolism | 26 | 2,781079 | 0,002708986 | 21 | 5 |
| hsa00240 | Pyrimidine metabolism | 43 | 2,765164 | 0,002844766 | 30 | 13 |
| hsa00910 | Nitrogen metabolism | 12 | 2,745626 | 0,003019833 | 10 | 2 |
| hsa00590 | Prostaglandin and leukotriene metabolism | 13 | 2,690598 | 0,003566252 | 11 | 2 |
| hsa00360 | Phenylalanine metabolism | 15 | 2,612624 | 0,004492541 | 13 | 2 |
| hsa00450 | Selenoamino acid metabolism | 14 | 2,542448 | 0,005503972 | 11 | 3 |
| hsa00220 | Urea cycle and metabolism of amino groups | 15 | 2,499032 | 0,006226668 | 13 | 2 |
| hsa00790 | Folate biosynthesis | 11 | 2,489504 | 0,006396086 | 9 | 2 |
| hsa00640 | Propanoate metabolism | 14 | 2,479671 | 0,006575193 | 11 | 3 |
| hsa00310 | Lysine degradation | 33 | 2,456827 | 0,007008514 | 27 | 6 |
| hsa05010 | Alzheimer's disease | 21 | 2,415657 | 0,007853421 | 17 | 4 |
| hsa00563 | Glycosylphosphatidylinositol(GPI)-anchor biosynthesis | 15 | 2,38544 | 0,008529342 | 13 | 2 |
| hsa00020 | Citrate cycle (TCA cycle) | 16 | 2,378603 | 0,008689179 | 11 | 5 |
| hsa00300 | Lysine biosynthesis | 7 | 2,366432 | 0,008980222 | 7 | 0 |
| hsa00626 | Nitrobenzene degradation | 7 | 2,366432 | 0,008980222 | 7 | 0 |
| hsa00642 | Ethylbenzene degradation | 7 | 2,366432 | 0,008980222 | 7 | 0 |
| hsa00860 | Porphyrin and chlorophyll metabolism | 13 | 2,341169 | 0,009611712 | 10 | 3 |