Understanding the differentiation of human embryonic stem cells

Technical report HS- IKI -TR-06-003

PhD research proposal
School of Humanities and Informatics
University of Skövde

25 April 2006

Student: Jane Synnergren (University of Skövde)

Supervisors: Anders Lindahl (University of Gothenburg),
Björn Olsson (University of Skövde),
Patric Nilsson (University of Skövde)

Abstract

The proposed research project will apply an information fusion approach to various types of experimental data in order to increase our understanding of the differentiation of human embryonic stem (hES) cells into various specialized cell types. Gene expression profiles from hES cells in different stages of differentiation will be analysed to identify significantly over- and underexpressed genes. The purpose of the analysis is to find genes that might be important in the differentiation process and that are crucial for directing stem cells into specialized cell types. The project will focus on the endoderm development and one issue will be to increase our knowledge about two different types of endoderm development occurring in humans; primitive and definitive endoderm and their derivates. Another issue will be to compare gene expression profiles from cells that lack chromosome 13 with a sub-clone with normal karyotype (Heins *et al.*, 2004). A comparison of gene expression profiles from *in vitro* derived specialized cells with gene expression profiles from adult cells from the same tissue type will also be conducted. The project will start, however, with an investigation and validation of "housekeeping genes", which will be used for normalization and calibration of the gene expression levels in the subsequent analyses. The novelty of the project arises from the fact that most previous research in understanding the differentiation of ES cells has been done on animals (Yamada *et al*., 2002; Stainier, 2002; Asahina *et al*., 2004), while very little has been done on human ES cells. Since substantial differences both in morphology and in the gene expression pattern are well known between ES cells from these two organisms, it is important to characterize the genetic regulation of the processes responsible for these differences. This project will work with hES cells supplied by Cellartis, a company specialized in hES cell technologies.

Project description

The discovery and subsequent availability of human stem cells have rendered a new exciting area of research with many promising applications, such as treatment of currently incurable diseases with cell therapy, replacement of animal testing, and *in vitro* drug tests on human cells. Human stem cells can be divided into human adult stem cells and human embryonic stem (hES) cells depending on how they are extracted. Adult stem cells are immature cells found in differentiated tissue. They are difficult to proliferate and have a limited ability to differentiate. Embryonic stem cells can proliferate in culture and are pluripotent, which means that they can differentiate into a great variety of cell types. They are extracted from the inner cell mass of supernumerary *in vitro* fertilized (IVF) eggs, donated for research (Heins *et al.,* 2004). It has been shown that hES cells can differentiate into all the three germ layers; ectoderm, endoderm, and mesoderm (Shivdasani, 2002; Grapin-Botton, 2000).

This project will focus on increasing our understanding of how the differentiation of hES cells into various specialized cell types is regulated. The regulation of the differentiation of hES cells is extremely complex and difficult to understand, but this knowledge is crucial in order to succeed in directing *in vitro* hES cells into specialized cell types like hepatocytes, cardiomyocytes or insulin-producing beta cells. One way to increase our understanding of the regulation of hES cell differentiation is to perform gene expression analysis of cells from different stages in the development process. By comparing gene expression profiles from different cell populations, one may be able to identify genes that are involved in the differentiation process of various specialized cell types. Microarrays will be used to obtain gene expression profiles from various cell populations and as they generate such large amounts of data, efficient bioinformatics algorithms will be needed for the extraction of plausible biological hypotheses that can be experimentally tested.

In various gene expression analyses it is crucial to have stable control genes to use for normalization and calibration of results. Commonly, housekeeping genes (HKGs) are used for this purpose. HKGs are required for basal cellular function and maintenance and are assumed to be expressed at relatively stable levels across different cell types and experimental conditions (Watson *et al*., 2003). With the advent of genome-wide expression profiling, HKG mRNA levels were observed to vary extensively between different cell types (Lee *et al*., 2002), which emphasizes the importance of careful selection and validation of HKGs before using them as control genes in normalization (Abruzzo *et al*., 2005; Hoerndli *et al*., 2004). Commonly used HKGs in studies of somatic cells include glyceraldehydes-3-phosphate dehydrogenase, albumin, actins, tubulins, cyclophilin, and 18S rRNA (Eisenberg & Levanon, 2003; Hsiao *et al.,* 2001; Lee *et al.,* 2002; Warrington *et al.,* 2000) and in general, investigators have used the traditional HKGs (e.g. GAPDH, β-tubulin, β-actin) also in studies of hES cells (Bhattacharya *et al.,* 2005; Cai *et al.,* 2005; Yang *et al.,* 2005). However, it is well known that the expression of several of these genes vary considerably in adult tissues, and their suitability as HKGs in hES cells remains to be investigated. In this regard, the RNA levels of hypoxanthine phosphoribosyl transferase and β-tubulin were shown to vary substantially in differentiating mouse ES cells (Murphy & Polack, 2002). Therefore, the project will start by investigating the

stability of commonly used HKGs (see also section *Experiment 1*). If these are found not to be appropriate as control genes, an attempt will be made to identify a novel set of HKGs, which are stable in hES cells and can therefore be used as control genes in future gene expression studies on hES cells.

The next study (see also section *Experiment 2*) that will be pursued is to investigate the gene expression effects of trisomy of chromosome 13. Autosomal trisomy is a common cause of human miscarriage, malformations, and learning disability, and it is therefore a highly interesting research area. Primary gene-dosage effects have been confirmed by recent transcriptome analyses (FitzPatric *et al*., 2002). A comparison of transcriptome effects between trisomy in human and trisomy in mouse has been done and a model representing four different categories of effects was presented in (FitzPatric, 2005). Many serious diseases have been shown to be correlated to abnormalities of chromosome 13. To our best knowledge, however, no study of trisomy has included hES cells. Therefore, in this study gene expression profiles from hES cells that have three copies of chromosome 13 will be compared with a sub-clone from the same cell line having a normal karyotype.

After these two introductory studies, the focus will be on the main aim of the whole project, i.e. to study the differentiation of hES cells into what is believed to be functional hepatocytes (see also section *Experiment 3*). In drug discovery it would be incredibly valuable to have access to hepatocytes derived *in vitro*, because the primary hepatocytes from adult tissues that are used in drug discovery today are only functional for a very limited time period. They very quickly, often within 24 hours, lose their ability to produce most of the important metabolites typically expressed in the liver. To increase the understanding of the regulation of differentiation of hES cells into hepatocyte-like cells is crucial to be able to efficiently derive hepatocytes *in vitro* in a large scale manner. Therefore, the differentiating cells will be studied by harvesting cells for mRNA measurements at different time points, and conducting comparisons of the gene expression profiles between these time points. The focus of this study will be to investigate the regulation of a set of genes known to be expressed in hepatocytes.

The last part of the project will focus on increasing our understanding of endoderm development and their derivates (see also section *Experiment 4*), like e.g. hepatocytes and beta cells. But because the same bioinformatics techniques can be applied to the analysis of differentiation of other germ layers it will also be possible to switch the focus of the project if necessary. Endoderm cells can be divided into definitive endoderm (that develops into e.g. liver and pancreas) and primitive endoderm (found in the yolk sack of the embryo). It appears that it is easier to obtain primitive endoderm from *in vitro* differentiation of hES cells (Asahina *et al.*, 2004; Fujikura *et al.*, 2002; Yoshida-Koide *et al.*, 2003; and Heins and Semb, unpublished), but these cells are probably more limited in their functions than definitive endoderm. No gene expression profiling of such ES cell-derived cell types has been reported. Therefore it is extremely important to increase our knowledge about the gene expression signatures of primitive and definitive endoderm, in order to be able to discriminate between these cell types and to understand more about their biology.

Previous work

Sato and co-workers compared the molecular signature of hES cells to mouse ES cells and performed a global transcriptional profile of the state of "stemness" in hES cells (Sato *et al*., 2003). They identified a set of 918 genes enriched in undifferentiatied hES cells compared with their differentiated counterparts. Interestingly they also found that a significant number of the enriched genes in hES cells intersect with enriched genes in mouse ES cells, which indicates that a core molecular program is shared between pluripotent cell lines from different species.

The lack of stability of commonly used HKGs in stem cells is highlighted in a work by Murphy & Polack (2002) were the suitability of three commonly used HKGs, (GAPDH, HPRT and beta-tubulin) as internal standards for measuring RNA levels are investigated. In their study neither HPRT nor beta-tubulin are found to be suitable as control genes, since they vary substantially during early differentiation in the experiment. However, their study used mouse ES cells, while the degree of stability of common HKGs in hES cells remains to be investigated. Microarrays have previously been used to identify novel HKGs in somatic cells (Warrington *et al*., 2000; Eisenberg *et al*., 2003). Eisenberg and co-workers analysed cells from 47 different tissues and identified a set of 575 genes that were stable across all investigated tissue types. Warrington *et al*. analysed gene expression data from 11 different human and fetal tissues and identified 535 genes involved in cellular maintainence. However, none of these studies included hES cells in their experiments.

FitzPatric and co-workers used microarrays to perform transcriptome analyses of primary cultures of human fetal cells from pregnancies affected with trisomy 21 and trisomy 13 (FitzPatric *et al*., 2002). They explored the chromosomal expression levels and could clearly show a misregulation of all the chromosomes, and not only the trisomic one, in cells affected by trisomy. They presented a list of genes that were significantly up- or downregulated in the trisomic cells, compared with cells with normal karyotype. In their study they used amniocyte cultures extracted from pregnancies affected by trisomy 13 or trisomy 21. The present study will analyse if hES cells with trisomy 13 are affected in similar ways as the investigated amniocytes.

Researchers have for several years used mouse models to learn how to direct ES cells into functional specialized cell types, usable for therapeutic transplantation. Milne *et al.* (2005) generated insulin-expressing cells from mouse ES cells. They suggest that their insuline producing cells are probably derived from the extra-embryonic endoderm origin, which *in vivo* develops to become the yolk sack. Recently a highly interesting work by D'Amour and co-workers was published, were they tried to direct hES cells to definitive endoderm (D'Amour *et al*., 2005). They showed that differentiation of hES cells in the presence of activin A and low serum produced cultures consisting of up to 80% definitive endoderm cells. Their results increase the possibility to succeed in directing hES cells into functional cells, usable in therapeutic applications.

Microarray data analysis

An overview of the general data analysis workflow in a typical microarray analysis project is shown in Figure 1. Before the gene expression profile of a cell type can be analysed, the microarray data must be pre-processed by processing of missing values, filtering out housekeeping genes, and normalization. These are well-studied problems, and standard methods will be used (for details, see e.g. Bolstad *et al*., 2003; Bø *et al*., 2004; Hill *et al*., 2001).

After the pre-processing steps, genes with altered expression levels between the experiments will be identified using Tusher *et al*.'s (2001) method for significance analysis of microarrays (SAM). Unlike conventional statistical methods such as *t* tests, SAM has been adapted specifically for microarray experiments, and solves issues such as the large number of false positives produced when *t* tests, fold change, or pairwise fold change are applied to datasets of thousands of genes. This is achieved by taking into account that variation is gene-specific, and therefore defining a statistic based on the ratio of expression change for each gene to the standard deviation for the same gene in the data ("gene-specific scatter"). SAM also computes an estimated false positive rate from permutations of the expression data, so that the significance threshold can be chosen in such a way that the expected number of false positives is controlled.

Apart from individual differentially expressed genes, also groups of genes with small but coordinated changes will be identified, using gene set enrichment analysis (GSEA) and pathway level analysis of gene expression (PLAGE) (Mootha *et al*., 2003; Tomfohr *et al.,* 2005). As demonstrated by both Mootha *et al*., and Tofohr et al., this is necessary in cases where the difference in expression is very modest for each individual gene in a pathway, while at the same time the difference in expression is consistent across all genes in the pathway. As known from metabolic control theory, small (and statistically insignificant) adjustments in many sequential steps of a pathway can lead to substantial change in the total flux, while large changes in a single step may have no measurable effect. It is therefore necessary to try to identify expression changes, which manifest at the level of pathways rather than at the level of individual genes.

Since several signalling pathways (such as Wnt, Stat, Notch, and Nodal) have been shown to be involved both in maintaining hES cells in an undifferentiated state and in differentiation of hES cells into specialized cell types (Brandenberger *et al*., 2004; Sato *et al*., 2004), we will carry out a pathway-level analysis of the expression data. Sato *et al*. (2004) have shown that Wnt pathways are up-regulated in undifferentiated hES cells and that they might be involved in maintaining the important pluripotency of hES cells. Terami *et al*. (2004) suggest that the Wnt11 pathway plays a crucial role in control of morphogenesis in several tissues and is possibly involved in cardiac differentiation of ES cells. Statistical methods, described above, will be used to analyse expression values of known pathways in a systematic and rigorous manner.

Materials and methods used in gene expression profiling

Several experiments will be performed during the project. Below follows a short description of each of the planned experiments, and the biological material and microarray equipment to be used.

Experiment 1

The first experiment will be to investigate if commonly used HKGs in somatic cells are stable also in gene expression data from hESCs and therefore useful as control genes in subsequent analyses. The study will also include identifying a novel set of HKGs that are stably expressed in undifferentiated and early differentiation of hESCs. Data from three different cell lines (SA001, SA002, and SA002.5; Cellartis hES cell lines are further described in Heinz *et al*. 2004) are used in the experiment, and RNA from five different differentiation samples per cell line are extracted.

- SA001 is a hES cell line with normal karyotype.
- SA002 is a hES cell line with an abnormal karyotype that has an identified triplicate of chromosome 13.
- SA002.5 is a sub-clone of line SA002 having two chromosome 13, and thus has a normal karyotype.

An overview of the sample sets is shown in Figure 2.

The materials used in the experiment are hES cells in different stages of differentiation: 1) undifferentiated hES cells 5 days after passage[1]; 2) cells grown in medium on plate for 11 days without passage; 3) cells grown in medium on plate for 25 days without passage; 4) cells grown in medium on plate for 5 days and then cut in aliquots and kept in suspension for 6 days to allow formation of embryonic bodies (EBs), and develop a 3D structure that facilitates differentiation; and 5) cells from plated EBs 14 days after plating.

The microarray equipment used in this experiment is the CodeLink microarray platform manufactured by Amersham Bioscience (General Electric Healthcare). CodeLink arrays differ from other commercial platforms in several ways. It is a one channel, single probe platform that uses a three-dimensional (3-D) surface, which is claimed to have a larger number of potential attachment sites than the more common 2-D surfaces, and to thereby facilitate the interaction between probe and target (Ramakrishnan *et al*., 2002, Amersham Inc. [online][2]). They use a 3-D polyacrylamide aqueous gel matrix that immobilizes amine-terminated oligonucleotide probes, which is claimed to allow greater target access to probes and better sensitivity than with a 2-D matrix. CodeLink microarrays have 30mer oligonucleotide probes instead of 25mer probes, which is used by the well-known competitor Affymetrix, claiming that they obtain better specificity with longer probes.

Experiment 2

The second experiment will be to further analyse the data generated in experiment 1 from a new perspective. The research question is how a trisomy of chromosome

[1] Cutting the cell culture in small aliquots which are re-plated.
[2] http://www5.amershambiosciences.com

13 affects the gene expression profile in hES cells. The data generated from experiment 1 will be very suitable for this analysis as it consists of two normal and one trisomic cell line, where SA002 and SA002.5 are derived from the same individual and therefore have exactly the same genetic material. With this set of biological material, the chances increase that any identified differences actually derive from the trisomic chromosome and not from genetic variations between individuals.

Experiment 3

The third experiment will be to analyse microarray data from two customized microarrays where: 1) genes that are particularly interesting in drug metabolism in the liver will be represented; and 2) specific drug targets in hepatocytes like cholesterol- and bile acid transporters will be represented. In the pharmaceutical industry, it is crucial to gain an increased understanding of metabolic pathways that are central for metabolism of various drugs. Furthermore, liver metabolism and the interplay between liver and other organs are important drug targets for metabolic and dyslipidemic diseases. Therefore, if one could direct the hES cells to develop into functional hepatocytes, these would be invaluable tools for the pharmaceutical industry by allowing improved target finding and evaluation and *in vitro* testing of drug metabolism. The hepatocyte-like cells cultured at Cellartis will be used for these microarray experiments to investigate if they have the desired functionality. Customized microarrays from the Affymetrix platform will be used for these experiments.

Experiment 4

As illustrated in Figure 3, the fate of a cell is determined during early embryogenesis through the gastrulation, where the cell differentiates towards one of the three germ layers; mesoderm, ectoderm, or endoderm (Shivdasani, 2002; Grapin-Botton, 2000). Despite the importance of the fate of the cell at this early stage in development, little is known of how this fate is controlled by gene regulation. It is therefore interesting to analyze the differences in gene expression during these very early developmental stages and to identify genes having expression patterns that differ between the three germ layers.

The endoderm that gives rise to mature tissue cells such as liver and pancreas cells is termed *definitive, or embryonic* endoderm. In addition, there is also endoderm development in the yolk sack, termed *primitive, or extraembryonic* endoderm. One purpose of the primitive endoderm is to support the embryo with metabolic functions during the very early stages of development, when the liver is not yet developed (Shivdasani, 2002). It is assumed that the primitive endoderm is not as fully functional as the definitive endoderm but this hypothesis needs further investigation. Therefore, the intension is to compare gene expression profiles from the two endoderm types and analyse any identified differences in gene activity.

To attain meaningful data, homogenous cell populations are required where primitive and definitive endoderm cells are isolated. As mentioned in Luo *et al*., (2003) the majority of genes involved in regulation of cell differentiation are often only expressed for a very short period, and therefore several time points with short time intervals will be measured to identify variations in the gene expression profiles between the primitive and definitive endoderm. Several cell lines will be

used in this experiment to facilitate comparison between different hES cell lines. It is of interest to investigate if any differences between the cell lines can be identified concerning how they spontaneously differentiate into one of the three germ layers. The Affymetrix whole genome microarray, which has probes representing ~42,000 known human genes, will be used for this experiment. It is important to use a wide range array because the purpose is to identify genes that are not yet known to have an impact on the differentiation process.

Information fusion approach

Whereas initial steps in microarray data analysis are based on more-or-less standard statistical methods, as described in the previous section, additional steps in the analysis are based on the application of various data mining techniques and on the integration of many different data sources. Although clustering (see fig 1) has become a well-studied standard approach in microarray data analysis (Eisen *et al*., 1998; Ben-Dor *et al*., 1999; Yeung and Ruzzo, 2001; Dougherty *et al*., 2002; McLachlan *et al*., 2002; and many other), the application of other data mining techniques is still an open research question and methods for biological relevance analysis are still rudimentary, at best. This means that any results from the application of such data mining algorithms must be evaluated against other sources. For this reason, information gained through various sources will be combined into an overall understanding of the differentiation processes studied in the project. Examples of information sources that will be used include:

- Data mining results gained by analysis of the gene expression data produced in the experiments that are part of the project (see "Materials and methods used in gene expression profiling").
- Results from analysis of published gene expression data from other work both from hES cells and mouse ES (mES) cells. For example, Sato *et al*. (2003) collected whole genome gene expression data for undifferentiated hES cells and for differentiated hES cells, co-cultured into neurons using a co-culturing system described by Kawasaki *et al*. (2002). Ramalho-Santos *et al*. (2002) have collected microarray gene expression data for mES cells.
- Results from a large scale sequencing project performed by Brandenberger *et al*. (2004) where expressed sequence tags (ESTs) where generated from hES cells. Through EST frequency analysis, where they compared the frequency of ESTs for each gene in different cell populations, genes that where differentially expressed in undifferentiated and differentiated hES cells where identified. Their results were verified in several ways, e.g. by rt-PCR.
- Results from gene expression analyses with QPCR in collaboration with TATAA (see www.tataa.com). QPCR is useful for verifying results from microarray experiments since it provides a higher level of accuracy than microarrays for the investigated genes. However, the genes of interest must be known and QPCR can therefore be useful as a complement to microarrays.
- DNA- and amino acid sequence databases, which can be used as sources of functional annotation and the loci of genes.

- Annotation ontologies, such as Gene Ontology (Ashburner *et al*., 2000), useful for analysis of the molecular function of genes.
- Pathway databases, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehis and Goto, 2000), which can be used to identify groups of genes, involved in a particular pathway. A prerequisite when applying the GSEA and the PLAGE algorithms (see Microarray data analysis section) is that the genes in the investigated pathway are known in advance (Mootha *et al*., 2003, Tomfohr *et al,* 2005)
- Systematic methods for evaluation of the biological plausibility of data mining results, such as the method under development by Gamalielsson *et al*. (Gamaliesson et al, 2006).

Figures

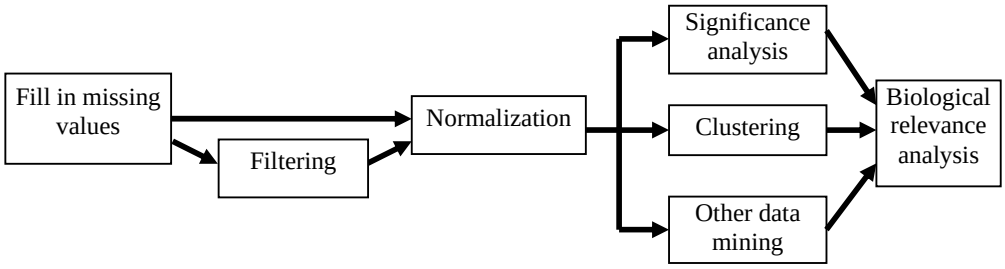

Figure 1. Overview of microarray data analysis steps. For explanations, see sections "Microarry data analysis" (missing values, filtering, normalization, significance analysis) and "Information fusion approach" (clustering, other data mining, biological relevance analysis).

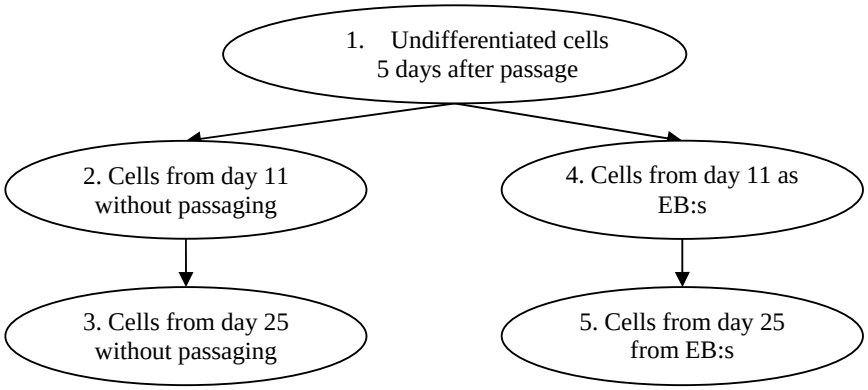

Figure 2. Overview of experiment 1, on comparing differentiation protocols.

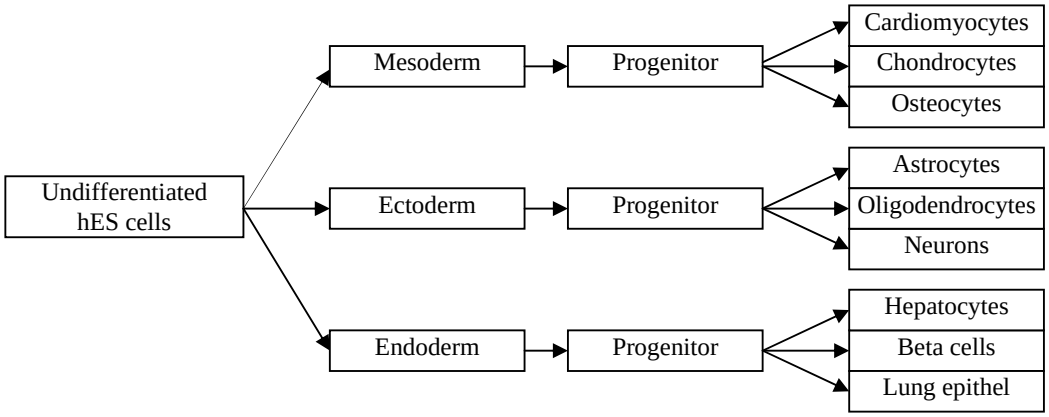

Figure 3. Overview of the main paths from undifferentiated stem cells to specific tissues. The listed cell types are examples of cells derived from the different germ layers.

References:

Abruzzo, L.V., Lee, K.Y., Fuller, A., Silverman, A., Keating, M.J., Dedeiros, L.J., and Coombes, K.R. (2005) Validation of oligonucleotide microarray data using microfluidic low-density attays: a new statistical method to normalize real-time RT-PCR data. *Biotechniques* 38(5):785-92.

Asahina, K., Fujimori,H., Shimizu-Saito, K., Kumashiro, Y., Okamura, K., Tanaka, Y., Teramoto, K., Arii, S., and Teraoka, H. (2004) Expression of the liver-specific gene Cyp7a1 reveals hepatic differentiation in embryoid bodies derived from mouse embryonic stem cells. *Genes to Cell* 9:1297-1308.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., and Sherlock, G. (2000) Gene ontology: tool for the unification of biology. *Nature Genetics* 25(1): 25-29.

Ben-Dor, A., Shamir, R., and Yakhini, Z. (1999) Clustering gene expression patterns. *Journal of Computational Biology* (3/4): 281-297.

Bhattacharya, B., Cai, J., Lou, Y., Miura, T., Mejido, J., Brimble, S.N., Zeng, X., Schulz, T.C., Rao, M.S., and Puri, R.K. (2005) Comparison of the gene expression profile of undifferentiated human embryonic stem cell lines and differentiation embryoid bodies. *BMC Developmental Biology 5(22).*

Bolstad, B.M., Irizarry, R.A., Åkerstrand, M., and Speed, T.P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19(2): 185-193.

Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J., Li, Y.,Xu, C., Fang, R., Guegler, K., Rao, M.S., Mandalam, R., Lebkowski, J., and Stanton, Lawrence, W.S. (2004) Transcriptome characterization elucidates signalling networks that control human ES cell growth and differentiation. *Nature Biotechnology* 22(6): 707-716.

Bø, T.H., Dysvik, B., Jonassen, I. (2004) LSimpute: accurate estimation of lmissing values in microarray data with least squares methods. *Nucleic Acids Research* 32(3): e34.

Cai, J., Cheng, J., Liu, Y., Miura, T., Luo, J., Loring, J.F., Freed, W.J., Rao, M.S., and Zeng, X. (2005) Assessing self-renewal and differentiation in hESC lines. *Stem Cells* Nov 17; [Epub ahead of print]

D'Amour, K.A., Agulnick, A.D., Eliazer, S., Kelly, O.G., Kroon, E., and Baetge, E.E. (2005) Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nature Biotechnology 23(12):* 1534-41

Dougherty, E.R., Barrera, J., Brun, M., Kim, S., Cesar, R.M., Chen, Y., Bittner, M., and Trent, T.M. (2002) Inference from clustering with application to gene-expression microarrays. *Journal of Computational Biology* 9: 105-126.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the USA* 95: 14863-14868.

Eisenberg, E., and Levanon, E. (2003) Human housekeeping genes are compact. *TRENDS in Genetics* 19(7)*:* 362-365.

Fujikura, J., Yamato, E., Yonemura, S., Hosoda, K., Masui, S., Nakao, K., Miyazaki, J., and Niwa, H. (2002) Differentiation of embryonic stem cells is induced by GATA factors. *Genes & Development 16*: 784-789.

FitzPatric, D.R., Ramsay, J., McGill, N., Shade, M., Carothers, D.A., and Hastie, N.D. (2002) Transcriptome analysis of human autosomal trisomy. *Human Molecular Genetics* 11(26): 3249–3256

FitzPatric, D.R. (2005) Transcriptional consequences of autosomal trisomy: primary gene dosage with complex downstream effects. *Trends in Genetics* 21(5): 249-253.

Gamalielsson, J., Olsson, B., and Nilsson, P. (2006) A Gene Ontology based Method for Assessing the Biological Plausibility of Regulatory Hypotheses. *In Proceedings of the 2:nd International Workshop on Bioinformatics Research and Applications (IWBRA 2006)*, Reading, Great Britain, May 2006.

Grapin-Botton, A. (2000) Endoderm development from patterning to organogenesis. *Trends in Genetics* 16(3): 124-130.

Heins, N., Englund, M.C.O., Sjöblom, C., Dahl, U., Tonning, A., Bergh, C., Lindahl, A., Hansson, C., Semb, H. (2004) Derivation, characterization, and differentiation of human embryonic stem cells. *Stem Cells* 22: 367-376.

Hill, A.A, Brown, E.L., Whitley, M.Z., Tucker-Kellogg, G., Hunter, C.P., and Slonim, D. (2001) Evaluation of normalization procedures for oligonucleotide array data based on spiked cRNA controls. *Genome Biology* 2(12): research 0055.

Hoerndli, F.J., Toigo, M., Schild, A., Gotz, J., and Day, P.J. (2004) Reference genes identified in SH-SY5Y cells using custom-made gene arrays with validation by quantitative polymerase chain reaction. *Analytical Biochemistry* 335(1): 30-41.

Hsiao, L.L., Dangond, F., Yoshida, T., Hong, R., Jensen, R.V., Misra, J., Dillon, W., Lee, K.F., Clark, K.E. Haverty, P., Weng, Z., Mutter, G.L., Frosch, M.P., Macdonald, M.E., Milford, E.L., Crum, C.P., Bueno, R., Pratt, R.E., Mahadeyappa, M., Warrington, J.A., Stephanopoulos, G., and Gullans, S.R. (2001) A compendium of gene expression in normal human tissues. *Physiological Genomics* 7(2):97-104.

Kanehisa, M. and Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28: 27-30.

Kawasaki, H., Suemori, H., Mizuseki, K., Watanabe, K., Urano, F., Ichinose, H., Haruta, M., Takahashi, M., Yoshikawa, K., Nishikawa, S., Nakatsuji, N., and Sasai, Y. (2002) Generation of dopaminergic neurons and pigmented epithelia from primate ES cells by stromal cell-derived inducing activity. *Proceedings of the National Academy of Sciences of the USA 99*: 1580-1585.

Lee, P.D., Sladek, R., Greenwood, C.M.T., and Hudson, T.J. (2002) Contrrol Genes and Variability: Absence of Ubiquitous Reference Transcripts in Diverse Mammalian Expression Studies. *Genome Research* 12(2): 292-297.

Lipshutz, R.L., Fodor, P.A., Gingeras, T.R., Lockhart, D.J. (1999) High desity synthetic oligonucleotide array. *Nature Genetics* 21: 20-24.

Luo, Y., Cai, J., Ginis, I., Sun, Y., Lee, S., Yu, S.X., Hoke, A., Rao, M. (2003) Designing, testing, and validating a focused stem cell microarray for characterization of neural stem cells and progenitor cells. *Stem Cells* 21: 575-587.

McLachlan, G.J., Bean, R.W., and Peel, D. (2002) A mixture model-based approach to the clustering of microarray expression data. *Bioinformatics* 18(3): 413-422.

Milne, H.M., Burns, C.J., Kitsou-Mylona, I., Luther, M.J., Minger, S.L., Persaud, S.J., and Jones, P.M. (2005) Generation of insulin-expressing cells from mouse embryonic stem cells. Biochemical *and Biophysical Communications 328:* 399-403.

Mootha, V.K., Lindgren, C.M., Eriksson, K-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., Houstis, N., Daly, M.J., Patterson, N., Mesirov, J.P., Golub, T.R., Tamayo, P., Spiegelman, B., Lander, E.S., Nirschhorn, J.H., Altshuler, D., and Groop, L.C. (2003) *Nature Genetics* 34(3): 267-273.

Murphy, C.L., and Polack, J.M. (2002) Differentiation Embryonic Stem Cells: GAPDH, But Neither HPRT Nor β-Tubulin Is Suitable as an Internal Standard for Measuring RNA Levels. *Tissue Enginering 8(4):* 551-558.

Ramakrishnan,R., Dorris,D., Lublinsky,A., Nguyen,A., Domanus,M., Prokhorova,A., Gieser,L., Touma,E., Lockner,R., Tata,M., Zhu,X., Patterson,M., Shippy,R., Sendera,T.J., & Mazumder, A. (2002) An assessment of Motorola CodeLink TM microarray performance for gene expression profiling applications. *Nucleic Acid Research* 30(7): e30.

Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R.C., and Melton, D.A. (2002) "Stemness": Transcriptional profiling of embryonic and adult stem cells. *Science* 298 (5593): 597-600.

Sato, N., Sanjuan, I.M., Heke, M., Uchida, M., Naef, F., and Brivanlou, A.H. (2003) Molecular signature of human embryonic stem cells and its comparison with the mouse. *Developmental Biology* 260: 404-413.

Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., and Brivanlou, A.H. (2004) Maintance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nature Medicine* 10(1): 55-63.

Shivdasani,R.A. (2002) Molecular Regulation of Vertebrate Early Endoderm Development. *Developmental Biology* 249: 191-203.

Stainier, D.Y.R. (2002) A glimpse into the molecular entrails of endoderm formation. *Genes & Development* 16: 893-907.

Terami, H., Hidaka, K., Katsumata, T., Iio, A., Morisaki, T. (2004) Wnt11 facilitates embryonic stem cell differentiation to Nkx2,5-positive cardiomyocytes. *Biochemical and Biophysical Research Communication* 325: 968-975.

Tomfohr, J., Lu, J., Kepler, T.B. (2005) Pathway level analysis of gene expression using singular value decomposition. *BMC Bioinformatics* 6(225).

Tuscher, V.G., Tibshirani, R., and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the USA* 98(9): 5116-5121.

Warrington, J.A., Nair, A., Mahadevappa, M., and Tsyganskaya, M. (2000) Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. *Physiol Genomics 2:* 143-147.

Watson, J.D. (2003) *Molecular Biology of the Gene*. Benjamin/Cummings Pub.Co. USA

Yamada, T., Yoshikawa, M., Kanda, S., Kato, Y., Nakajima, Y., Ishizaka, S., and Tsunoda, Y. (2002) In vitro differentiation of embryonic stem cells into hepatocyte-like cells identified by cellular uptake of indocyanine green. *Stem Cells* 20: 146-154.

Yang, A.X., Meijdo, J., Luo, Y., Zeng, X., Schwartz, C., Wu, T., Thies, R.S., Bhattacharya, B., Han, J., Freed, B., Rao, M.S., and Puri, R.K. (2005) Development of a focused microarray to assess human embryonic stem cell differentiation. *Stem Cells Development* 14: 270-284.

Yeung, K.Y. and Ruzzo, W.L. (2001) Prinicipal component analysis for clustering gene expression data. *Bioinformatics* 17: 763-774.

Yoshida-Koide, U., Matsuda, T., Saikwa, K., Nakanuma, Y., Yokota, T., Asashima, M., and Koide, H. (2003) Involvement of Ras in extraembryonic endoderm differentiation of embryonic stem cells. *Biomedical and Biophysical Communications* (*BBRC*) 313: 475-481.