We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



19

Microarray Analysis of Undifferentiated and Differentiated Human Pluripotent Stem Cells

Jane Synnergren[1,2] and Peter Sartipy[3]
[1]*University of Skövde,*
[2]*Sahlgrenska University Hospital,*
[3]*Cellartis AB*
Sweden

1. Introduction

Over the last decade, a tremendous progress has been made regarding our understanding of the molecular program involved in early human development. The main reason behind these advancements can be ascribed to the successful isolation of human embryonic stem cell (hESC) lines in the late 1990's. Based on their fundamental properties of pluripotency and unlimited proliferation, these unique cells have provided the possibility to study early human developmental processes in vitro. However, there are many obstacles to overcome before the potential of these cells can be fully realized. One important issue is to increase the understanding about the gene regulatory mechanisms that control the differentiation of hESCs. A wide variety of tools and technologies have been used to manipulate and study basic hESC characteristics and functions. Furthermore, the parallel analysis of functional derivatives of hESCs has provided important insights into the mechanisms that govern their differentiation into specific cell lineages. Global transcriptional changes in cells and tissues can be studied using molecular techniques such as DNA microarray, EST-enumeration, MPSS profiling, and SAGE. The results from such experiments provide a snapshot of the status of the cells under study. This approach has proven well suited for characterization of the "stemness" state of hESCs, but also for the identification of crucial pathways involved in their differentiation. Large scale gene expression databases have been generated using various hESC lines and technical platforms, and subsequently the information has been analyzed using different bioinformatic approaches. A discrete set of genes has been identified which are highly expressed in hESCs, and these genes are considered to be involved in preserving the pluripotency and self-renewal capacity of the undifferentiated cells. Furthermore, several studies have focused on characterizing the molecular signature of specific differentiation processes. Again, global expression analysis has proven to be a very suitable tool since novel important mechanisms can be revealed in the context of such experiments. More recently, the possibility to analyze also the global expression profile of microRNA (miRNA) has been realized, and microarray based platforms designed specifically for the detection of miRNA species are now available. The concurrent analysis of the global mRNA and miRNA expression profiles of hESCs and their differentiated progenies are anticipated to provide additional insights into the regulatory pathways which are active in the cells in the undifferentiated and differentiated states. In the present chapter, we will

discuss issues regarding the platforms used and the bioinformatic tools applied to interpret the results and also review some important studies on transcriptional profiling of hESCs.

2. Gene transcription and protein translation

Gene expression is the process by which information from a gene is copied from the gene to an mRNA sequence, which is then used in the synthesis of a functional gene product, a protein. The properties of the expression products give rise to the phenotype of an organism. When conducting gene expression studies, it is important to understand the basic concepts behind these processes, for a proper interpretation of the data. The genetic code is mediated by the gene expression, and the process from transcription of a gene to a functional protein involves several steps, such as transcription of the gene in the nucleus, and transport of the mRNA to the cytoplasm where translation to a protein is carried out aided by ribosomes. What is usually measured with microarrays is the amount of mRNA in the cell. However, in recent years, the microarray technology has also been successfully used for assessment of miRNA expression and moreover, protein arrays have begun to enter the market. By means of gene regulation the cell has control over its structure and function, and this is the basis for cellular differentiation, morphogenesis, and the versatility and adaptability of any organism. Transcriptional regulation is also essential for evolutionary changes, since control of the timing, location, and amount of gene expression often have profound effects on the functions of the gene in a cell.

2.1 Transcriptional regulation

The transcription of genes involves intricate dynamic dependencies which makes it challenging to study. Several mechanisms have been shown to be critical for the initiation of transcription, the rate of transcription, and the subsequent processing of the mRNA. These regulatory mechanisms control when the transcription occurs and the amount of mRNA produced. The transcription of a gene is carried out by RNA polymerase and the process is regulated by several components (Alberts et al., 2004):

- *Specificity factors* control the ability for the RNA polymerase to bind to a specific promoter or set of promoters.
- *Repressors* bind to non-coding regions, close to or overlapping with the promoter for a gene, and impede the RNA polymerase's progress along the DNA strand, thus hampering the transcription of the gene.
- *General transcription factors* aid in positioning the RNA polymerase at the start of a protein coding sequence.
- *Activators* enhance the interaction between the RNA polymerase and the specific promoter.
- *Enhancers* are sites on the DNA helix that are bound to by activators in order to loop the DNA and bring a specific promoter to the initiation complex.

As microarrays provide the opportunity to globally monitor the transcriptome and the expression of interacting molecules, they provide a powerful tool for studies of transcriptional regulation.

2.2 Splicing of mRNAs

Splicing is a modification of an RNA sequence post-transcription, in which introns are removed and exons are joined together (Fig. 1.) Hence, after transcription of a gene, the pre-

mRNA is spliced to mRNA, which is often performed in a series of reactions. RNA splicing allow for packing of more information into every gene as the transcripts from one single gene can be spliced in various ways to produce different mRNAs, depending on the cell type in which the gene is being expressed or the stage of the development of the organism (Alberts et al., 2004). As a consequence, different proteins can be produced by the same gene and it is estimated that 60% of the human genes undergo such alternative splicing (Alberts et al., 2004). Thus, RNA splicing increases the already enormous coding potential of eukaryotic genomes, at the same time as it complicates the studies of gene transcription. This is because the complexity increases dramatically when there, as in many cases, are several different transcripts transcribed by one single gene.

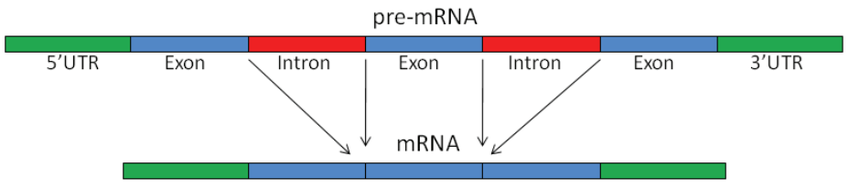


Fig. 1. Splicing of pre-mRNA where introns are removed before formation of the mRNA sequence. Different sets of exons can be selected to form the mRNA, which means that one pre-mRNA can give rise to several variants of mRNA sequences.

2.3 Translation to protein

After the splicing, the mRNA transcript is transported from the nucleus to the cytoplasm where the translation occurs by means of ribosomes. The same mRNA sequence can be translated many times, and therefore the period of time that a mature mRNA molecule persists in the cell, influences the amount of protein that is produced. The lifetime of mRNAs differs considerably and cannot be assessed by using single time point microarray analyses, as these only gives a snap-shot of the amount of mRNA at a specific time point. The mRNA lifetime is dependent on a multitude of factors, such as the nucleotide sequence of the mRNA itself, as well as the type of cell in which the mRNA is produced. The typical lifetime for mRNA molecules in eukaryotic cells ranges from 30 minutes up to 10 hours (Alberts et al., 2004). In the ribosome the nucleotide sequence is translated into an amino acid sequence by means of the *genetic code*. The sequence of nucleotides in the mRNA is read in groups of three, *codons*, and each codon specifies one amino acid. Depending on where in the sequence the de-coding begins, each mRNA sequence can be translated in three different, non-overlapping, reading frames but only one of these is the correct one (Alberts et al., 2004). To control that correct reading frame are used, the translation of an mRNA begins with a specific start codon (AUG), and is then performed in the direction 5′ cap to 3′ end. The end of a protein coding mRNA is indicated by the presence of one of three stop codons (UAA, UAG, UGA) which signal to the ribosome to stop the translation.

3. MicroRNAs

An additional level of cellular regulation involves a family of tiny molecules, known as miRNAs. These are 19–25 nucleotides non-coding RNAs that bind to the 3′ untranslated region of target mRNAs through imperfect matching. In mammalian genomes miRNAs are predicted to regulate the expression of approximately 30% of the protein-coding genes

(Bartel, 2004). Knowledge about the biological functions of most miRNAs identified thus far is still lacking, but it has been shown that they play important roles in embryo development, determination of cell fate, cell proliferation, and cell differentiation (Sartipy et al., 2009, Wang et al., 2009). MicroRNAs are derived from approximately 70 nucleotide long precursors, encoded by introns or intergenic regions, and are expressed in most organisms ranging from plants to humans. Many miRNAs appear to be expressed at different levels in various tissues, and the maturation and function of the tissues seem to be influenced by their presence. Interestingly, results from recent studies have indicated important roles for miRNAs in the control of diverse aspects of heart formation and cardiac function (Ivey et al., 2008, van Rooij & Olson, 2007). It is also known that miRNAs are involved in various types of cancer by targeting tumor suppressing genes (Lu et al., 2005, Zhu et al., 2008). MicroRNAs bind to their target mRNAs and negatively regulate their expression, either by repression of translation or by degradation of the mRNA (Bartel, 2004). Increased expression levels of miRNAs can also result in up-regulation of previously suppressed target genes either directly, by decreasing the expression of inhibitory proteins and/or transcription factors, or indirectly, by inhibiting the expression levels of inhibitory miRNAs (Gregory et al., 2008). Depending on the state of the cell, miRNAs have also been observed to affect the translation of target mRNAs by regulation of their stability (Gregory et al., 2008, Vasudevan et al., 2007). Moreover, it has been shown that combinatorial regulation by miRNAs is common, which enables complex regulatory programs that are exceptionally challenging to dissect (Zhou et al., 2007).

4. Global transcriptional profiling techniques

There are several high throughput techniques for measuring gene expression at the large scale, such as expressed sequence tags (EST)-enumeration, Serial Analysis of Gene Expression (SAGE), Massively Parallel Signature Sequencing (MPSS) and different types of microarrays (described in more detail below). In EST-enumeration the expression levels are assessed by counting the number of ESTs for a particular gene, randomly selected from a cDNA library derived from the sample. The ESTs are clustered into groups of sequences originating from the same transcript, and subsequently assembled to achieve a longer consensus sequence, which is then aligned to the genome to find the matching gene sequence. Both SAGE and MPSS are sequencing based techniques that use tags to identify and count the mRNAs, but the biochemical manipulation and the sequencing approaches differ substantially between these techniques. Both methods are based on the principle that a short sequence tag contains sufficient information to uniquely identify a transcript, provided that the tag is obtained from a specific position within each transcript. In SAGE, short tags, usually 9-10 base pairs in length, are extracted from each mRNA, at a defined position. These tags are then linked together to form long serial molecules that can be cloned and sequenced. The quantification is performed by counting the number of times a specific tag is observed in the sequenced molecule. Finally, the tags are matched to the corresponding genes.

In MPSS, the extracted signatures are longer, 17-20 base pairs. Every one of these signatures is cloned into a vector and labeled with a unique 32 base pair oligonucleotide tag. The tag is then attached to one of millions of microbeads, by hybridization of the tag to a complementary sequence on the bead. The signatures on the microbeads are then sequenced and matched to the corresponding genes, and subsequently quantified by counting the

number of beads. The longer tag sequences, used in MPSS, provide higher specificity compared to SAGE. Another advantage of MPSS is the larger library size. One disadvantage that applies to both SAGE and MPSS is the loss of certain transcripts due to lack of restriction enzyme recognition sites, and ambiguity in tag annotation. Compared to microarray techniques, sequencing techniques, which are not based on hybridizations, give on the other hand a more exact quantitative value. This is because the number of transcripts is counted directly, instead of quantifying spot intensities which is prone to background noise. Another advantage is that the mRNA sequences do not need to be known beforehand, and therefore also previously unknown transcripts can be detected. Nevertheless, microarray experiments are much cheaper to perform and are therefore usually used in large scale experiments.

4.1 Microarray technology

The microarray technology has been around since the early 1990s, and during the last two decades the precision of the technology has increased considerably and, at the same time, the cost has decreased. Microarrays render the possibility to monitor the expression of thousands of genes simultaneously, which make them exceptionally useful for transcriptional studies on the global scale. Investigators are using the microarray technology to try to understand fundamental aspects of growth and development as well as to explore the underlying genetic causes of many human diseases. By monitoring the cells at various time points during a biological process or at specific biological conditions, one can take snapshots of the global transcriptional profile at different stages. The principle behind the microarray technology is base pairing of DNA/RNA. When two complementary sequences come together, such as the immobilized probe on the array and the mobile target in the sample, they will lock together (hybridize). The microarray consists of a surface on which millions of probes are immobilized. The surface is divided into features (locations), and each feature on the microarray has a superfluous number of probes that correspond to a specific transcript. When labeled target transcripts are hybridized onto the microarray, these bind complementary to their probes. The general procedure for performing a microarray experiment (which varies somewhat depending on the type of system) includes a series of steps. Initially, the RNA is reverse transcribed, usually to cDNA, and labeled with a fluorophore, and then the solution is hybridized onto the array. After the hybridization, the arrays are thoroughly washed, rinsed, and dried to remove non-hybridized transcripts from the surface. They are subsequently scanned to measure the fluorescence intensity for each spot on the array and these intensities are then translated into expression values. The spot intensities are directly proportional to the number of transcripts corresponding to each gene, and thus to the expression level of the gene.

4.2 Different types of microarrays

There are many different types of microarrays and the broadest distinction is whether the probes are spatially arranged on a slide made of glass, silicon or plastic or, if they are coded on microscopic polystyrene beads. They can be fabricated using different techniques, where the most common ones are robotic printing of the spots on the array (spotted arrays) or synthesis of the probes in situ using techniques such as photolithography. Moreover, the arrays vary in the way the signals are detected, and they are designed for hybridization of either one or two samples on the same array (one- or two-channel arrays) (Fig. 2).

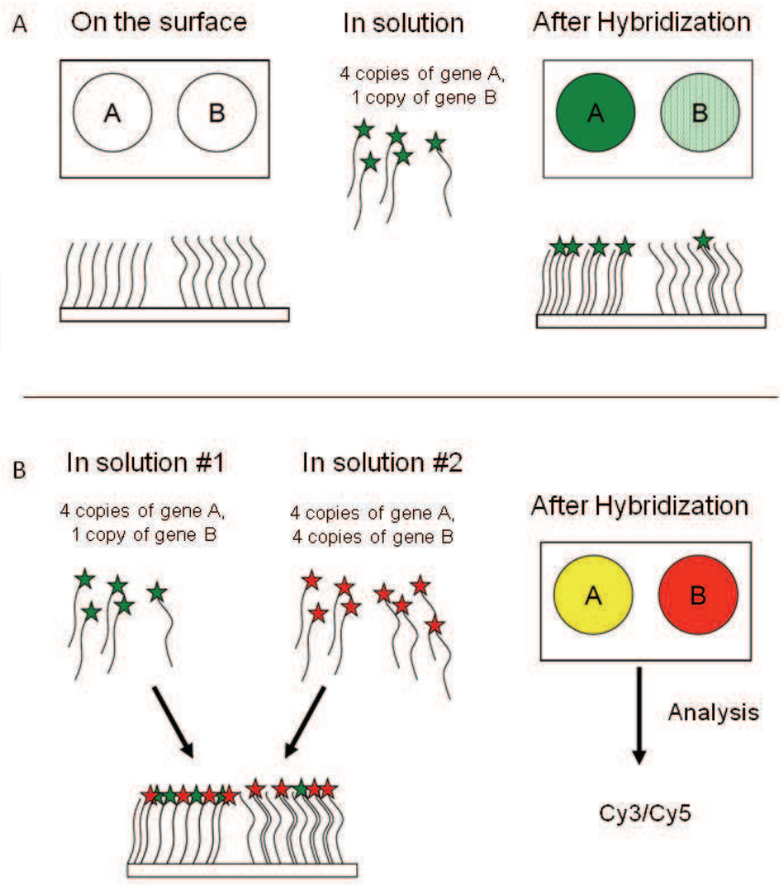


Fig. 2. Schematic picture of one-channel hybridization (A) and two-channel hybridization (B). Round-shaped features contain superfluous identical probes that hybridize with labeled targets from the samples. The shape of the features may vary between different microarray platforms. The intensity of the color is proportional to the number of probes that are hybridized to that feature. Yellow color means equal amounts of red and green labeled targets.

On one-channel arrays (also called oligonucleotide arrays) only one sample can be hybridized on each array, and the intensity levels are measured rather than the ratio between two intensities. Therefore, comparison of two conditions requires two separate single-dye hybridizations. On two-channel arrays two samples are labeled with two different fluorophores, typically Cy3 and Cy5, which have different fluorescence emission wavelengths. The two Cy-labeled cDNA samples are mixed and hybridized to a single microarray. Since the fluorophores have different excitation wavelengths it is possible to split the two signals during the scanning and calculate the intensities of each fluorophore, and use this in ratio-based analysis to identify up- and down-regulated genes. One benefit of one-channel arrays is that the data is more easily compared to data from different experiments, as long as batch effects have been accounted for. However, using the one-color system may require twice as many microarrays to compare samples within an experiment than with the two-color system. Depending on which system is used, the experimental design, and the generated data, the subsequent data analysis may differ.

4.3 Affymetrix microarrays

The Affymetrix platform is the most widely used commercial platform, providing a whole range of different types of arrays and covering various species. Affymetrix arrays are in situ synthesized, applying the photolithography technology to synthesize thousands to millions of 25-mer cDNA oligonucleotides in parallel. By using light-sensitive masking agents, a sequence is "built" one nucleotide at a time across the entire array. Typical for Affymetrix arrays are the multiple (11–20) probe pairs for each transcript. A new type of Affymetrix arrays which recently have entered the market is the Whole Transcript arrays, including both Gene ST 1.0 and Exon ST 1.0 arrays (Pradervand et al., 2008). The characteristics of these arrays are that they have an increased number of probes targeting exons along the whole transcript and not only in the 3′ end. The Gene ST 1.0 array has 1-2 probes per exon and the more comprehensive Exon ST 1.0 has four probes per exon.

4.4 Reliability and reproducibility of microarray data

The microarray technology has had tremendous impact on gene expression analysis during the last decade. However, publications of studies with dissimilar or even contradictory results have raised concerns regarding the reliability of this technology (Draghici et al., 2006, Kuo et al., 2002, Tan et al., 2003). For example, several global gene expression studies of stem cells have shown poor overlap (Fortunel et al., 2003, Ivanova et al., 2002, Ramalho-Santos et al., 2002). To address these and other concerns, such as performance and data analysis issues, the MicroArray Quality Control project (Chen et al., 2007, Shi et al., 2006) was initiated by the US Food and Drug Administration. Using an impressive number of laboratories, this comprehensive study showed both intra-platform consistencies across laboratories and a high level of inter-platform concordance in terms of genes identified as differentially expressed genes. Nevertheless, there are several issues to be aware of when using this technology, and which can introduce substantial biases in the final results. Examples of such issues to consider are:

- Cross-hybridization: There is a risk that some mRNAs may cross-hybridize probes on the array that are supposed to detect other mRNAs.
- Fold change compression: Due to various technical limitations, such as limited dynamic range and signal saturation, a certain level of FC compression is expected for array data compared to e.g. RT-PCR data (Wang et al., 2006, Yuen et al., 2002).
- Poor sensitivity for low expressed transcripts: Problems with relatively poor sensitivity in detecting small FCs have been reported for several microarray platforms (Wang et al., 2006).
- Cross-platform inconsistency: Inconsistent probe annotations across platforms, which leads to difficulties to ascertain that probes on various platforms aimed at the same gene do in fact quantify the same mRNA transcript (Draghici et al., 2006).
- Dye-biases: In two-channel systems the fluorescent dyes usually have different dynamic ranges and quantum yields, which is partially adjusted for by appropriate normalization, but may not be completely eliminated.
- Non-biological variations: There is always a risk that variations may be introduced during the experimental procedure (e.g. different persons performing the experiment, minor variations in temperature or duration) (Frantz, 2005) and these sometimes add substantial noise to the system. However, this source of variation is not unique to microarray experiments but is also an issue in other reverse transcription reactions (Wang et al., 2006).

5. Bioinformatic and statistical analysis

Large scale gene expression experiments generate enormous datasets that are computationally demanding to analyze. Therefore, gene expression analysis is one research area where bioinformatic methods have had important impact. These datasets are often challenging to analyze because of complex dependencies of interacting molecules and the data is often fragmented, incomplete, and noisy. Today, there are a lot of tools and software available, both commercially and open source, for solving various bioinformatic problems, such as identification of differentially expressed genes, clustering of data, and identification of interaction networks.

5.1 Data analysis of microarray data

The raw data from microarray experiments need to be pre-processed in several steps, before conducting any high level data analysis. Depending on the array type and the platform, these pre-processing steps vary, but basically involve subtraction of background and normalization for removal of non-biological variations. The data is also typically $\log_2$-transformed to achieve roughly normally distributed data, and potential outliers are excluded before performing the high level analysis. Due to the large amounts of data generated in microarray experiments, advanced bioinformatic algorithms (described below) are required for efficient interpretation of the data into valuable biological information. In the area of gene expression analysis there are e.g. algorithms for:

- identification of differentially expressed genes
- clustering of gene expression data
- pathway analysis
- derivation of protein interaction networks
- functional annotation of regulated genes

Although requiring some programming skills, the freely available R software environment[1] is highly recommended for various analyses of microarray data. This software has packages for normalization/standardization and statistical computing, as well as graphics. R can be used as a powerful standalone programming language, but the most prominent advantages are indeed all the implemented functions that are ready to use, and which make the R environment both flexible and extendible.

5.2 Identification of differentially expressed genes

There are several approaches on how to determine which genes that are differentially expressed in large gene expression datasets. Traditionally, researchers have applied different statistical tests and used the p-values as selection criteria. However, when dealing with microarray data one has to consider the *multiple testing problem*, as these datasets usually contain thousands of genes and the statistical test is applied for each of these genes. In other words, the multiple testing problem means that thousands of hypotheses are tested simultaneously, which leads to an increased chance of false positives. Individual p-values of a specified significance threshold no longer correspond to significant findings and thus, there is a need to adjust for multiple testing when assessing the statistical significance of genes in large microarray datasets. One such example of adjustment is Bonferroni

[1] http://www.r-project.org

correction, where the significance threshold is divided by the number of tests performed. However, this is often too strict criteria to be useful, and therefore other methods that are more adopted for analysis of microarray data have been developed. Examples of such methods are Significance Analysis of Microarray Data (SAM) (Tusher et al., 2001) and Empirical Bayes Analysis of Microarrays (EBAM) (Efron & Tibshirani, 2002), both included in the Siggenes package[2]. SAM and EBAM are statistical methods that controls for the false discovery rate (FDR), which is the fraction of false positives in the total set of genes that are selected as differentially expressed. The FDR is determined by using permutations of the repeated measurements to estimate the percentage of genes identified by chance. Although statistical methods usually are preferable since they provide a significance measure, the simple Fold Change (FC) method which calculates the ratio between two samples, are also frequently used. However, since FC provides no statistics regarding the significance of the results, combinations of FC and statistical methods are sometimes applied.

5.3 Clustering of gene expression data

To reduce the dimensionality and facilitate interpretation of microarray data one can apply different clustering techniques, such as hierarchical clustering, K-means (Świniarski et al., 1998), principle component analysis (PCA) (Jolliffe, 1986) or self-organizing maps (SOMs) (Tamayo et al., 1999), to group transcripts with similar transcriptional profiles. The purpose of clustering is to identify co-regulated and functionally related genes in large data sets. Clustering can also be used as a quality control tool to check the reproducibility of replicated arrays. For this purpose the agglomerative hierarchical clustering approach is most common. It starts with clusters containing a single item, and iteratively links and merges the two closest clusters together based on a distance measure. After each step, all the distances between the newly formed clusters are re-calculated. The output is a relationship tree (dendrogram) where the branches represent similarity and where the replicated arrays are expected to be grouped tightly together.

5.4 Pathway analysis

There are two main approaches for identification of pathways that are differentially expressed across various experimental conditions. These are Individual Gene Analysis (IGA) methods and Gene Set Analysis (GSA) methods (Nam & Kim, 2008). IGA is the most widely used approach and evaluates the significance of individual genes between two groups of compared samples. Methods using this approach typically yield a list of differentially expressed genes from a cutoff threshold, and evaluate this list for the enrichment of genes participating in specific pathways from a pathway database. A limitation with IGA approaches is that the final result is significantly affected by the selected threshold, which is often arbitrarily chosen. Notably, many genes with moderate, but biologically meaningful expression differences, are discarded by a strict cutoff threshold, which implies a reduction in statistical power. The GSA approach directly scores pre-defined pathways or gene sets based on differential expression, and specifically aims to identify pathways with subtle but coordinated expression changes that cannot be detected by IGA methods (Mootha et al., 2003, Nam & Kim, 2008). It is based on the principle that even weak expression changes for groups of related genes can have important effects. From

[2] http://www.bioconductor.org

a biological perspective, GSA methods are promising because functionally related genes often display coordinated expression (Nam & Kim, 2008). Various bioinformatic resources are available for carrying out pathway analysis such as WebGestalt[3] and DAVID[4] for the IGA approach, and GSEA[5] for the GSA approach.

5.5 Protein interaction networks

Protein–protein interactions are of central importance for virtually every biological process in a living cell. Typically, signal transduction, where mechanical/chemical stimuli to a cell are converted into specific cellular responses, plays a fundamental role in many biological processes and in many diseases. To investigate the putative interactions among proteins from the significantly up- or down-regulated genes identified from a biological experiment, protein interaction networks can be computationally generated by combining the experimental data with information from interaction databases. Several tools to aid derivation of protein interaction networks are available and currently one of the most comprehensive is STRING[6], which is a freely available database and web resource for experimentally determined and predicted protein-protein interactions (Jensen et al., 2009, von Mering et al., 2007). STRING includes both physical and functional interactions, and it weights and integrates information from numerous sources, including experimental repositories, computational prediction methods and public text collections. Thus, STRING is acting as a meta-database that maps all interaction evidence into a common set, which is then graphically visualized in a protein interaction network (Fig. 3).

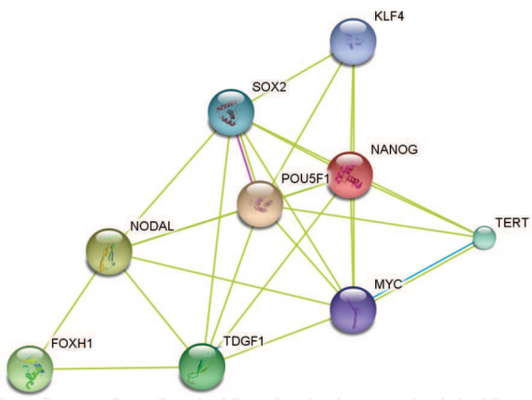


Fig. 3. Example of a protein interaction network generated by STRING, using pluripotency marker genes as input.

5.6 Functional annotation of differentially expressed genes

To increase our understanding of the biological properties of differentially expressed genes and further explore their functional properties, one can also use annotation information from Gene Ontology (Ashburner et al., 2000) which has annotation terms describing the

[3] http://bioinfo.vanderbilt.edu/webgestalt/

[4] http://david.abcc.ncifcrf.gov/

[5] http://www.broadinstitute.org/gsea/

[6] http://string-db.org/

genes or gene products. Gene Ontology consists of three categories of annotation terms, Biological Processes, Molecular Functions and Cellular Components. By comparing with a reference list, overrepresentation of annotations among sets of genes can be calculated by dividing the observed number of genes holding a specific annotation with the expected number of genes with that annotation. A common tool for Gene Ontology enrichment analysis is AmiGO[7], but also e.g. WebGestalt[8] and DAVID[9] provide similar functions.

6. Problems to address with transcriptional profiling

Presently, one of the main bottlenecks in stem cell research is insufficient yield and purity of the final cell preparations and immature phenotypes of the differentiated cells. There are substantial gaps in our understanding of the molecular programs that govern early cellular differentiation and maturation, which limits the possibilities to improve the culturing protocols. In order to fully realize the potential of hESCs, a better understanding of the regulatory mechanisms that control their differentiation towards specific lineages is needed. An extensive characterization of hESCs and their derivates is also urgently required. To address these challenges, information is needed about which genes that are differentially expressed at specific stages during differentiation. Some of these are likely candidate genes in regulatory mechanisms, important for transitions of cells between different developmental stages. We also need precise characterization methods to determine the identity of cells during the differentiation process, and therefore more reliable marker genes for specific developmental stages are required. In next chapter we describe how microarray technology can been applied to analyze global gene expression patterns in hESCs and their differentiated progenies. Below are some examples studies, with focus on cardiac- and hepatic lineages, described in more detail.

7. Examples of transcriptional profiling of hESCs and their derivatives

Several global gene expression studies have been conducted on hESCs and differentiated progenies thereof. Depending on the research questions addressed, these studies have had different experimental designs and they have used different analysis approaches.

7.1 Transcriptional profiling of hESC-derived cardiomyocytes

The first example is a study where hESC-derived cardiomyocyte clusters (CMCs) were characterized at the gene expression level and their global transcriptional pattern was investigated (Synnergren et al., 2008). This required only a rather simple design with not more than two groups to compare, undifferentiated (UD) hESCs and hESC-derived CMCs. The material consisted of one pooled sample of UD hESCs and two different biological replicates of pooled hESC-derived CMCs, harvested at a number of time points up to 22 days after initiation of differentiation (Fig. 4).

[7] http://amigo.geneontology.org/cgi-bin/amigo/go.cgi
[8] http://bioinfo.vanderbilt.edu/webgestalt/
[9] http://david.abcc.ncifcrf.gov/

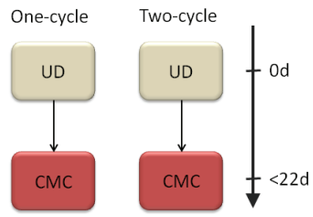


Fig. 4. Experimental design where two different groups (UD and CMC) were included and the experiment was repeated twice using one- and two-cycle amplification respectively.

The hESC line SA002 (Cellartis AB, Göteborg) was used in this experiment. Due to technical issues, two separate sets of microarray experiments were conducted. In the first, one-cycle amplified RNA was used, while in the second set of experiments two-cycle amplified RNA was used due to the limited amount of available RNA. Even though no obvious differences between the two data sets could be observed, all subsequent calculations between samples were conducted within each experiment separately. The quality of the RNA and cRNA, labeled by in vitro transcription, was tested and the fragmented cRNA was then hybridized to the microarrays. Each sample was hybridized to duplicate arrays from the Affymetrix microarray platform (GeneChip 133 Plus 2.0) (Affymetrix, Santa Clara, CA). Extraction of expression values and scaling of data was performed using the MAS5 algorithm and transcripts flagged as 'Absent' on all arrays were filtered before the data analysis. The SAM statistical algorithm was used to identify significantly up- and down-regulated genes between the CMCs compared to UD cells. In total 530 genes were identified as up-regulated and 40 genes were down-regulated in the CMCs (Synnergren et al., 2008). These sets of regulated genes were further analyzed using various bioinformatic tools. To further explore the biology of the significantly up-regulated genes in hESC-derived CMCs, Gene Ontology annotations were used to group the genes according to biological process, molecular function, and cellular component. To investigate possible interactions among proteins from the significantly up-regulated genes in hESC-derived CMCs, the search tool STRING was applied to derive protein interaction networks. These networks were used to identify hub-proteins, with many interactions to other genes. Moreover, differentially expressed pathways were assessed in the CMCs using the WebGestalt tool.

7.2 Investigation of putative correlation between mRNA and miRNA expression

This second example describes how transcriptional profiling can be applied to further our understanding of the regulatory mechanisms of transcription and translation, by investigation of putative correlation between mRNA and miRNA expression. Thus, mRNA and miRNA microarray experiments were designed where matched samples from hESCs and hESC-derived CMCs were collected for global mRNA and miRNA profiling. Using cell line SA002 (Cellartis AB, Göteborg) total RNA was extracted with a method which preserves small molecules. The RNA was split into two aliquots, and microarray experiments were conducted in parallel to measure both miRNA and mRNA expression of paired samples.

As illustrated in Fig. 5, the material consisted of samples of UD cells and hESC-derived CMCs, cultured for 3 (CMC3w) and 7 weeks (CMC7w) after onset of differentiation. In addition, samples from fetal heart (FH) and adult heart (AH) were included as reference material. Significantly up- or down-regulated miRNAs and mRNAs were identified using the SAM statistical algorithm. The differentially expressed miRNAs were grouped

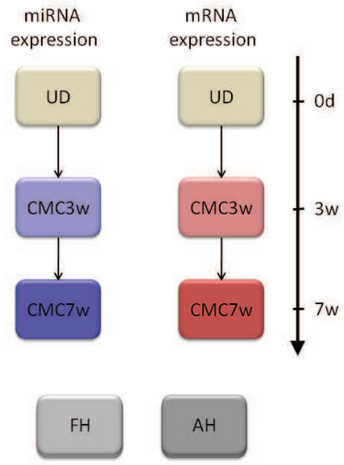


Fig. 5. Experimental design of a parallel miRNA and mRNA study. Three time points (UD, CMC 3 weeks, and CMC 7 weeks) were analyzed and fetal heart (FH) and adult heart (AH) were included as reference samples.

according to their expression profiles using hierarchical clustering. Putative target genes of up- and down-regulated miRNAs were predicted using the tool microT[10], and to investigate if the predicted target genes are present in the set of differentially expressed mRNAs, the overlap between predicted target genes and differentially expressed mRNAs were calculated. To further explore the molecular biology of the regulated target genes these were also analyzed for enrichment of Gene Ontology annotations related to cardiac development.

7.3 Transcriptional comparison between two endoderm differentiation protocols

The final example describes a transcriptional comparison between hESCs differentiated through the endoderm, either definitive endoderm (DE) or primitive endoderm (PrE), as well as a global transcriptional characterization of endoderm, hepatocyte progenitors, and hepatocyte-like cells (Synnergren et al., 2010a). A comprehensive experimental design was applied in this work including three cell lines (SA002, SA167, and SA461) and four time points, as well as the two separate differentiation protocols (Fig. 6).

The hepatocellular carcinoma cell line (HepG2) was included as a reference sample in the experiment. Each sample was cultured and harvested in biological duplicates. The RNA was extracted and assessed for quality before generation of cRNA, and subsequently hybridized to the arrays. The raw data was extracted and normalized using MAS5 and filtered and $\log_2$ transformed before subsequent data analysis. In this experiment, the FC method was applied to identify differentially expressed genes at various stages. At the early time points these cells showed relatively large variations in the magnitude of up- or down-regulation of genes across the cell lines. However, the trend of the regulation was consistent in all three cell lines and therefore an interesting observation. In such situation, the statistical methods have problems to select genes of real biological interest, and consequently FC may be a preferable alternative. The DE differentiation pathway to generate hepatocytes mimics the development of hepatocytes in the embryo and the expression profiles in these samples

[10] http://diana.cslab.ece.ntua.gr/microT/

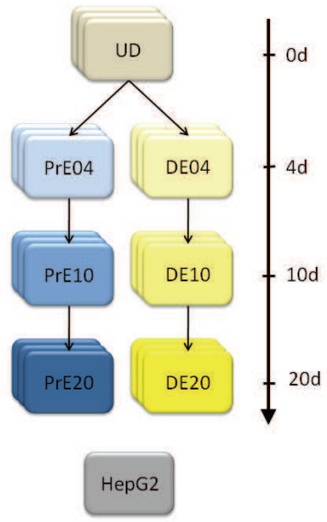


Fig. 6. Experimental design of the protocol comparison experiment. Four time points (UD, 4 days, 10 days, and 20 days) and two differentiation protocols (PrE and DE) were included in the experiment which was run in duplicates and repeated using three different cell lines. HepG2 was included as a reference sample in the study.

were therefore characterized in more detail. The genes that showed up-regulation in DE20 samples compared to UD samples were further analyzed for enrichment of Gene Ontology annotations and differentially expressed pathways were also identified using the DAVID bioinformatic resource. Moreover, protein interaction networks were derived using the STRING tool and based in these, hub proteins were identified among the up-regulated genes in DE20 samples.

8. Results and discussion

Extensive characterization of hESC-derived functional cell types, such as cardiomyocytes and hepatocytes has been performed by several investigators, with the purpose to explore the transcriptional programs that are activated during differentiation along these specific lineages. Results from these studies have identified large sets of genes that showed differential expression at specific developmental stages. The sets of regulated genes have been further explored by various bioinformatic analyses, to learn more about the molecular functions of the differentially expressed genes and understand their biology. Here we discuss results from studies on cardiomyocyte and hepatocyte differentiation, performed by us and others, and report on genes and pathways that showed up- or down-regulation particularly during these processes.

8.1 Molecular signature of hESC-derived cardiomyocyte clusters

Despite the substantial progress made by different investigators during recent years, the understanding of the molecular signature of hESC-derived cardiomyocytes (CMs) and the factors that induce cardiogenesis during embryonic development still remains limited. However, important knowledge about the transcriptional program that is activated during CM differentiation has been gained through transcriptional profiling of these cells

(experimental set-up described in section 7.1). Selected colonies of hESC-derived contracting clusters of CMs were manually dissected, and pooled for subsequent microarray analysis. These samples were compared with samples from UD hESCs. In total 530 up-regulated and 40 down-regulated genes were identified in the CMCs (Synnergren et al., 2008). Among the up-regulated genes, there were several that have been used before to characterize hESC-derived CMs e.g., MYH6, MYH7, PLN, TNNT2, NPPA, GATA4, and MEF2C (Kehat et al., 2001, McDevitt et al., 2005). The functional properties of the up-regulated genes in the hESC-derived CMCs were further investigated, using available Gene Ontology annotations. Among the enriched annotations were 'muscle contraction', 'development of mesoderm and muscle', 'cellular differentiation', 'calcium ion binding', and 'tropomyosin binding'. Moreover, several induced cellular pathways where identified, that may be important for cardiogenic induction of hESCs but also for sustaining the CM phenotype. Results reported by other investigators show high overlap with our data, despite the fact that direct comparisons of results between different microarray studies are sometimes difficult to make. Different experiments often have major discrepancies in differentiation models, microarray platforms, cell lines used, and experimental set-ups, which partly may explain observed problems with poor overlap between published results from different stem cell studies (Fortunel et al., 2003, Ivanova et al., 2002, Ramalho-Santos et al., 2002).

To overcome some of these problems one should preferably re-analyze the data from the raw data files from each experiment, using a consistent data mining approach for all the data sets, when comparing results from multiple microarray studies. However, as an alternative, we have compared published lists of significantly enriched genes from similar studies, to explore the overlap of differentially expressed genes during CM differentiation. Importantly, in addition to the above mentioned challenges when comparing microarray data from different experiments, the final cell populations that have been analyzed in these studies differ in their composition (Cao et al., 2008, Xu et al., 2009). Nevertheless, and in contrast to previous findings (Fortunel et al., 2003, Ivanova et al., 2002, Ramalho-Santos et al., 2002), notable similarities were identified across results from the four global expression studies that so far have been published on hESC-derived CM-like cells (Beqqali et al., 2006, Cao et al., 2008, Synnergren et al., 2008, Xu et al., 2009). Importantly, this strengthens the reliability of the microarray technology and verifies that hESC-derived CMs express a uniform transcriptional profile, despite different cell lines and major differences in how these cells are derived.

Comparing results from our data (Synnergren et al., 2008) with the study performed by Beqqali et al. (Beqqali et al., 2006), where hESC-derived CMs were generated by co-culture with END-2 cells (Passier et al., 2005), 15 genes were reported as enriched in their hESC-derived CMs and in fetal heart tissue. Notably, eight (53%) of these genes are also up-regulated in our hESC-derived CMs (e.g. TNNT2, PLN, and MYL7). Another study published on hESC-derived CMs (Cao et al., 2008) report on analyses made on material from hESCs, hESC-derived beating embryoid bodies (EBs), hESC-derived CMs which were percol purified to 40-45% CMs, and purified CMs from fetal heart (FH) tissue samples. Notably, their study focused on transitions from one stage to the next one, and consequently they compared hESCs-EBs, EBs-CMs, and CMs-FH. In our work we compared hESCs with hESC-derived CM clusters and the corresponding direct comparison of CMs and hESCs was not done by Cao et al. (Cao et al., 2008) which hampers the comparison of our results with theirs. Nevertheless, we found that 33% of our up-regulated genes in the CMCs were in their study identified as enriched already at the EB stage. Five of our up-regulated genes in

the CMCs (EPAS1, ITGB3, PLD1, MSRB3, EMP1) were significantly enriched in FH compared to the CM-sample, and six of our genes that were enriched in the CMCs (CLIC5, RUNX1, COL8A1, LONRF2, MSRB3, CAV2) were up-regulated in CMs compared to EBs. A similar comparison was made regarding the repressed genes across these two studies and 17 (43%) of the 40 significantly down-regulated genes in our data were already repressed at the EB stage in Cao et al., and one gene was among the significantly down-regulated genes between FH and CMs. Again, no comparison was made between CMs and hESCs (Cao et al., 2008) regarding down-regulated genes, but such a comparison is anticipated to generate a higher overlap with our list of genes that were down-regulated in CMCs.

The most recent work on global gene expression of hESC-derived CMs used a transgenic cell line with a construct comprising the CM-restricted α-myosin heavy chain (α-MHC) promoter (Xu et al., 2009). They applied antibiotic selection to purify their population of hESC-derived CMs and achieved a 99% pure population. Fetal and adult heart tissue were used as reference samples but notably, these samples were not purified, but contained a mixture of the cell types present in heart tissue. Despite substantial differences, such as different cell lines, differentiation protocols, purity of CMs, sampling day etc, a prominent overlap was observed between our data and the data from Xu and colleagues (Xu et al., 2009). In total 147 (27%) of the 540 genes that were up-regulated in our data were also identified as significantly up-regulated in their population of CMs, when compared to UD and EB samples. Remarkably, 115 (78%) of these 147 genes also show up-regulation in the FH and AH samples in data from Xu et al. Strikingly, a subset of 57 genes that show up-regulation in our hESC-derived CM clusters is also overlapping with the up-regulated genes both in Cao et al. (Cao et al., 2008) and in Xu et al (Xu et al., 2009). All of these 57 genes also show significantly up-regulation in FH and AH. Furthermore, three (RBM24, TCEA3, and FHOD3) of the four novel candidate cardiac markers, which by Xu and co-workers were validated by *in situ* hybridization during early mouse development, were indeed significantly up-regulated in our study of hESC-derived CMCs (Synnergren et al., 2008). The fourth one (C15orf52) was not present on the arrays we used. Interestingly, TCEA3 is also among the 57 genes that overlapped across all three studies (Cao et al., 2008, Synnergren et al., 2008, Xu et al., 2009). Taken together, this suggests that there are substantial similarities between the CM cell populations obtained from hESCs, independent of differentiation protocols and cell lines used. The results summarized here provide valuable information about the molecular program that is active in hESC-derived CMs.

However, to further analyze the regulatory mechanisms that may control CM differentiation, an additional level of gene regulation have also been explored (experimental set-up described in section 7.2) and hESC-derived CMCs have been characterized with respect to their miRNA expression. Global microarrays were employed to measure the expression of both miRNA and mRNA in parallel in samples of CMC, harvested at two different time points, 3 weeks and 7 weeks after onset of differentiation, as well as in UD cells and in fetal and adult heart tissue samples. Differentially expressed miRNAs and mRNAs were identified in these datasets, by using UD cells as control sample. Notably there were more than twice as many up-regulated than down-regulated miRNAs in the samples of CMCs, indicating the importance of increased expression of specific miRNAs during cardiac development (Synnergren et al., 2010b). Furthermore, we also identified more differentially expressed miRNAs (both up- and down-regulated) in the CMC samples than in the fetal and adult heart tissue samples. To define a set of miRNAs of putative importance in cardiac-like cells, differentially expressed miRNAs in samples from CMC and

in samples from fetal and adult heart were compared, and an overlap of regulated miRNAs in all four samples was identified. In total 61 up-regulated and 24 down-regulated miRNAs were identified when investigating all four samples (Synnergren et al., 2010b). Moreover, possible correlations between differentially expressed miRNAs and mRNAs were investigated, by first conducting computational predictions for the differentially expressed miRNAs, and then determining putative concordance in miRNA expression and mRNA levels of the predicted target genes. Interestingly, a correlation between the global miRNA expression and corresponding target mRNA expression was observed. To further explore the biology of the predicted target genes of the differentially expressed miRNAs, enrichment of Gene Ontology annotations was determined and strikingly, several of the over-represented annotations relate to cardiac function and cardiac development. A number of induced cellular pathways were also identified among the predicted target genes, and several of these have been demonstrated to be important in cardiac development or functions e.g. 'NFAT and Hypertrophy of the heart', 'Wnt signaling pathway' and 'Calcium signaling pathway'. Results from this analysis provide an excellent starting point for further studies regarding the functional properties of the differentially expressed miRNAs in the context of cardiogenesis and regeneration of cardiac tissue.

Several other studies have also identified a number of miRNAs that are likely to play key roles during heart development and in cardiac function (Cordes & Srivastava, 2009, Divakaran & Mann, 2008, Sartipy et al., 2009, Thum et al., 2008, van Rooij & Olson, 2007, Zhang, 2008). It was recently demonstrated that miR-1 reinforces the expression of one of the earliest cardiac markers, NKX2.5, in both murine and human ESC lines and that it increases the fraction of contracting CMs compared to control samples (Ivey et al., 2008). Another group of miRNAs that are expressed during the stem cell state and progressively declines during differentiation are the nearly identical miRNAs miR-302a-d, collectively referred to as miR-302 (Rosa et al., 2009). By controlling the germ layer specification and promote mesendodermal fate specification while inhibiting neuroectoderm formation, it is suggested that miR-302 has a crucial role in embryogenesis (Rosa et al., 2009). In line with these reports, our data confirms that several variants of miR-302 are highly expressed in UD cells, and a substantial down-regulation is also observed in differentiated progenies as well as in fetal and adult heart tissue. Moreover, we identified miR-208a/b and miR-499 as significantly induced in all cardiac-like samples. Interestingly, these miRNAs are also reported as enriched in cardiac tissue by others (Adachi et al., 2010, Ji et al., 2009, Sluijter et al., 2010). Both miRNA-499 and miRNA-1 are suggested to regulate the proliferation of human CM progenitors and their further differentiation into CMs (Sluijter et al., 2010). In addition, miR-499 is also proposed as a marker for acute myocardial infarction in humans (Adachi et al., 2010). Moreover, miR-208 is suggested as a marker for myocardial injury in rat (Ji et al., 2009), and as a regulator of cardiac hypertrophy in mice (Callis et al., 2009).

Together with our results, these reports emphasize the importance of miRNAs for cardiac development as well as potentially useful markers in clinical applications. In the future, some of these miRNAs may also serve as prospective drug targets in various cardiac injuries. There is accumulating evidence supporting the importance of miRNAs for hESC self-renewal, pluripotency, and differentiation. Determining miRNAs that are associated with re-programming will yield significant insight into the specific miRNA expression patterns that are required for pluripotency. To further investigate which miRNAs that are associated with re-programming, investigators have now started to characterize the miRNA

expression during the re-programming of iPS cells (Wilson et al., 2009). Interestingly, result shows that miR-302 is up-regulated in both hESCs and in iPS cells.

8.2 Transcriptional profiling of hESCs differentiating to definitive and primitive endoderm and further towards the hepatic lineage

The endoderm lineage can be subdivided into the DE which further develops into liver, pancreas and lung, and the PrE which develops into the yolksack, where it forms the placenta. To explore transcriptional differences between these two subtypes, global gene expression data from DE and PrE differentiation was analysed and the transcriptional patterns in these two cell lineages were compared with UD cells, as well as with control samples from HepG2 (experimental set-up described in section 7.3). Interesting differences and similarities were identified between these two endodermal subtypes. We also thoroughly characterized the DE-derivatives, by identifying up- and down-regulated genes in each of the three differentiation time points 4 days (DE), 10 days (DE-Prog) and 20 days (DE-Hep). In total, we identified 167, 439, and 921 transcripts which were significantly up-regulated in DE, DE-Prog, and DE-Hep, respectively, when compared to UD samples. None of these transcripts were significantly enriched in the PrE derivatives. Well-known markers for DE, such as SOX17, CXCR4, CER1 and GSC, showed a distinct peak of expression in the DE time point in all the three investigated cell lines. At the final time point, when using the DE differentiation protocol, several genes known to be expressed in mature hepatocytes, e.g. ALB, DPP4, SERPINA7, TF, TM4SF1 and UBD (Chiao et al., 2008, Hay et al., 2008, Kon et al., 2006), showed increased mRNA levels. Notably, also CD44, known to be expressed in hepatocyte progenitors (Kon et al., 2006), showed high expression at 20 days, which indicates that the DE-Hep have an immature phenotype or also contain a fraction of hepatocyte progenitors. Interestingly, ALB, which is a well-known marker for mature hepatocytes (Duncan, 2003), showed 3-11 times higher expression in DE-Hep than in the corresponding PrE-derivatives, and the expression of ALB was about 1,000-fold higher in the DE-Hep than in the UD samples. Of special note from our results are the less reported genes TM4SF1 and UBD, which demonstrate highly interesting expression patterns. Their expression patterns were consistent across all three cell lines, indicating a putative importance of controlled expression during hepatocyte differentiation. Both these genes demonstrate increasing expression levels during the differentiation towards hepatocytes, with the peak expression in the most mature samples. Interestingly, and consistent with our observations, these genes have also previously been reported as reliable hepatocyte markers (D´Amour et al., 2007). Another observation from our results is that the global transcriptional activity increases dramatically as the cells differentiate, with a larger number of differentially expressed genes in the more mature samples.

To the best of our knowledge, besides our work (Synnergren et al., 2010a), global transcriptional profiling of hESC-derived hepatocytes has only been reported once (Chiao et al., 2008), in which AFP^+ cells were selected for gene expression analysis. However, the data analysis approach used in that study differs from ours and instead of generating lists of up- vs. down-regulated genes during hepatocyte differentiation, they applied Gene Set Enrichment Analysis (Mootha et al., 2003) to identify affected sets of genes during hepatic specification, e.g. molecular pathways etc. An overlap comparison of lists of enriched genes in different stages is therefore not possible. However, when investigating the expression of 30 marker genes, which specifically were reported as enriched in their AFP^+ population, we observed important similarities between the data sets. Examples of genes that are enriched

in the early DE stage in both studies are typical DE-markers such as SOX17, FOXA2 and MIXL1. Moreover, AFP, which is expressed during hepatocyte specification but also at very early stages of PrE, is up-regulated accordingly in our data. Additionally, CDH17 and KRT7, which are reported as expressed during hepatocyte specification (Chiao et al., 2008) are induced in our hepatocyte-like cells, which might indicate that this population also contains more immature cells. Also other genes such as ALB, NTN4, MET, and CEBPA, which are known to be expressed in mature hepatocytes, show induced expression patterns in both studies.

9. Conclusion and future perspectives

Human ESCs have a tremendous potential in many different applications such as drug development, regenerative medicine, and as a model system in basic research. However, to fully utilize the potential of these cells we need to better understand the regulatory mechanisms that control the differentiation of hESC into various functional cell types. Transcriptional profiling is a powerful approach to learn more about the global transcriptional pattern in differentiating hESC. In this chapter we have reviewed studies on global transcriptional profiling in hESCs and differentiated progenies of the cardiac- and hepatic lineages. The comprehensive datasets generated from these studies are especially useful for characterization purposes, identification of differentially expressed genes, and investigation of gene regulation. We, and others, have performed transcriptional profiling of hESC-derived CMs and hepatocytes, and sets of significantly up- and down-regulated genes in different stages during the differentiation processes have been identified. Interesting transcriptional patterns have been revealed, which confirms the expression of known marker genes as well as identifies large sets of novel genes that are differentially expressed at particular stages during the differentiation. The expected expression patterns of known marker genes is important information to confirm the efficiency of the differentiation, but most interesting are the novel sets of genes that not previously have been associated with differentiation or developmental processes. It is hypothesized that many of these genes have the potential to serve as novel markers and provide important information for optimization of the differentiation protocols, and further validation and investigation of their functional properties is therefore suggested.

Less studied than mRNA expression is the miRNA expression, which has shown to be important in stem cell specification. These tiny molecules provide an additional level of gene regulation by fine-tuning of the mRNA expression. An interesting approach to identify putative correlations between these interacting molecules is to explore the miRNA and mRNA expression in parallel, which has revealed interesting result when applied on data from hESC-derived CMs. MicroRNA expression may also provide a novel characterization of hESC-derivatives and results from studies on CMs show that specific miRNAs may serve as important complementary markers for this specific cell type. However, additional studies on the miRNA expression in other derivatives are needed to be able to compare the expression in different cell types, and ideally identify lineage specific miRNA patterns.

During the last decade the microarray technology has improved considerably and the recent development of whole transcript arrays has rendered the possibility to also explore different splicing variant genes. This is an important step forward since it is anticipated that more than half of all genes undergo alternative splicing. However, the data analysis of exon data sets are demanding and tools for proper interpretation of such data is still rare and

insufficient. Nevertheless, splicing variations are believed to be important in many disease and their function in stem cell differentiation remains to be investigated. The most recent progress in the field of human pluripotent stem cells is successful re-programming of differentiated somatic cells into induced pluripotent stem (iPS) cells (Takahashi et al., 2007). Human iPS-cells are generated from somatic cells by over-expression of specific factors and these cells share many characteristics of hESCs, including multi-lineage differentiation potential and infinitive proliferation capabilities *in vitro*. However, extensive characterization of the iPS cells and their differentiated progenies at the transcriptional level is still lacking, but it is nevertheless required to be able to assess their similarity to hESCs.
Taken together, transcriptional profiling of hESC and their derivatives offers great possibilities to explore global expression patterns that are activated during hESC differentiation, and provide a foundation for further dissection of the molecular mechanisms that control the stem cell specification.

10. References

Adachi, T., Nakanishi, M., Otsuka, Y., Nishimura, K., Hirokawa, G., Goto, Y., Nonogi, H. & Iwai, N. (2010) Plasma microrna 499 as a biomarker of acute myocardial infarction. *Clin Chem,* 56, 7**,** 1183-5.

Alberts, B., Bray, D., Hopkin, K., Johnson, A., Lewis, J., Raff, M., Roberts, K. & P, W. (2004) *Essential cell biology,* (Second edn) (Garland Science, Taylor & Francis Group, New York).

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. & Sherlock, G. (2000) Gene ontology: Tool for the unification of biology. The gene ontology consortium. *Nat Genet,* 25, 1**,** 25-9.

Bartel, D. P. (2004) Micrornas: Genomics, biogenesis, mechanism, and function. *Cell,* 116, 2**,** 281-97.

Beqqali, A., Kloots, J., Ward-Van Oostwaard, D., Mummery, C. & Passier, R. (2006) Genome-wide transcriptional profiling of human embryonic stem cells differentiating to cardiomyocytes. *Stem Cells,* 24, 8**,** 1956-67.

Callis, T. E., Pandya, K., Seok, H. Y., Tang, R. H., Tatsuguchi, M., Huang, Z. P., Chen, J. F., Deng, Z., Gunn, B., Shumate, J., Willis, M. S., Selzman, C. H. & Wang, D. Z. (2009) Microrna-208a is a regulator of cardiac hypertrophy and conduction in mice. *J Clin Invest,* 119, 9**,** 2772-86.

Cao, F., Wagner, R. A., Wilson, K. D., Xie, X., Fu, J. D., Drukker, M., Lee, A., Li, R. A., Gambhir, S. S., Weissman, I. L., Robbins, R. C. & Wu, J. C. (2008) Transcriptional and functional profiling of human embryonic stem cell-derived cardiomyocytes. *PLoS ONE,* 3, 10**,** e3474.

Chen, J. J., Hsueh, H. M., Delongchamp, R. R., Lin, C. J. & Tsai, C. A. (2007) Reproducibility of microarray data: A further analysis of microarray quality control (maqc) data. *BMC Bioinformatics,* 8**,** 412.

Chiao, E., Elazar, M., Xing, Y., Xiong, A., Kmet, M., Millan, M. T., Glenn, J. S., Wong, W. H. & Baker, J. (2008) Isolation and transcriptional profiling of purified hepatic cells derived from human embryonic stem cells. *Stem Cells,* 26, 8**,** 2032-41.

Cordes, K. R. & Srivastava, D. (2009) Microrna regulation of cardiovascular development. *Circ Res,* 104, 6**,** 724-32.

D´Amour, K., Baetge, Emmanuel (2007) Hepatocytes lineage cells. US Patent WO 07127454.

Divakaran, V. & Mann, D. L. (2008) The emerging role of micrornas in cardiac remodeling and heart failure. *Circ Res,* 103, 10**,** 1072-83.

Draghici, S., Khatri, P., Eklund, A. C. & Szallasi, Z. (2006) Reliability and reproducibility issues in DNA microarray measurements. *Trends Genet,* 22, 2**,** 101-9.

Duncan, S. A. (2003) Mechanisms controlling early development of the liver. *Mech Dev,* 120, 1**,** 19-33.

Efron, B. & Tibshirani, R. (2002) Empirical bayes methods and false discovery rates for microarrays. *Genet Epidemiol,* 23, 1**,** 70-86.

Fortunel, N. O., Otu, H. H., Ng, H. H., Chen, J., Mu, X., Chevassut, T., Li, X., Joseph, M., Bailey, C., Hatzfeld, J. A., Hatzfeld, A., Usta, F., Vega, V. B., Long, P. M., Libermann, T. A. & Lim, B. (2003) Comment on " 'stemness': Transcriptional profiling of embryonic and adult stem cells" And "A stem cell molecular signature". *Science,* 302, 5644**,** 393; author reply 393.

Frantz, S. (2005) An array of problems. *Nat Rev Drug Discov,* 4, 5**,** 362-3.

Gregory, P. A., Bert, A. G., Paterson, E. L., Barry, S. C., Tsykin, A., Farshid, G., Vadas, M. A., Khew-Goodall, Y. & Goodall, G. J. (2008) The mir-200 family and mir-205 regulate epithelial to mesenchymal transition by targeting zeb1 and sip1. *Nat Cell Biol,* 10, 5**,** 593-601.

Hay, D. C., Fletcher, J., Payne, C., Terrace, J. D., Gallagher, R. C., Snoeys, J., Black, J. R., Wojtacha, D., Samuel, K., Hannoun, Z., Pryde, A., Filippi, C., Currie, I. S., Forbes, S. J., Ross, J. A., Newsome, P. N. & Iredale, J. P. (2008) Highly efficient differentiation of hescs to functional hepatic endoderm requires activina and wnt3a signaling. *Proc Natl Acad Sci U S A,* 105, 34**,** 12301-6.

Ivanova, N. B., Dimos, J. T., Schaniel, C., Hackney, J. A., Moore, K. A. & Lemischka, I. R. (2002) A stem cell molecular signature. *Science,* 298, 5593**,** 601-4.

Ivey, K. N., Muth, A., Arnold, J., King, F. W., Yeh, R. F., Fish, J. E., Hsiao, E. C., Schwartz, R. J., Conklin, B. R., Bernstein, H. S. & Srivastava, D. (2008) Microrna regulation of cell lineages in mouse and human embryonic stem cells. *Cell Stem Cell,* 2, 3**,** 219-29.

Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., Bork, P. & Von Mering, C. (2009) String 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res,* 37, Database issue**,** D412-6.

Ji, X., Takahashi, R., Hiura, Y., Hirokawa, G., Fukushima, Y. & Iwai, N. (2009) Plasma mir-208 as a biomarker of myocardial injury. *Clin Chem,* 55, 11**,** 1944-9.

Jolliffe, I. T. (1986) *Principal component analysis,* (Springer-Verlag).

Kehat, I., Kenyagin-Karsenti, D., Snir, M., Segev, H., Amit, M., Gepstein, A., Livne, E., Binah, O., Itskovitz-Eldor, J. & Gepstein, L. (2001) Human embryonic stem cells can differentiate into myocytes with structural and functional properties of cardiomyocytes. *J Clin Invest,* 108, 3**,** 407-14.

Kon, J., Ooe, H., Oshima, H., Kikkawa, Y. & Mitaka, T. (2006) Expression of cd44 in rat hepatic progenitor cells. *J Hepatol,* 45, 1**,** 90-8.

Kuo, W. P., Jenssen, T. K., Butte, A. J., Ohno-Machado, L. & Kohane, I. S. (2002) Analysis of matched mrna measurements from two different microarray technologies. *Bioinformatics,* 18, 3**,** 405-12.

Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B. L., Mak, R. H., Ferrando, A. A., Downing, J. R., Jacks, T., Horvitz, H. R. & Golub, T. R. (2005) Microrna expression profiles classify human cancers. *Nature,* 435, 7043**,** 834-8.

Mcdevitt, T. C., Laflamme, M. A. & Murry, C. E. (2005) Proliferation of cardiomyocytes derived from human embryonic stem cells is mediated via the igf/pi 3-kinase/akt signaling pathway. *J Mol Cell Cardiol,* 39, 6**,** 865-73.

Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov, J. P., Golub, T. R., Tamayo, P., Spiegelman, B., Lander, E. S., Hirschhorn, J. N., Altshuler, D. & Groop, L. C. (2003) Pgc-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet,* 34, 3**,** 267-73.

Nam, D. & Kim, S. Y. (2008) Gene-set approach for expression pattern analysis. *Brief Bioinform,* 9, 3**,** 189-97.

Passier, R., Oostwaard, D. W., Snapper, J., Kloots, J., Hassink, R. J., Kuijk, E., Roelen, B., De La Riviere, A. B. & Mummery, C. (2005) Increased cardiomyocyte differentiation from human embryonic stem cells in serum-free cultures. *Stem Cells,* 23, 6**,** 772-80.

Pradervand, S., Paillusson, A., Thomas, J., Weber, J., Wirapati, P., Hagenbuchle, O. & Harshman, K. (2008) Affymetrix whole-transcript human gene 1.0 st array is highly concordant with standard 3' expression arrays. *Biotechniques,* 44, 6**,** 759-62.

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

Rosa, A., Spagnoli, F. M. & Brivanlou, A. H. (2009) The mir-430/427/302 family controls mesendodermal fate specification via species-specific target selection. *Dev Cell,* 16, 4**,** 517-27.

Sartipy, P., Olsson, B., Hyllner, J. & Synnergren, J. (2009) Regulation of 'stemness' and stem cell differentiation by micrornas. *IDrugs,* 12, 8**,** 492-6.

Shi, L., Reid, L. H., Jones, W. D., Shippy, R., Warrington, J. A., Baker, S. C., Collins, P. J., De Longueville, F., Kawasaki, E. S., Lee, K. Y., Luo, Y., Sun, Y. A., Willey, J. C., Setterquist, R. A., Fischer, G. M., Tong, W., Dragan, Y. P., Dix, D. J., Frueh, F. W., Goodsaid, F. M., Herman, D., Jensen, R. V., Johnson, C. D., Lobenhofer, E. K., Puri, R. K., Schrf, U., Thierry-Mieg, J., Wang, C., Wilson, M., Wolber, P. K., et al. (2006) The microarray quality control (maqc) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol,* 24, 9**,** 1151-61.

Sluijter, J. P., Van Mil, A., Van Vliet, P., Metz, C. H., Liu, J., Doevendans, P. A. & Goumans, M. J. (2010) Microrna-1 regulate differentiation and proliferation in human-derived cardiomyocyte progenitor cells. *Arterioscler Thromb Vasc Biol,* 30, 4**,** 859-68.

Świniarski, R., Cios, K. J. & Pedrycz, W. (1998) *Data mining methods for knowledge discovery,* (Kluwer Academic).

Synnergren, J., Akesson, K., Dahlenborg, K., Vidarsson, H., Ameen, C., Steel, D., Lindahl, A., Olsson, B. & Sartipy, P. (2008) Molecular signature of cardiomyocyte clusters derived from human embryonic stem cells. *Stem Cells,* 26, 7**,** 1831-40.

Synnergren, J., Heins, N., Brolen, G., Eriksson, G., Lindahl, A., Hyllner, J., Olsson, B., Sartipy, P. & Bjorquist, P. (2010) Transcriptional profiling of human embryonic stem cells differentiating to definitive and primitive endoderm and further toward the hepatic lineage. *Stem Cells Dev,* 19, 7**,** 961-78.

Synnergren, J., Améen, C., Lindahl, A., Olsson, B., Sartipy, B. (2010) Molecular signature of cardiomyocyte clusters. Physiol Genomics, 2010 Sep 14. [Epub ahead of print].

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. & Yamanaka, S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell,* 131, 5**,** 861-72.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E. S. & Golub, T. R. (1999) Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc Natl Acad Sci U S A,* 96, 6**,** 2907-12.

Tan, P. K., Downey, T. J., Spitznagel, E. L., Jr., Xu, P., Fu, D., Dimitrov, D. S., Lempicki, R. A., Raaka, B. M. & Cam, M. C. (2003) Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res,* 31, 19**,** 5676-84.

Thum, T., Catalucci, D. & Bauersachs, J. (2008) Micrornas: Novel regulators in cardiac development and disease. *Cardiovasc Res,* 79, 4**,** 562-70.

Tusher, V. G., Tibshirani, R. & Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A,* 98, 9**,** 5116-21.

Van Rooij, E. & Olson, E. N. (2007) Micrornas: Powerful new regulators of heart disease and provocative therapeutic targets. *J Clin Invest,* 117, 9**,** 2369-76.

Wang, Y., Barbacioru, C., Hyland, F., Xiao, W., Hunkapiller, K. L., Blake, J., Chan, F., Gonzalez, C., Zhang, L. & Samaha, R. R. (2006) Large scale real-time pcr validation on gene expression measurements from two commercial long-oligonucleotide microarrays. *BMC Genomics,* 7**,** 59.

Wang, Y., Keys, D. N., Au-Young, J. K. & Chen, C. (2009) Micrornas in embryonic stem cells. *J Cell Physiol,* 218, 2**,** 251-5.

Vasudevan, S., Tong, Y. & Steitz, J. A. (2007) Switching from repression to activation: Micrornas can up-regulate translation. *Science,* 318, 5858**,** 1931-4.

Wilson, K. D., Venkatasubrahmanyam, S., Jia, F., Sun, N., Butte, A. J. & Wu, J. C. (2009) Microrna profiling of human-induced pluripotent stem cells. *Stem Cells Dev,* 18, 5**,** 749-58.

Von Mering, C., Jensen, L. J., Kuhn, M., Chaffron, S., Doerks, T., Kruger, B., Snel, B. & Bork, P. (2007) String 7--recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res,* 35, Database issue**,** D358-62.

Xu, X. Q., Soo, S. Y., Sun, W. & Zweigerdt, R. (2009) Global expression profile of highly enriched cardiomyocytes derived from human embryonic stem cells. *Stem Cells,* 27, 9**,** 2163-74.

Yuen, T., Wurmbach, E., Pfeffer, R. L., Ebersole, B. J. & Sealfon, S. C. (2002) Accuracy and calibration of commercial oligonucleotide and custom cdna microarrays. *Nucleic Acids Res,* 30, 10**,** e48.

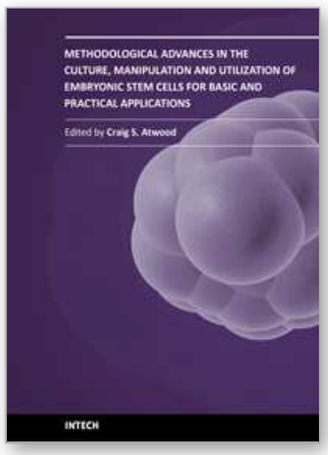
Zhang, C. (2008) Micrornas: Role in cardiovascular biology and disease. *Clin Sci (Lond),* 114, 12**,** 699-706.

Zhou, Y., Ferguson, J., Chang, J. T. & Kluger, Y. (2007) Inter- and intra-combinatorial regulation by transcription factors and micrornas. *BMC Genomics,* 8**,** 396.

Zhu, S., Wu, H., Wu, F., Nie, D., Sheng, S. & Mo, Y. Y. (2008) Microrna-21 targets tumor suppressor genes in invasion and metastasis. *Cell Res,* 18, 3**,** 350-9.

IntechOpen

IntechOpen



Methodological Advances in the Culture, Manipulation and Utilization of Embryonic Stem Cells for Basic and Practical Applications

Edited by Prof. Craig Atwood

ISBN 978-953-307-197-8
Hard cover, 506 pages
Publisher InTech
Published online 26, April, 2011
Published in print edition April, 2011

Pluripotent stem cells have the potential to revolutionise medicine, providing treatment options for a wide range of diseases and conditions that currently lack therapies or cures. This book describes methodological advances in the culture and manipulation of embryonic stem cells that will serve to bring this promise to practice.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Jane Synnergren and Peter Sartipy (2011). Microarray Analysis of Undifferentiated and Differentiated Human Pluripotent Stem Cells, Methodological Advances in the Culture, Manipulation and Utilization of Embryonic Stem Cells for Basic and Practical Applications, Prof. Craig Atwood (Ed.), ISBN: 978-953-307-197-8, InTech, Available from: http://www.intechopen.com/books/methodological-advances-in-the-culture-manipulation-and-utilization-of-embryonic-stem-cells-for-basic-and-practical-applications/microarray-analysis-of-undifferentiated-and-differentiated-human-pluripotent-stem-cells

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License, which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen