Computational Biology and Chemistry 64 (2016) 250–262

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

# An integrative approach predicted co-expression sub-networks regulating properties of stem cells and differentiation

Mousumi Sahu, Bibekanand Mallick*

*RNAi and Functional Genomics Laboratory, Department of Life Science, National Institute of Technology, Rourkela-769008, India*

ARTICLE INFO

*Article history:*
Received 4 March 2016
Received in revised form 22 June 2016
Accepted 15 July 2016
Available online 18 July 2016

*Keywords:*
Stem cells
Microarray
GSEA
Co-expression network
k-clique sub-network

ABSTRACT

The differentiation of human Embryonic Stem Cells (hESCs) is accompanied by the formation of different intermediary cells, gradually losing its stemness and acquiring differentiation. The precise mechanisms underlying hESCs integrity and its differentiation into fibroblast (Fib) are still elusive. Here, we aimed to assess important genes and co-expression sub-networks responsible for stemness, early differentiation of hESCs into embryoid bodies (EBs) and its lineage specification into Fibs. To achieve this, we compared transcriptional profiles of hESCs-EBs and EBs-Fibs and obtained differentially expressed genes (DEGs) exclusive to hESCs-EBs (early differentiation), EBs-Fibs (late differentiation) and common DEGs in hESCs-EBs and EBs-Fibs. Then, we performed gene set enrichment analysis (GSEA) followed by overrepresentation study and identified key genes for each gene category. The regulations of these genes were studied by integrating ChIP-Seq data of core transcription factors (TFs) and histone methylation marks in hESCs. Finally, we identified co-expression sub-networks from key genes of each gene category using k-clique sub-network extraction method. Our study predicted seven genes edicting core stemness properties forming a co-expression network. From the pathway analysis of sub-networks of hESCs-EBs, we hypothesize that FGF2 is contributing to pluripotent transcription network of hESCs in association with DNMT3B and JARID2 thereby facilitating cell proliferation. On the contrary, FGF2 is found to promote cell migration in Fibs along with DDR2, CAV1, DAB2, and PARVA. Moreover, our study identified three k-clique sub-networks regulating TGF-β signaling pathway thereby promoting EBs to Fibs differentiation by: (i) modulating extracellular matrix involving ITGB1, TGFB1I1 and GBP1, (ii) regulating cell cycle remodeling involving CDKN1A, JUNB and DUSP1 and (iii) helping in epithelial to mesenchymal transition (EMT) involving THBS1, INHBA and LOX. This study put forward the unexplored genes and co-expression sub-networks regulating stemness and different stages of differentiation of hESCs which will undoubtedly add to the comprehensive understanding of hESCs biology.



## 1. Introduction

Human Embryonic Stem Cells (hESCs) are characterized by continuous self-renewal and possess a parallel ability giving rise to differentiated progeny of all three germ layers. The maintenance of self-renewal and the gradual transition of hESCs to other cell types are dependent on a distinct hESC transcriptional network and its meticulous modulation in transiting to differentiating and terminally differentiated cells. The usual differentiation process of hESCs starts with formation of three-dimensional heterogeneous aggregations of hESCs, termed as Embryoid Bodies (EBs) which have the ability to produce different specialized cell types including hematopoietic (Wang et al., 2005), neuronal (Schulz et al., 2003; Rao, 2004), myogenic (Khoo et al., 2005) and cardiac muscle cells (Reppel et al., 2004; Segev et al., 2005) by responding to lineage-specific markers (Kurosawa, 2007). Thus, EBs represent an early stage in the lineage specification process possessing properties of hESCs and other differentiated progeny in their gene expression profiles. Further, the terminally differentiated fibroblast (Fib) cells are widely studied to uncover mechanisms underlying embryonic stem cells (ESCs) differentiation, adult cell reprogramming and transdifferentiation (Yu et al., 2007; Kidwai et al., 2013; Takahashi et al., 2007; Han et al., 2012). Thus, understanding the differences in transcriptional profiles and regulatory signatures contributing to cell-specific phenotypes of hESCs, EBs and Fibs are imperative which will help to unveil the molecular control driving the cellular conversion from hESCs to EBs and Fibs. In this study, we attempted to elucidate cellular

* Corresponding author.
*E-mail addresses:* vivek.iitian@gmail.com, mallickb@nitrkl.ac.in (B. Mallick).

http://dx.doi.org/10.1016/j.compbiolchem.2016.07.006
1476-9271/

differentiation of hESCs towards Fibs in a three-stage framework, i.e., hESCs, EBs, and Fibs to put forward unexplored regulatory nodes and networks of pluripotency and differentiation.

In hESCs maintenance, a set of transcription factors (TFs)—OCT4, SOX2, NANOG, KLF4, and c-MYC have been widely studied for their crucial role in self-renewal, pluripotency and cellular reprogramming (Marson et al., 2008; Anokye-Danso et al., 2011). In addition to maintaining self-renewal, and pluripotency of ESCs, these TFs are also known to suppress lineage specific markers in hESCs. Other than these TFs, histone methylations (e.g., H3K4m3, H3K27me3) also play a crucial role in stemness maintenance and cell fate decision by regulating transcription of genes (Lunyak and Rosenfeld, 2008; Wu and Sun, 2006). H3K4me3 at promoter or transcription start site (TSS) of a gene causes activation, whereas H3K27me3 within these regions causes inhibition of the gene. Intriguingly, the presence of both H3K4me3 and HK27me3 marks (bivalent domain) within the promoter region plays an essential role in hESCs maintenance and lineage specification by giving a poised expression to the associated gene (Voigt et al., 2013; Golebiewska et al., 2009). In this study, we used ChIP-Seq (Chromatin Immunoprecipitation sequencing) data to interlink the regulation of the key determinants of pluripotency, early differentiation to EBs and terminal differentiation of Fibs with core pluripotent TFs and two major histone methylation marks (H3K4me3 and H3K27me3) in hESCs.

Further, we built gene co-expression networks involved in important processes and pathways of pluripotency and differentiation. The gene co-expression networks are advantageous compared to commonly used biological networks such as protein–protein interaction networks, signaling networks, metabolic networks, and gene regulatory networks, etc. So, we constructed gene co-expression networks of differentially expressed key genes involved in pluripotency, early differentiation (hESCs to EBs) and late differentiation (EBs to Fibs) using publicly available gene co-expression data of human. Then, we adopted a clique-based method to identify sub-networks/communities within the gene co-expression networks. The cliques are densely connected closed sub-networks. These are thought to be conserved across cell types executing key biological processes and genes of a clique are believed to perform similar function. In summary, this study will help in deciphering crucial co-expressing gene modules explicitly dictating the processes/pathways of hESC pluripotency and Fib differentiation.

## 2. Materials and methods

The workflow of the study is provided in Fig. 1.

### 2.1. Gene expression analysis

In this study, we downloaded two microarray gene expression datasets, accession no: GSE25970 and GSE23034 profiled on two different platforms, Affymetrix HT Human Genome U133A and Affymetrix Human Gene ST 1.0 Custom RefSeq CDF version 12 respectively from GEO database of NCBI. These are the only two datasets that contain transcription profiles of all three cell types (hESC, EB, and Fib), and of a variety of cell lines. We used GeneSpring GX 12.6 (Agilent Technologies, Inc.) to analyze these two raw data sets independently. Pre-processing of the probes was carried out using gene-level expression analysis applying Robust Multichip Averaging (RMA) algorithm. We used percentile shift algorithm for summarization of the log-summarized values for each probe set creating a gene-level experiment. Further, filtration was performed to remove probe sets with lowest 20 percentile of

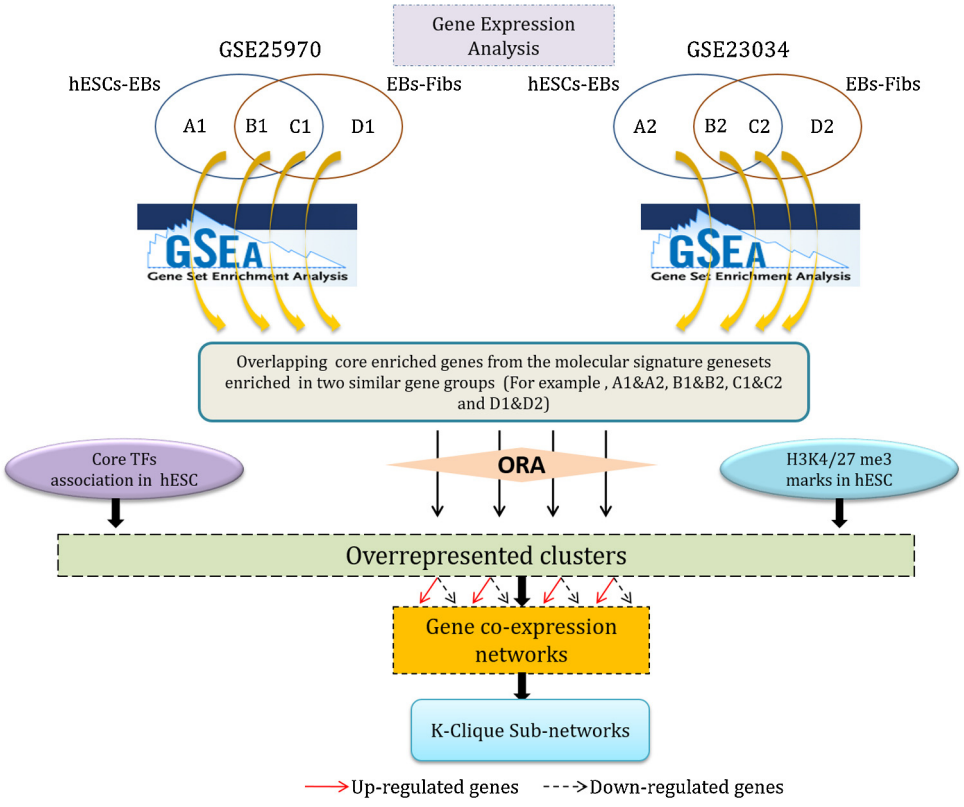


**Fig. 1.** Workflow of the study.
1 and 2 indicate two datasets, GSE25970, and GSE23034 respectively. A1 and A2: DEGs exclusive to hESCs-EBs; B1 and B2: DEGs in hESCs-EBs and EBs-Fibs; C1 and C2: DEGs in hESCs-EBs and EBs-Fibs; D1 and D2: DEGs exclusive to EBs-Fibs. ORA: Overrepresentation Analysis.