Selection of stable housekeeping genes in differentiating human pluripotent stem cells

Gustav Holmgren[1,2], Xianmin Zeng[3], Anders Lindahl[2], Peter Sartipy[1,4], Jane Synnergren[1]

[1] Systems Biology Research Center, School of Bioscience, University of Skövde, Skövde, Sweden

[2] Department of Clinical Chemistry and Transfusion Medicine, Institute of Biomedicine, The Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden

[3] Buck Institute for Research on Aging, 8001 Redwood Blvd Novato, CA 94945, USA

[4] Cellectis AB, Gothenburg, Sweden

Introduction

Housekeeping genes (HKGs) are involved in basic functions of the cell, and are assumed to be constitutively expressed at a constant level. Based on these features, HKGs are frequently used for normalization of gene expression data. However, recent studies show that the expression levels of several of the commonly used HKGs vary in different cell types[1] as well as after different biological treatments[2], making the choice of appropriate HKGs challenging. Compared to somatic cells, HKGs in stem cells are poorly studied[3] and even less explored in human pluripotent stem cells (hPSCs).

This work presents a comprehensive study on hPSC, including nine global gene expression datasets from both hESC and hiPSCs, obtained from studies of differentiation towards endoderm, mesoderm, and ectoderm. These datasets have been carefully mined for putative HKGs in hPSCs during different differentiation regimes.

Methods

A total of nine different datasets, including 113 microarrays have been analyzed in this study. Four of these datasets are generated using cell lines from Cellectis (www.cellectis.com). To remove background expression, the 25th quartile of the lowest expressed genes were filtered from these gene sets and only genes with an official gene symbol (HUGO) were included in the analysis. Genes with coefficient of variation (CV) <15 % across a dataset were defined as stably expressed.

Ref/Accession nr.	Diff. regime	Celline/ Origin	Nr. of samples	Nr. of arrays	Days of diff.	Nr. of stably expressed genes
Sivertsson *et al.* 2013	Endoderm	hESC	7	21	26	4,587
Unpublished data Synnergren, J.	Endoderm	hESC	4	12	28	5,015
Unpublished data Synnergren, J.	Endoderm	hiPSC	5	15	35	4,749
Synnergren *et al.* 2011	Mesoderm	hESC	3	9	49	5,926
E-GEOD-33622	Mesoderm	hiPSC	3	5	2	16,027
E-GEOD-28191	Mesoderm	hiPSC	6	12	11	9,558
E-GEOD-26629	Ectoderm	hiPSC	3	6	32	8,238
Unpublished data Zeng, X.	Ectoderm	hiPSC	4	4	49	2,434
E-GEOD-34551	Ectoderm	hESC	10	29	7	10,982

Table 1: Summary of the nine global microarray datasets that were used in this study.

The global microarray datasets used in this study (Table 1) represents differentiation experiments from all three germ layers endoderm, mesoderm, and ectoderm, and data from both hESCs and hiPSCs differentiation systems were included in the analysis.

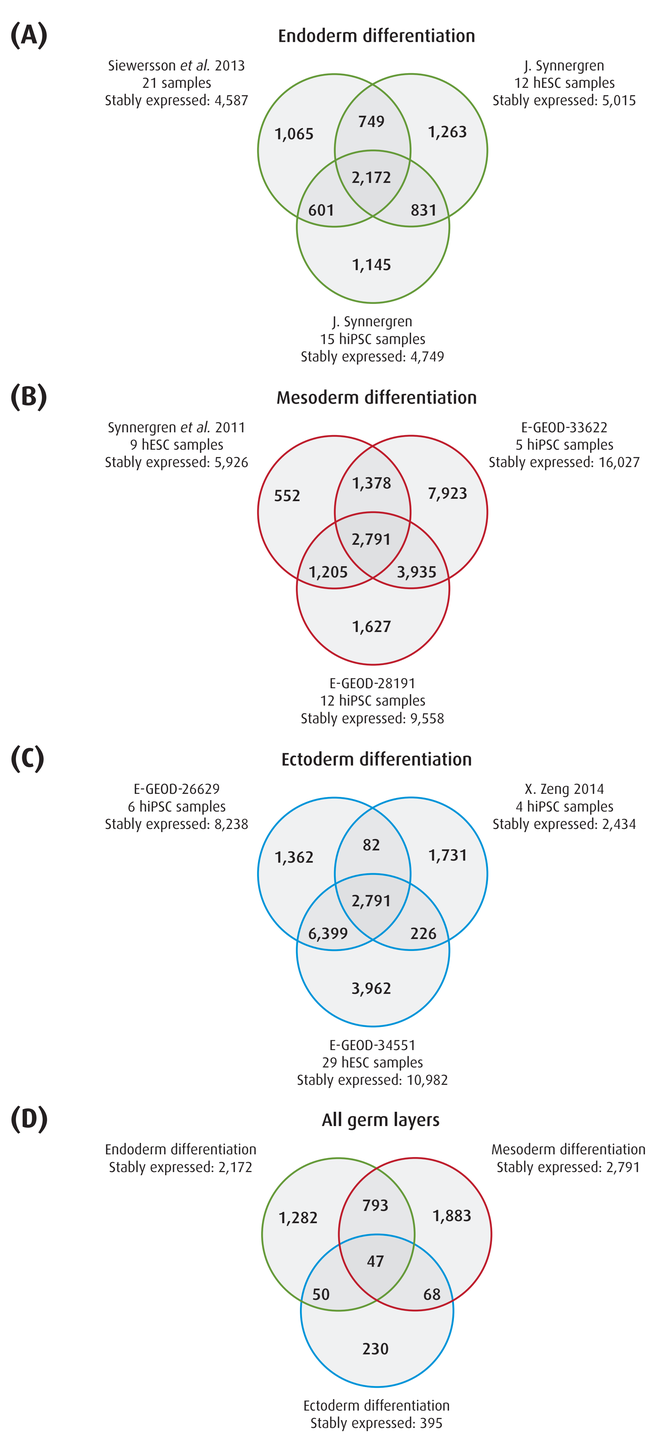

Figure 1: Venn diagrams showing the number of stably expressed genes in each of the three differentiation regimes separately (Panel A, B, and C) and the number of stably expressed genes that overlap between the differentiation regimes (Panel D). In total 47 genes show stable expression (CV<15%) in all nine datasets including 113 different samples from hPSCs and their derivatives.

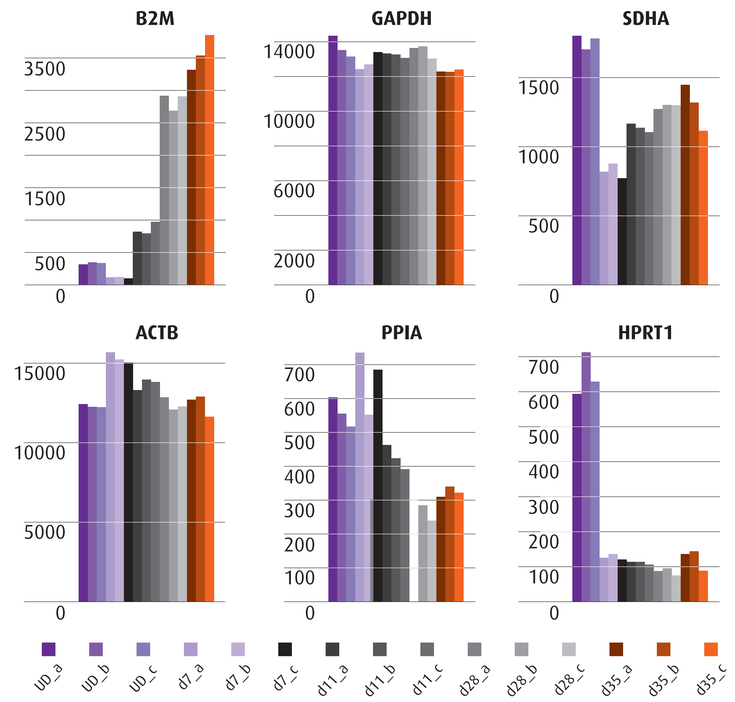

Figure 2: Expression profiles of six commonly used HKGs in 5 samples from hiPSC, representing definitive endoderm differentiation towards the hepatic phenotype.

Unigene ID	Gene Symbol	Gene Name
Hs.1545	CDX1	caudal type homeobox
Hs.629246	HDGFL1	hepatoma derived growth factor-like 1
Hs.134989	EN2	engrailed homeobox 2
Hs. 47472	DNAI2	dynein, axonemal, intermediate chain 2
Hs.503048	IGHMBP2	immunoglobulin mu binding protein 2
Hs.236646	HOXD9	homeobox D9
Hs.279259	EPX	eosinophil peroxidase
Hs.516855	CENPB	centromere protein B 80dkDa
Hs.436055	ALX4	ALX homeobox 4
Hs.147762	CCDC108	coiled-coil domain containing 108

Table 2: A selection of 10 genes that show the lowest variation in expression across all nine investigated datasets.

Conclusions

* The stability of HKGs varies substantially between different tissues and cell types.
* Several of the commonly used HKGs in somatic cells show unstable expression in differentiating hPSCs.
* There are different sets of genes that are stably expressed in the three investigated differentiation regimes and these only partly overlap.
* A set of 47 genes were identified that show stable expression during differentiation to all three germ layers.

References:

1. Lee, P.D. *et al.* Genome Res 12, 292-7 (2002).
2. Haller, F. *et al.* Anal Biochem 335, 1-9 (2004).
3. Eisenberg, E. & Levanon, E.Y. Trends Genet 19, 362-5 (2003).

Contact details:
Peter Sartipy
Cellectis AB
Arvid Wallgrens Backe 20, 413 46 Gothenburg, Sweden
email: peter.sartipy@cellectis.com
http://www.cellectis-bioresearch.com

UNIVERSITY OF SKÖVDE
1977