Polymorphic Dynamics of Ribosomal Proteins' Gene Expression during Induced Pluripotency

Prashanth kumar Guthikonda[1], Sumitha Prameela Bharathan[2], Janakiram Rayabaram[2], Sailu Yellaboina[3] , Shaji R Velayudhan[2] & Sreenivasulu Kurukuti[1*]

1.Department of Animal Biology, School of Life Sciences, University of Hyderabad, Hyderabad-500046, India.

2. Centre for Stem Cell Research, Christian Medical College, Vellore-632002, India.

3. CR Rao Advanced Institute of Mathematics, Statistics and Computer Sciences, UOH campus, Hyderabad-046, India

*Correspondence: Sreenivasulu Kurukuti Email: skurukuti@uohyd.ac.in

Abstract:

Reprogramming of somatic cells into Induced pluripotent stem cells (iPSCs) by defined factors has been well established. iPSCs were shown to be very similar to Embryonic Stem Cells (ESCs) but recent studies have reported that early stage iPSCs from various somatic cell sources differ in their differentiation potency into other cell-types. However, it was noted

that such iPSCs have differentiation bias to their donor cell types than to any other cell-types. Epigenetic memory in terms of DNA methylation/histone methylation of iPSCs has been attributed to the observed phenomenon but other mechanisms underlying this process remain to be explored. In the light of cell-type specific specialized ribosomes, by re-analyzing the publically available gene expression datasets among ESCs and various sources of iPSCs , here we report that transcripts of Ribosomal Protein (RP) subunit composition differ between ESCs and iPSCs. Further, ribosome protein subunit transcripts among various iPSCs are quite different and are dynamic. We studied the dynamic patterns of RP gene expression during different stages of pluripotency. We also discussed the possible outcomes/effects of deriving various cell types from these iPSC, in the context of ribosomopathies, and potential ways to overcome them. Our results provide a informatics' framework for researchers working on deriving iPSC and the implications will have profound impact on regenerative medicine and iPSC derived patient specific cell transplantation.

Keywords: iPSC, Ribosomopathies, Specialized Ribosomes, Regenerative medicine

INTRODUCTION

Embryonic stem cells (ESCs) are extremely worthy resources in regenerative medicine and serve as *in vitro* model systems to study early embryonic development and disease processes [1]. Induced pluripotent stem cells (iPSCs), in which ectopic expression of few transcription factors, generally Oct4, Sox2, Klf4 and cMyc (OSKM factors), into any adult somatic cells would reprogram them into an embryonic stem cell state [2]. Initially, iPSCs were thought to be very similar to ESCs, but later found to be substantially different in their gene expression patterns [3]. Irrespective of source of donor cell-types, iPSCs were shown to be less efficient in their differentiation potency to other cell-types, but they were shown to easily differentiate back into its respective donor cell-types, highlighting the influence of donor cell-type specific epigenetic signatures in this process [4]. However, it was noted that continuous passaging of these cells would attenuate these differences between iPSCs of different sources [5]. Donor cell-lineage specific factors, incomplete DNA methylation,

incomplete repression and reactivation of multiple genes[6], persistent donor cell-type specific gene expression or unique gene expression pattern [7] have been attributed to these observed phenomena. The transcriptional/post transcriptional regulation of these aberrant/unique epigenetic signatures of iPSCs is poorly understood but errors arising during reprogramming or incomplete reversion to pluripotency could be a cause. Since, the potential application of iPSCs in regenerative medicine and disease modelling depends on successful cell-type specific differentiation of iPSC, one need to investigate mechanisms behind this differentiation bias and methods to overcome it.

Apart from the above mentioned epigenetic determinants, other crucial cell-type specific protein complexes might also influences restricted differentiation of iPSCs. In this regard occurrence of cell-type specific ribosome composition (specialized ribosome protein composition) has attracted our attention. Researchers have reported that ribosome composition is tissue specific and expression levels of different ribosomal proteins are different in different tissues/cell-types [8-10]. Interestingly, decrease in concentration of a specific ribosomal protein was shown to affect a spectrum of translated mRNAs without affecting overall protein synthesis in a given cell. This explanation could account for the fact that mutations in some ribosomal proteins cause abnormality in particular tissue or cell-type but doesn't affect the whole body of an organism [reviewed in11]. Recently, Slavov et al., 2016 did mass spectrometric studies in RPs to further support the existence of ribosomes with distinct protein composition and physiological function [12]. Based on these observations, we hypothesized that heterogeneity in cell type specific ribosome composition could serve as one of the important determinants that might restrict iPSCs to achieve complete pluripotency and might explain differentiation biases.

Here we analysed expression pattern of ribosomal protein genes during different days of reprogramming of four somatic cells to respective iPSCs and compared them with that of ESC in humans. Our analysis identified unusual polymorphic behaviour of various ribosomal protein gene expressions during this process, in which not only a few patterns correlated well with cell-type specific ribosomal composition restricting differentiation capacity of

those derived iPSCs, but other unique patterns in iPSC of different founder cells showed up.We referred initial days of reprogramming as early stage and late stage ipsc as late stage.These results highlight the importance of ribosome composition and future work on it in efficient generation of pluripotent iPSCs equivalent to that of ESCs.

Materials and methods

Bioinformatics analysis of publically available gene expression data sets:

We studied and re-analysed publically available microarray gene expression data sets from ESCs, as well as iPSCs derived from four different somatic cell-types viz. Human dermal fibroblasts (HDF), Human astrocytes (HA), Normal human bronchial epithelium (NHBE), Human prostate epithelial cell (PREC) with early, intermediate and late days of reprogramming, including gene expression data sets under the accession number GSE50206 [13]. 75th percentile normalised expression data from all samples were analysed. Initially we defined the possible significant patterns based on visual observations. Then we used computational tools to derive the list of genes in each cell-type across the stages of reprogramming corresponding to the defined patterns. We set defined thresholds for an expression value to be considered as 'high expression'/ 'low expression'. Heat maps were generated using Cluster3 and Java TreeView tools.

Microarray gene expression data analysis:

75th percentile normalised expression data from all samples were analysed. We divided expression values of each dataset which consists of ESC, donor cell and its iPSCs passages of a particular cell type into two parts. One with a range of -0.5 to +0.5 (range1) and the remaining ones in another part (range2), since most of the genes have expression values in range1. We designed a PERL program to identify whether particular gene is expressed or not, such that if the gene is expressed, it will be greater than 0.3 in range1 and 0.4 in range2 as well as if the gene is not expressed or down regulated it should be less than -0.2 in range1 and less than -0.3 in range2. We selected these thresholds as such because at these

values, the patterns can be clearly seen under a heat map. We did not consider lesser values because they may hinder the significance of these results. We then set minimum threshold to consider a value as 'not expressed/'very less expressed' and upper threshold to consider as 'highly expressed'. Then we set the parameters for each of the patterns. We applied this program with defined thresholds for all the cell types.Heat maps were generated using Cluster3 and Java TreeView tools.

Generation of iPSCs from HDF:

To facilitate ecotopic retroviral-mediated reprogramming of HDFs, they were transduced with lentiviruses to express mouse *Slc7a1*, and the transduced cells were selected with Blasticidin S. Slc7a1^{+} hADFs were seeded at a count of 8x10^{5} cells in fibroblast medium on a 10 cm dish. The cells were subjected to two rounds of transduction with pools of freshly prepared pMXs-OSKM retroviral supernatants at 1:1:1:1 ratio in an interval of 48 hrs between the first and the second transductions. On day 4, 5x10^{5} transduced hADFs were seeded on a 6 well plate containing mitomycin-C treated SNL feeder cells in fibroblast medium. Two days later, the medium was changed to hiPSC medium. The reprogramming cells were fed daily with fresh medium and were observed for the emergence of iPSC colonies. For the derivation of iPSC lines by retroviral method (retro iPSC), colonies with ESC-like morphology were manually picked up in the 3rd – 4th weeks of reprogramming. The isolated clones were mechanically broken into small clumps and seeded on mitomycin C-treated SNL feeder layer in the iPSC medium. When the confluency reached 60-70% (4-5 days), the clones were subsequently passaged onto feeders by enzymatic treatment with 1mg/ml Collagenase IV (Life Technologies) for 45 minutes in 5% CO_2 incubator.

RNA isolation and Real time PCR analysis:

RNA was extracted from fibroblasts and iPSCs using Tri-reagent (Sigma-Aldrich). 1 µg of total RNA was used for reverse transcription reaction using high capacity cDNA reverse transcription kit (Life Technologies) according to the manufacturer's instructions. Quantitative RT-PCR was set up with SYBR Premix Ex Taq II (Takara Bio) using specific primers (ref Table) and analyzed with ABI 7500 (Life Technologies) or QuantStudio12K Flex

(Life Technologies) real-time PCR systems. The raw data was normalized with actin (Actb) gene expression.

Results:

We observed different ribosomal gene expression patterns with significance in reprogramming cells from published datasets and categorised into following nine patterns.(fig1&2)(1) The late stage iPSCs were similar to that of ESCs, but early stages were similar to that of donor cell type. The donor cell memories slowly get erased over time. (2) Both late and early stage iPSCs retained donor cell-type transcription profile. In this, most of the donor transcriptional memories are not erased during the course of reprogramming and were retained up to the last stages. These genes seem to be persistent in expression no matter how long the cells were passaged. (3) Late stage iPSCs retaining expression at similar levels as that of donor cell-type but immediate early stage hiPSCs showed higher levels of expression. This is a variation of pattern#2, where the late stage iPSC retain the donor memories, but immediately after induction these genes are overexpressed. (4)Same as #3 but late stage iPSCs showed ESC type composition. Donor memories initially increase but late stages convert into ESC type. This is a different pattern where initially the donor memories get overexpressed but as the time passes, the expression is converted back into ESC type. These genes might help in achieving pluripotency (5) ESC with intermediate expression, donor and immediate iPSC high expression, late iPSC low expression. This variation of pattern#1, but the late stage iPSC expression is lower than ESC. this may be beneficial or deleterious to the cell. (6) Genes that are not expressed in ESCs, donor cells or late iPSCs but were expressed in the intermediate stages. (7) Those only appear high in iPSC. They are neither expressed high in donor nor in ESC. (8) Genes, whose expression in less in ESC, more in donor but more than donor in late iPSC. (9) The expression of late stage iPSC is always less than that of ESC.

Through, our analysis has shown some other patterns but we did not consider them as so significant. At global level, most of the genes showed pattern# 1 (Fig1a), wherein the donor cell memory is retained in immediate stages but gets erased over the time and show ESCs type expression at late stage passages. But not all the genes followed the similar pattern.

We have listed out the genes which follow the above said patterns in other cell types in the same dataset (fig 4). Without considering these differences, if we derive iPSCs and differentiate into another cell-type, the resulting tissues will most probably result in abnormalities in their functions. Moreover some mitochondrial ribosomal proteins have also shown these trends suggesting the importance in iPSC. For example, genes like *Mrpl24*, *Mrps10* and *Mrpl53* show pattern #4. Interestingly, we noted that *Rps6ka4*, a ribosomal protein kinase, has moderate expression in donor cells but has high levels of expression in initial and intermediate passages.

Based on these bioinformatics study, we further validated observed expression patterns by real time PCR (qPCR) in iPSC lines derived from Human dermal fibroblasts for some genes in early and late passages of hiPSC (fig3) Most of them are following the expected patterns and in well agreement with our analysis , if not all.

Discussion:

Our results and lists are more of information to researchers working on pluripotency. We provided information about some unique patterns in iPSCs (fig4) For example, the genes that follow pattern# 2 could be repressed so that the late stage passages attain more pluripotency. We have observed that many ribosomal proteins are deficient in late stage iPSC passages. We are discussing the probable phenotypes one can observe if target tissues are derived from the late stage iPSCs with RP deficiencies.

Surprisingly, in all of the cells, *Rpl15* and *Rpl22* (except in PREC where *Rpl21* and *Rpl22*) followed pattern #9 (data not shown). Mutations in Rpl15 are associated with Diamond Blackfan Anemia 12 (DBA12) [14]. So, if blood cells are derived from this iPSC, they are most

likely to be associated with DBA phenotype. On the other hand, *Rpl22* deficiency selectively arrested development of alpha beta-lineage T cells at the beta-selection checkpoint by inducing their death [15]. So, T cells derived from these iPSC are most probably prone to the above said phenotype.

Interestingly, *Rps6ka3,* a protein kinase of *Rps6*, was observed to be deficient in all the late stage iPSC passages. The protein made by the *Rps6ka3* gene appears to play an important role in the brain. The protein is involved in cell signalling pathways that are required for learning, the formation of long-term memories, and the survival of nerve cells.

More than 125 mutations in the *Rps6ka3* gene have been identified in people with Coffin-Lowry syndrome, a condition associated with intellectual disability and skeletal abnormalities. All of these mutations severely reduce or eliminate the activity of the Rps6ka3 protein [16].

In PREC, a total of nine ribosomal proteins follow pattern#5, Depletion of *Rps9* by small interfering RNA in human cancer cell lines resulted in decreased global protein synthesis in association with induction of p53 target genes followed by cell cycle arrest or apoptosis depending on cell type [17]. Similarly, mutations in *Rps5* was shown to negatively affect translational accuracy in yeast [18]. *Rps2* was shown to be positive regulator of transferase activity cellular protein metabolic process [19]. Similarly *Rpl4* was also shown to be associated with DBA [20]. *Rpl15* was seen in pattern#8, where it is only expressed in late stage iPSC. This gene has been shown to be overexpressed in some oesophageal tumours compared to normal matched tissues [21]. So it is very likely that iPSC derived from PREC show the above phenotypes.

Coming to pattern#8, where the expression is only seen high in late stages of iPSC, there are some commonalities among the four cell-types. For instance, *Rps27* follows the pattern in

NHBE,HDF and PREC. It is interesting to note that over expression of *Rps27* shown to be linked to many cancers [22]. Similarly, *Rps6* follows the pattern in HDF and HA, *Rpl36A* and *Rps26* in both NHBE and PREC.

Moreover, many ribosomal genes such as *Mrpl27*, *Mrpl33* and *Mrpl52* are observed to be deficient in the late stage iPSCs . These defects in the mitochondrial ribosomal proteins may result in defects in mitochondrial functions such as oxidative phosphorylation.

It has been reported that small change in ribosomal protein expression may lead to drastic changes in regulation. Here in our study, we can say that the heterogeneity in ribosome composition among various induced pluripotent stem cells (iPSCs) should be taken into consideration while deriving iPSCs and target tissue from those iPSCs and is correlated with their differentiation potential and donor cell-type. One can predict the donor cell type by looking at early iPSCs passages. This may provide another clue that ribosomal protein composition play role in cell-type specific gene regulation and highlights he role of specialized ribosomes in iPSC.

List of Abreviations:

iPSCs: Induced Pluripotent Stem Cells; ESCs: Embryonic Stem Cells; RP: Ribosomal Protein;OSKM: Oct4, Sox2, Klf4 and cMyc; HDF: Human dermal fibroblasts; HA:Human astrocytes; NHBE: Normal human bronchial epithelium; PREC:Human prostate epithelial cell ; hADFs: human Adult Dermal Fibroblasts; DBA: Diamond Blackfan Anemia

Funding: Funding from UPE-II is greatly acknowledged. PK is grateful to UGC for providing fellowship as a part of UGC-JRF. SB acknowledge CSIR for research fellowship.

Acknowledgments: We would like to thank Ms. Subhalaxmi Mohanthy for her help in PERL programming. Help from BIF facility, School of Life Sciences, University of Hyderabad.

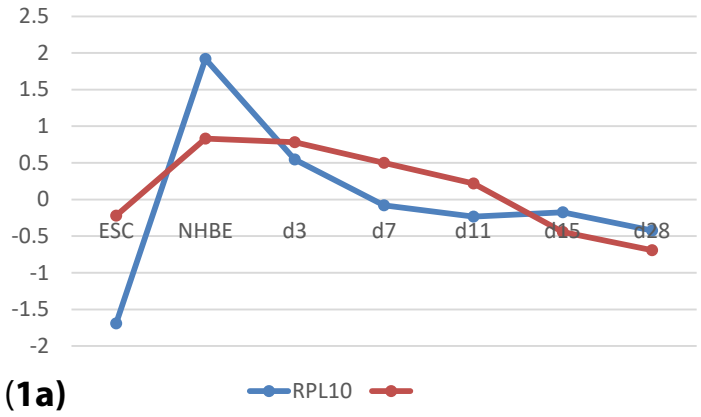
Author contributions: PK and SK conceived the idea. PK & SY analysed the microarray datasets. SP, JR & SV contributed iPSC and real time PCR analysis. PK, SV and SK wrote the paper.

References:

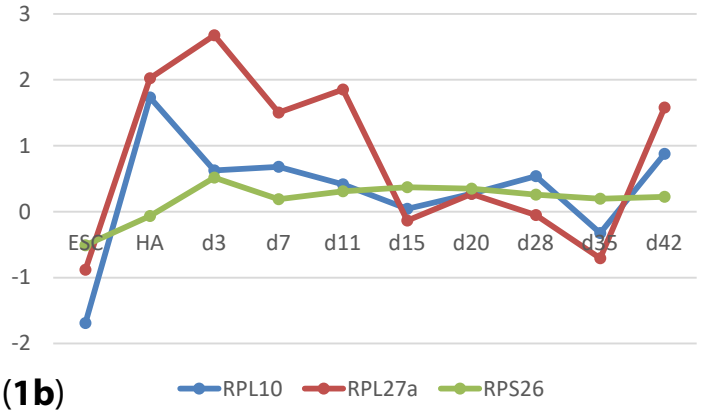
1. Keller, Gordon. "Embryonic stem cell differentiation: emergence of a new era in biology and medicine." *Genes & development* 19.10 (2005): 1129-1155.
2. Takahashi, Kazutoshi et al. "Induction of pluripotent stem cells from adult human fibroblasts by defined factors." *cell* 131.5 (2007): 861-872
3. Chin, Mark H et al. "Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures." *Cell stem cell* 5.1 (2009): 111-123
4. Kim, K et al. "Epigenetic memory in induced pluripotent stem cells." *Nature* 467.7313 (2010): 285-290.
5. Polo, Jose M et al. "Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells." *Nature biotechnology* 28.8 (2010): 848-855
6. Ohi, Yuki et al. "Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells." *Nature cell biology* 13.5 (2011): 541-549.
7. Ghosh, Zhumur et al. "Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells." *PloS one* 5.2 (2010): e8975.

8. Xue, Shifeng, and Maria Barna. "Specialized ribosomes: a new frontier in gene regulation and organismal biology." *Nature reviews Molecular cell biology* 13.6 (2012): 355-369.
9. Zhang, WeiYan et al. "Decreased expression of ribosomal proteins in human age-related cataract." *Investigative ophthalmology & visual science* 43.1 (2002): 198-204
10. Bortoluzzi, Stefania et al. "Differential expression of genes coding for ribosomal proteins in different human tissues." *Bioinformatics* 17.12 (2001): 1152-1157.
11. McCann, Kathleen L, and Susan J Baserga. "Mysterious ribosomopathies." *Science (New York, NY)* 341.6148 (2013): 849.
12. Slavov, Nikolai et al. "Differential stoichiometry among core ribosomal proteins." *Cell reports* 13.5 (2015): 865-873.
13. Takahashi K, Tanabe K, Ohnuki M, Narita M et al. Induction of pluripotency in human somatic cells via a transient state resembling primitive streak-like mesendoderm. *Nat Commun* (2014) 5:3678.
14. Landowski, Michael et al. "Novel deletion of RPL15 identified by array-comparative genomic hybridization in Diamond–Blackfananemia." *Human genetics* 132.11 (2013): 1265-1274.
15. Anderson, Stephen J et al. "Ablation of ribosomal protein L22 selectively impairs αβ T cell development by activation of a p53-dependent checkpoint." *Immunity* 26.6 (2007): 759-772.
16. Field, M et al. "Mutations in the RSK2 (RPS6KA3) gene cause Coffin–Lowry syndrome and nonsyndromic X-linked mental retardation." *Clinical genetics* 70.6 (2006): 509-515.
17. Lindström, Mikael S, and Monica Nistér. "Silencing of ribosomal protein S9 elicits a multitude of cellular responses inhibiting the growth of cancer cells subsequent to p53 activation." *PLoS One* 5.3 (2010): e9578.Galkin,
18. Galkin, Oleksandr et al. "Roles of the negatively charged N-late stage extension of Saccharomyces cerevisiae ribosomal protein S5 revealed by characterization of a yeast strain containing human ribosomal protein S5." *Rna* 13.12 (2007): 2116-2128.
19. Choi, Seeyoung et al. "PRMT3 inhibits ubiquitination of ribosomal protein S2 and together forms an active enzyme complex." *BiochimicaetBiophysicaActa (BBA) - General Subjects* 1780.9 (2008): 1062-1069.

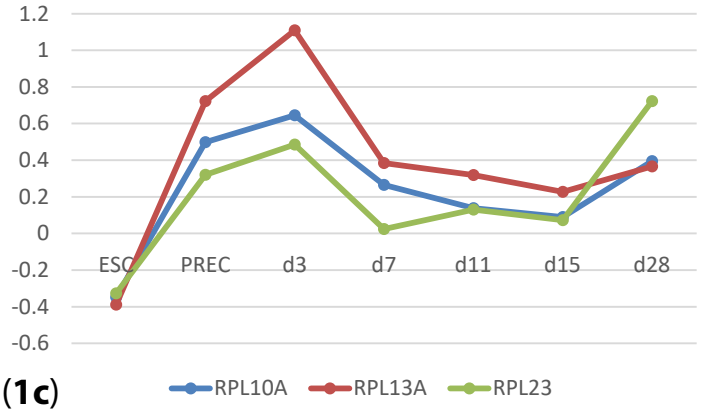
20. Koga, Yuhki et al. "Reduced gene expression of clustered ribosomal proteins in Diamond-Blackfananemia patients without RPS19 gene mutations." *Journal of pediatrichematology/oncology* 28.6 (2006): 355-361.
21. http://www.ncbi.nlm.nih.gov/gene/6138.
22. Fernandez-Pol, J Alberto. "Conservation of Multifunctional Ribosomal Protein Metallopanstimulin-1 (RPS27) through Complex Evolution Demonstrates its Key Role in Growth Regulation in Archaea, Eukaryotic Cells, DNA Repair, Translation and Viral Replication." *Cancer Genomics-Proteomics* 8.3 (2011): 105-126.



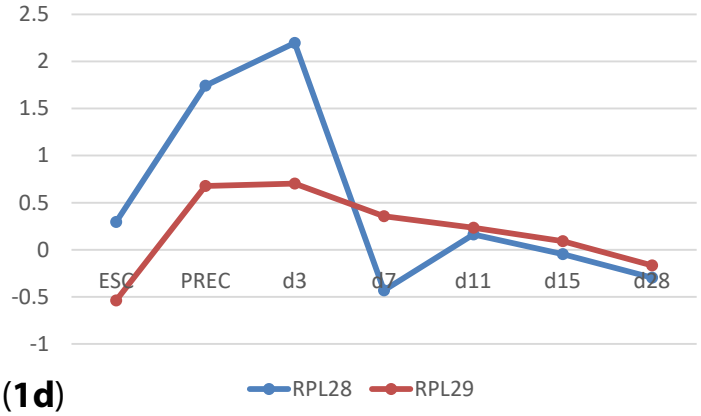
(1a)



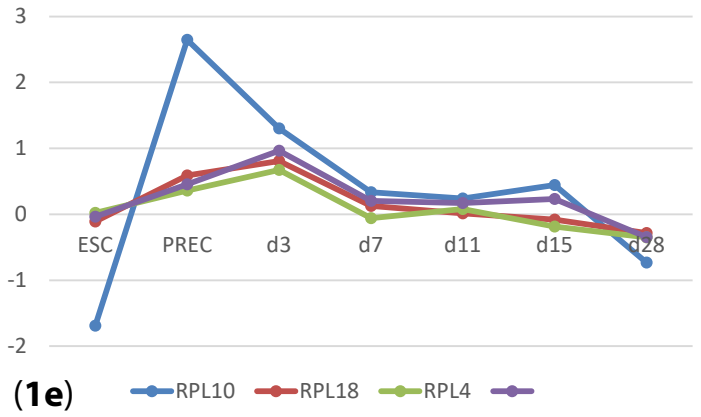
(1b)



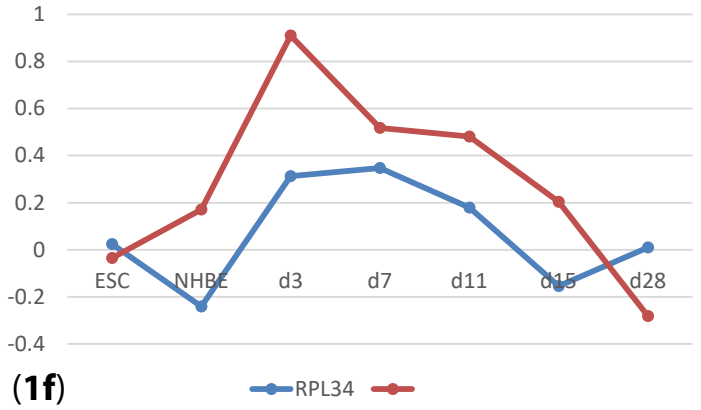
(1c)



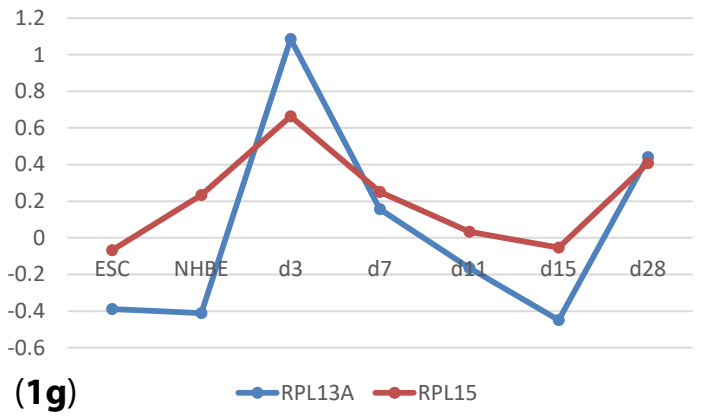
(1d)



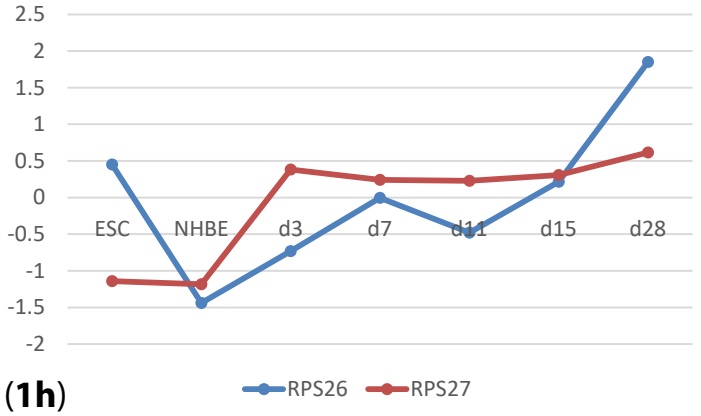
(1e)



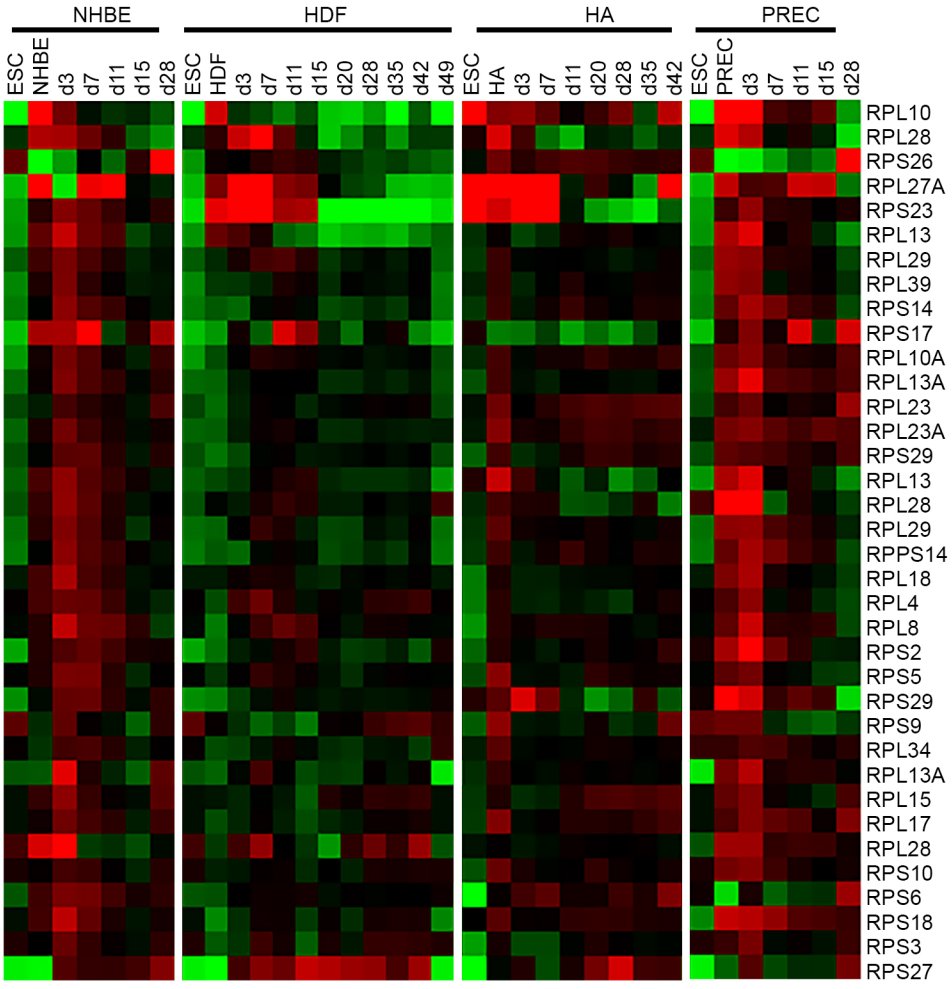
(1f)

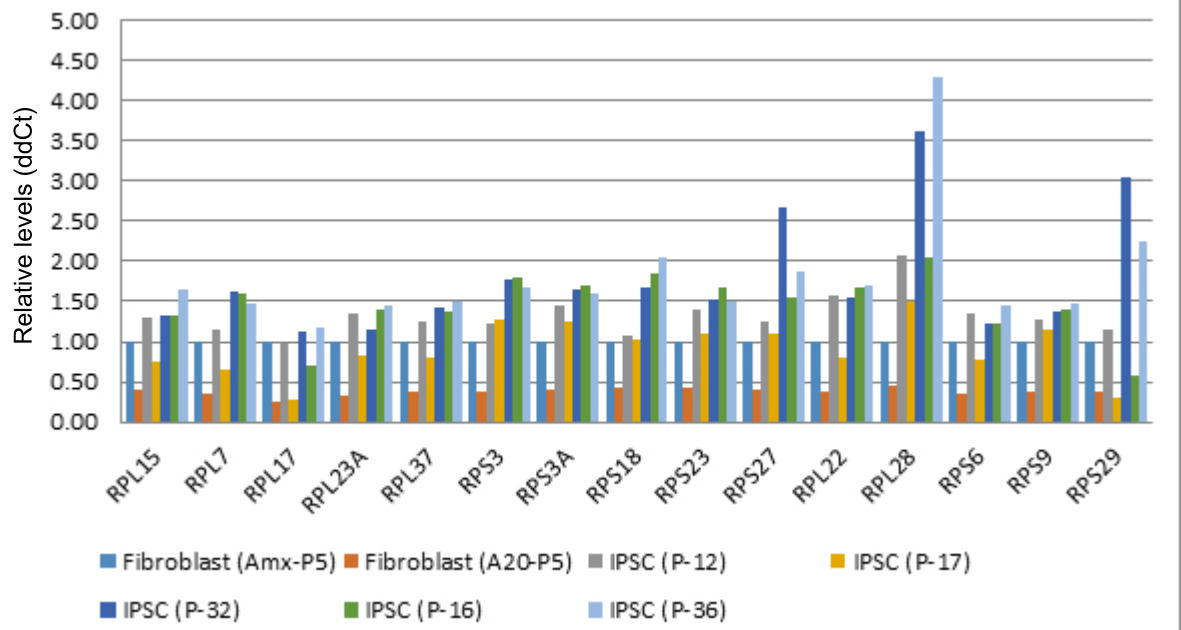


(1g)



(1h)


NHBE
HDF
HA
PREC
ESC
NHBE
d3
d7
d11
d15
d28
ESC
HDF
d3
d7
d11
d15
d20
d28
d35
d42
d49
ESC
HA
d3
d7
d11
d20
d28
d35
d42
ESC
PREC
d3
d7
d11
d15
d28
RPL10
RPL28
RPS26
RPL27A
RPS23
RPL13
RPL29
RPL39
RPS14
RPS17
RPL10A
RPL13A
RPL23
RPL23A
RPS29
RPL13
RPL28
RPL29
RPPS14
RPL18
RPL4
RPL8
RPS2
RPS5
RPS29
RPS9
RPL34
RPL13A
RPL15
RPL17
RPL28
RPS10
RPS6
RPS18
RPS3
RPS27


Relative levels (ddCt)
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
RPL15
RPL7
RPL17
RPL23A
RPL37
RPS3
RPS3A
RPS18
RPS23
RPS27
RPL22
RPL28
RPS6
RPS9
RPS29
Fibroblast (Amx-P5)
Fibroblast (A20-P5)
IPSC (P-12)
IPSC (P-17)
IPSC (P-32)
IPSC (P-16)
IPSC (P-36)

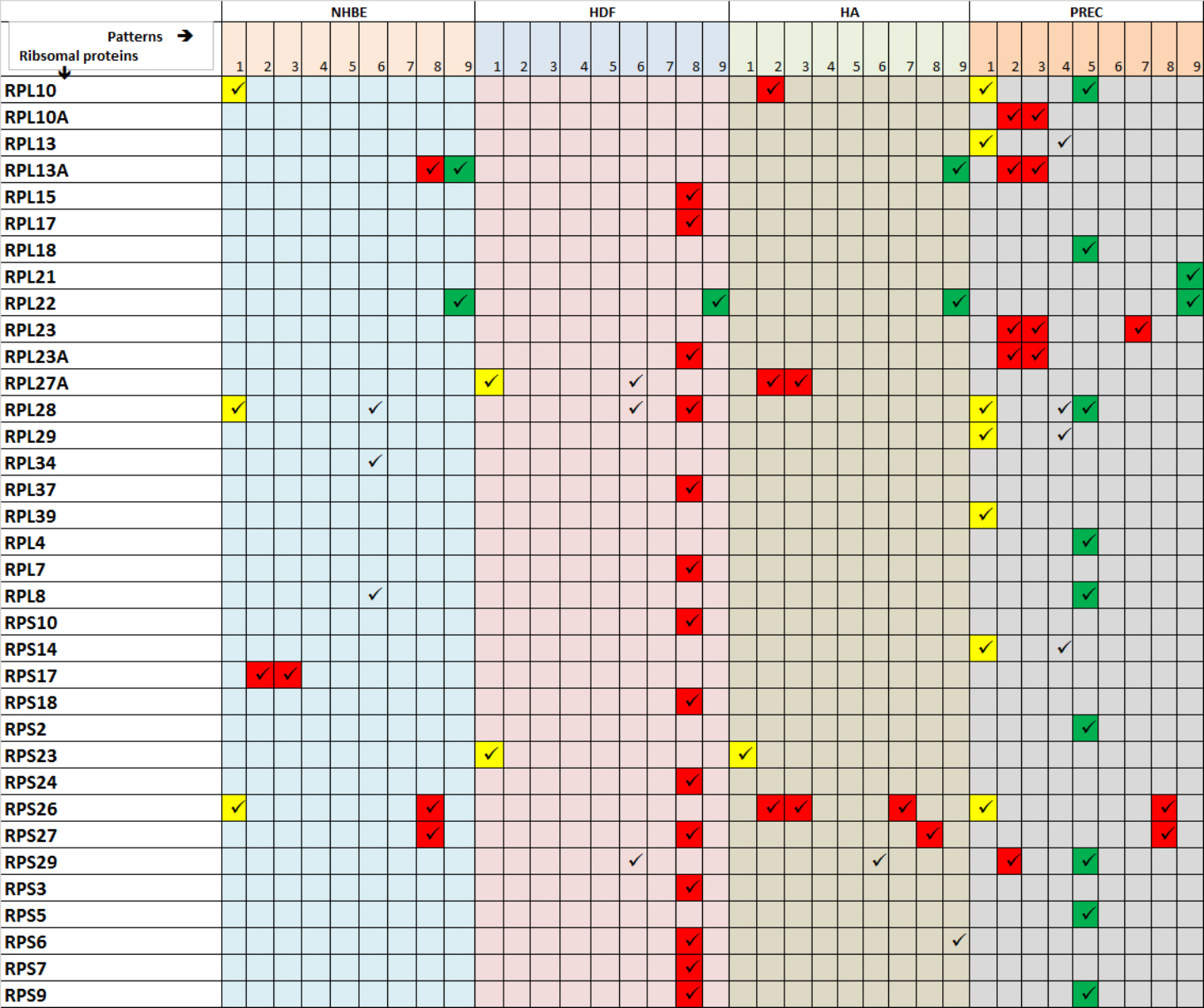

Patterns
Ribsomal proteins
NHBE
HDF
HA
PREC
1 2 3 4 5 6 7 8 9
RPL10
RPL10A
RPL13
RPL13A
RPL15
RPL17
RPL18
RPL21
RPL22
RPL23
RPL23A
RPL27A
RPL28
RPL29
RPL34
RPL37
RPL39
RPL4
RPL7
RPL8
RPS10
RPS14
RPS17
RPS18
RPS2
RPS23
RPS24
RPS26
RPS27
RPS29
RPS3
RPS5
RPS6
RPS7
RPS9

Figure legends

Figure 1: Line Graphs representing different patterns. a-h: The expression of Ribosomal genes and the dynamics during different patterns showing ESC, Donor cell type and different days of differentiation. The figures and patterns respectively are : a-1,b-2,c-3,d-4,e-5,f-6,g-7,h-8.

Figure 2: Heatmap showing patterns. All the ribosomal genes' expression which follow the patterns and their corresponding values in other cell types . colour scale: green (low expression)- black (Intermediate expression) – red(intermediate expression).

Figure 3: qPCR Validation. Real-time PCR validation of selected genes in established hiPSC lines at early and late passages. RP gene expression in iPS cells (Normalized to Actin and fibroblast - ddCt method)

Figure 4: Summary Table. summarizing the information about the genes following patterns. Those highlighted in Yellow show the typical expected expression pattern upon induction. Those highlighted in red are to be knocked down and those in green needs to be over expressed.