## MATERIALS AND METHODS

### TRANSCRIPTOME DATA PROCESSING

In this study we compared transcriptional profiles from 13 iPS cell lines from 8 published reports hiPS1_8 (Masaki et al., 2008), hiPS2_4 (Masaki et al., 2008), hiPS3_2 (Masaki et al., 2008), BJiPS12 (Maherali et al., 2008), p-iPS01 (Kim et al., 2009), rv-iPS01 (Kim et al., 2009), iPS-PDB-1lox (Soldner et al., 2009), iPS-PDB-2lox (Soldner et al., 2009), Hips7 (Lowry et al., 2008), hiPSC2 (Lowry et al., 2008), hiPS20B1 (Takahashi et al., 2007), hAFF-(4PU)-iPS9 (Zhao et al., 2008), KiPS4F2 (Aasen et al., 2008). Each iPS cell line was compared with the ES cell lines from the same experiment and platform in GEO database (http://www.ncbi.nlm.nih.gov/geo/). Accession number for each DNA chip and induction conditions are summarized in Supplementary Table 5. Raw data files were normalized on GeneSpring (Agilent). First, signal intensities less than 0.01 were set to 0.01, and then each chip was normalized to the 50 percentile of the measurements taken from that chip.

The fold change was calculated relative to the expression level of each gene in the ES cell line, and genes, which are more than 2.0 fold up- or down- regulated in human iPS cells were selected for further analysis. Selected genes were then annotated by their presence ("P") or absence ("A") on the respective DNA chip in somatic cells and only genes with consistent status over all cell lines were used for the further analysis.

### TRANSCRIPTION REGULATION ANALYSIS

Regulatory analysis was performed on Induced Genes category with motif discovery tool MEME (Bailey et al., 2006) and ExPlain3.0 suit (Wingender et al., 1996; Kel et al., 2006). Sequences to 5000 base pairs upstream and 500 base pairs downstream from the transcriptional start site (TSS) of the integrated TRANSPro database were used (Wingender et al., 1996).

Human position-specific scoring matrices (PSSM) for OCT4, SOX2, and NANOG were constructed from the regulatory regions of experimentally verified promoters using MEME motif discovery tool (Polouliakh et al., 2005; Bailey et al., 2006). We took this step because human PSSM for those transcription factors do not exist in any database. Following genes were used for the matrices construction: OCT4 matrix from NANOG, OSR2, MSC, KCNN2, PCTK2, RORB; SOX2 matrix from GREG2, SCN3A, NELL1, THBS2, HIST1H4D, INHBA), and NANOG matrix from ONECUT1, GSC, PRKCDBP, FOXB1, FGFR2 (Boyer et al., 2005). Those genes were experimentally verified to allocate respective transcription factor binding sites (TFBS) in their promoters (Boyer et al., 2005). Motif consensus derived from generated matrices are: OCT4 ("TTTGCATT"), SOX2 "(A/G)ACAA(A/T)G", and NANOG "TAATTG". KLF4 matrix was acquired from JASPAR database (http://jaspar.cgb.ki.se) and for c-Myc human/mouse TRANSFAC matrix was used.

Matrices of five transcription factors were combined then in one "Master_gene" profile to be used in F-match software within ExPlain3.0 suit (Wingender et al., 1996). F-match evaluates the set of promoters and for each matrix tries to find two thresholds: one, *th*-max, which provides maximum ratio between the frequency of matches in the promoter of in focus (control set, "C") and background promoters (background, "B") (over-represented sites); and the second threshold, *th*-min, that minimizes the same ratio (under-represented sites). A binominal distribution of the sites between each control promoter dataset and respective background dataset is calculated and the *p*-value is assigned. For each cell line promoter dataset of up- and down- regulated genes three background datasets of the same size were created using the ExPlain3.0 housekeeping genes set (997 genes) through random selection procedure. Over-represented consensus motifs appeared in three housekeeping background datasets was used as background motifs to calculate *p*-value for the respective motifs in control set using F-match (Wingender et al., 1996).

The selection criteria for the promoter to be accepted, as "*possessing predicted binding motif*" was that the promoter should have at least two out of three main transcription factor binding motifs OCT4, SOX2, NANOG under *p*-value less than 0.001.

### ANNOTATIONS "*in silico*"

Epigenetic features, such as CpG density and the histone H3 modification status in human ES cells were assigned based on publications (Pan et al., 2007; Weber et al., 2007; Meissner et al., 2008). GO annotations were performed on the manually curated Gene Ontology database within ExPlain3.0 (Wingender et al., 1996). Each cell line was analyzed independently, thus results summarized can have multiple inclusions of the same gene in several groups of Induced Genes category. Pathway search was done on KEGG database using DAVID functional annotation tool (Huang et al., 2009). *T*-test and ANOVA statistical analysis has been performed for calculating virus-type correlation with epigenetic features in down-regulated Inherited Genes category.

## RESULTS

### GLOBAL EXPRESSION COMPARISON OF iPS AND ES CELL LINES

Comparison of expression profiles has been done on 9961 genes shared by 12 cell lines from 7 experiments (**Figures 1A–G**) and on 6357 genes for cell line KiPS4F2 (Aasen et al., 2008), as it has less genes shared with other cell lines (Materials and Methods). Genes found over-expressed more than two-folds (FC $\geq$ 2.0) between iPS and ES cells in more than two cell lines were subjected for categorization into Induced Genes and Inherited Genes according to their status (absence/presence) in the somatic cell of origin.

### INDUCED GENES FALL IN FOUR GROUPS BY TRANSCRIPTION ANALYSIS

In order to investigate the binding likelihood of transcription factors OCT4, SOX2, NANOG, KLF4, and c-Myc to the promoters of genes in Induced category transcription regulation analysis were performed. For up-regulated Induced Genes category we had 278 genes shared in more than two cell lines and for down-regulated Induced Genes category we had 128 genes shared by more than two cell lines.

Transcription regulation analysis revealed existence of four groups: genes possessing predicted transcription factor binding motifs without significant hit to the particular GO category over a group. We call this group as "master" gene group, and we have in it 42 up-regulated and 25 down-regulated genes shared

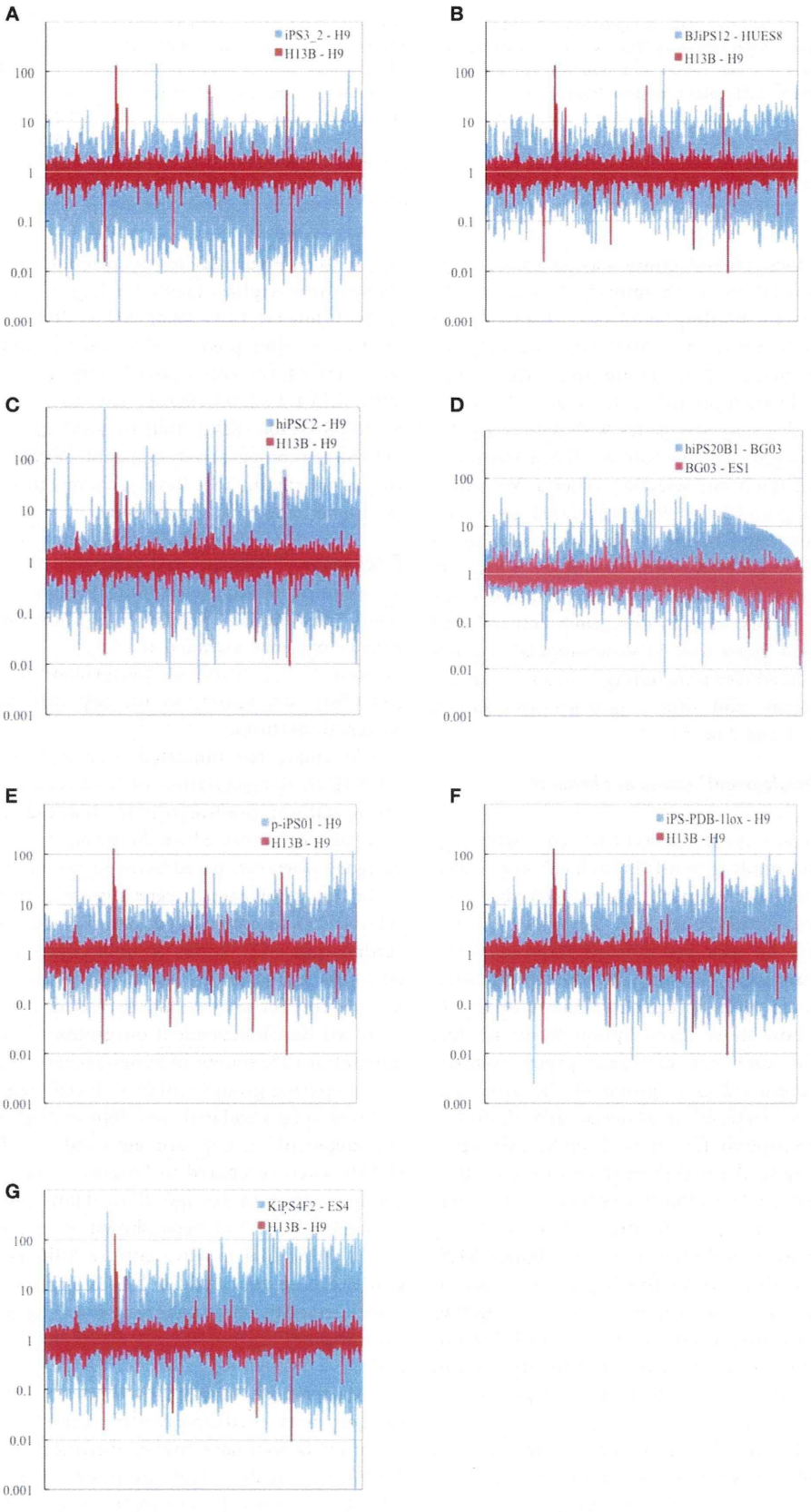


**FIGURE 1 | Continued**

**FIGURE 1 | Logarithm of fold change (FC) comparison of gene expression profiles between representative iPS cell from each of 7 experiments and respective ES cell (blue) and two ES cell lines H13B and H9 obtained from GEO platform 570 (red).** Eight experiments had less number of shared genes (6357) (not shown). In total 9961 genes (15,758 probes) shared among 7 cell lines are shown. **(A)** retorvirus-induced iPS3_2 cell line (Masaki et al., 2008), **(B)** lentivirus-induced BJiPS12 cell line (Maherali et al., 2008), **(C)** retrvorvirus MPx vector-induced hiPS02 cell line (Lowry et al., 2008), **(D)** retrovirus-induced hiPS20B1 cell line (Takahashi et al., 2007), **(E)** virus-free p-iPS01 cell line (Kim et al., 2009), **(F)** cre-recombinase excisable virus-induced iPS-PDB-lox1 cell line (Soldner et al., 2009), **(G)** Mouse stem cell virus-induced KiPS4F2 cell line (Aasen et al., 2008). Genes on each sub-figure are shown in the descending order of expression intensities in human ES cell line BG03. List of cell lines used is represented in Supplementary Table 5.

by more than two cell lines. Second group was genes with significant hit ($p$-value $\leq 0.05$) to "development" GO term and predicted transcription factor binding motif (see "Transcription Regulation Analysis" in Materials and Methods). We call this group as "master_development" gene group and have 36 up-regulated genes and 8 down-regulated genes shared by more than two cell lines. Third group was genes with significant hit to GO "development" category and without predicted transcription factor binding motif upon our selection criteria. We call it "development" gene group and have 89 up-regulated genes and 44 down-regulated genes found in more than two cell lines in it. Fourth group was those genes without significant hit to any particular GO term over a group and without predicted transcription factor binding motifs. We call this group "others" and obtained 111 up-regulated genes and 51 down-regulated genes for this group. List of Induced Genes including "master," "master-development," "development," and "others" gene groups is shown in Supplementary Tables 1 and 2 (a–d).

### *"Master" and "master_development" genes are found in each cell line*

We were interested to know the proportion of genes with predicted transcription factor binding motifs in each cell line ("master" and "master development" genes). We noticed that 18% (SD ± 11.01) of up-regulated genes and 23% (SD ± 9.2) of down-regulated genes in each cell line of Induced Genes category belong to "master" and "master_development" groups, i.e., the proportion of such genes in each cell line is similar in size. Then we checked how those transcription factor binding motifs are distributed in each cell line gene group comparing to the random background gene group of the same size ("Transcription Regulation Analysis" in Materials and Methods). **Figure 2** represents transcription factors with motif overrepresentation score comparing to the background set in each cell. It is interesting to note that binding affinities of transcription factors in down-regulated groups is significantly higher comparing to up-regulated groups, as it is shown below in **Figures 2A,B**. Larger fraction of genes with predicted binding motifs in down-regulated "master" and "master-development" groups together with observed stronger binding affinity of those motifs lead us to the conclusion that down-regulation by hypermethylation of CpG islands via the reprogramming must be a major problem to be overcome (Ohi et al., 2011).

NANOG was predicted in the promoters of up-regulated groups of 13 cell lines and in down-regulated groups of 7 cell lines. The fact that NANOG was not one of the reprogramming factors implies the possibility of its ectopic activation in the course of reprogramming, which was also confirmed in other studies (Jiang et al., 2011). **Figure 2C** depicts promoter model with the approximated transcription factor binding allocation identified in our study. While OCT4 [−4800 (SD ± 200) to −3600 (SD ± 200)], OCT4_rev-comp.pos1 [−3700 (SD ± 1000) to −2700 (SD ± 800)], OCT4_rev-comp.pos2 [−400 (SD ± 100) to + 500 (SD ± 0.0)], KLF4 [−2000 (SD ± 600) to −1600 (SD ± 700)], and NANOG [−400 (SD ± 100) to +400 (SD ± 100)] allocations to the TSS are comparatively constant, SOX2, and c-Myc positioning was not related to TSS. Reverse-complementary predicted sites of OCT4 were found in all cell lines.

### *Epigenetics of Induced Gene category*

To examine differences among groups and categories regarding the histone H3 epigenetic modifications and CpG promoter density of genes included (Pan et al., 2007; Weber et al., 2007; Meissner et al., 2008) we performed chi-square independence test (Yates correction) to identify correlations based on these epigenetic features.

Chi-square test indicated prevalence of genes with bivalent (H3K4K27, trimethylation of both lysines K4 and K27) modification status of promoters of the Induced Gene category ($p$-value $\leq 0.002$) as compared to the promoters of the Inherited Gene category when compared between two categories.

In Induced Genes category, genes with bivalent modification (H3K4K27), unmethylated genes and genes with non-defined methylation "ND' status in total constitute same proportion of 60% (SD ± 0.1) in up-regulated genes in each cell line and similar proportion of 71% (SD ± 7.1) of down-regulated genes in each cell line, independent on group. Gene lists with respective annotations are shown in Supplementary Tables 1 and 2 (a–d).

Comparing groups within Induced Genes category we identified that as up-regulated "development" group as down-regulated "development" group are enriched in High CpG promoters (HCP) when compared to "master" and "master_development" groups with $p$-values less than 0.001 and less than 0.01, correspondingly. It has been shown in previous study (Meissner et al., 2008) that *in vitro* culture induces hypermethylation of housekeeping genes coded by HCP promoter. These genes are associated with ubiquitous housekeeping genes and key developmental genes (Saxonov et al., 2006); both are highly expressed in ESCs.

"Master" and "master_development" groups of Induced Genes category with predicted binding sites for reprogramming transcription factors have higher inclusion of Low CpG promoters (LCP), Intermediate CpG promoter (ICP) and promoters with non-defined CpG density (ND), in comparison with "development" and "other" groups under $p$-value of less than 0.05, where transcription factor binding motifs were not predicted.

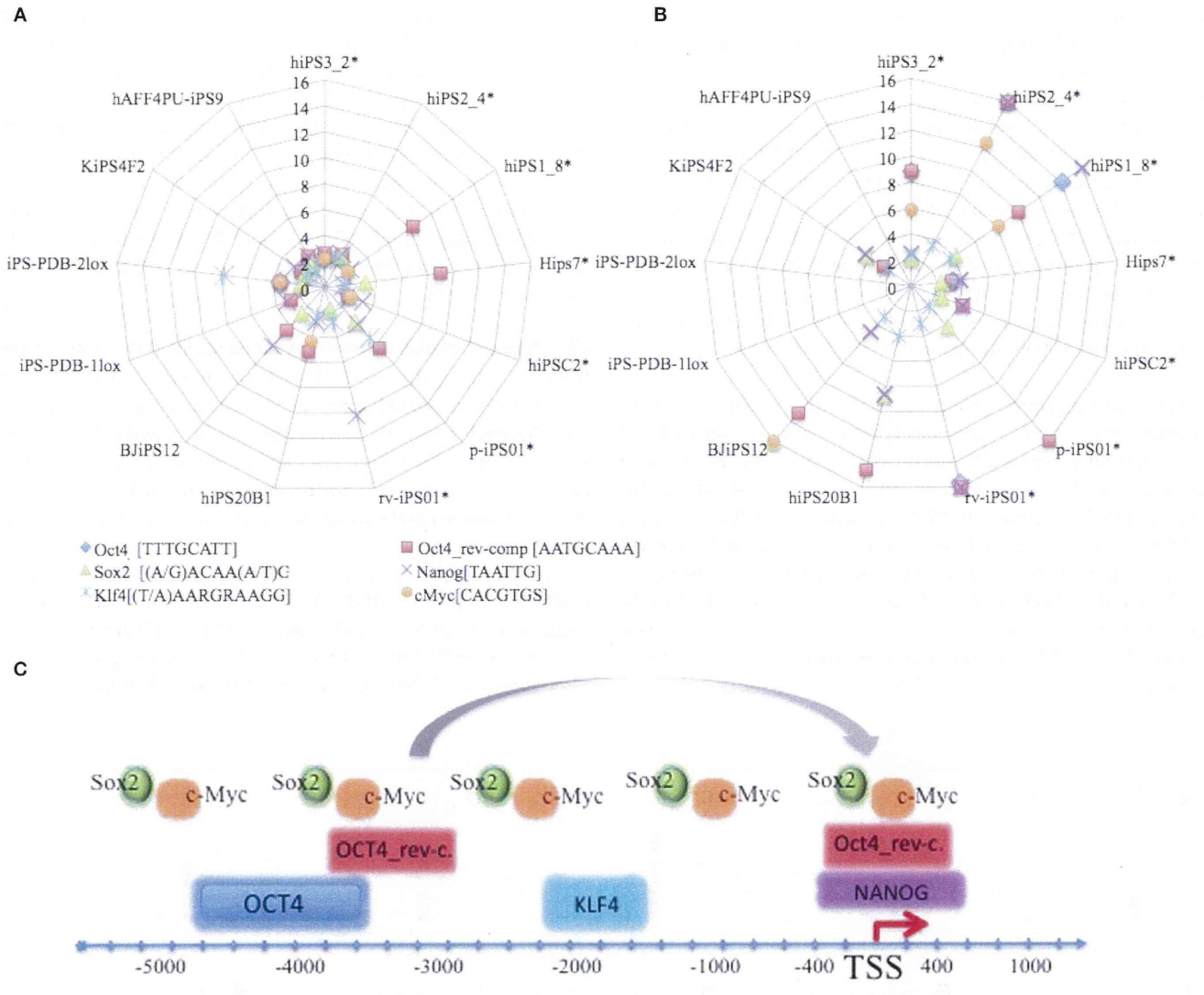


**FIGURE 2 | Distribution of predicted binding motifs over-representation score in the promoters of up- (A) and down-(B) regulated genes comprising "master" and "master_development" groups in 13 cell lines used in our study.** Asterisks (*) reference neonate cell lines. Symbols on the graphs correspond to the values represent over-representation score of the motif in the respective cell line (control) comparing to the random set of housekeeping genes (background) (Materials and Methods). Motifs consensus is shown aside the motif name. Number of genes for each cell line can be referenced from Supplementary Tables 1 and 2. IUPAC symbol R and S connoting A/G and C/G, respectively. Panel **(C)** represents promoter regulation model based on the allocations of predicted transcription factor binding sites by F-match tool in ExPlain3.0 suite. TSS denotes transcriptional start site.

We conclude that transcriptional activity status of CpG-poor promoters hypermethylated in somatic cells is not precluded on the first stage of the transfection, and consequently these genes are showing more plastic response to re-programming in comparison with High CpG less methylated and inactive promoters (Pan et al., 2007).

### *Pathways in Induced Genes category*

Following pathways (Huang et al., 2009) were detected in up- and down-regulated or either group in Induced Genes category: calcium signaling pathway (4.00E-03, up-, down-), cell adhesion molecules CAMs pathway (1.27E-02, up-), PPAR signaling pathway (1.83E-02, up-), Tight junction (2.69E-02, down-), neuroactive ligand-receptor interaction (3.45E-02, down-) pathways. Calcium related genes in both up- and down-regulated groups might imply the possibility for induction of differentiation and development.

## INHERITED GENES OF SOMATIC TRANSCRIPTIONAL MEMORY

Inherited Genes are those expressed in somatic cell and up- or down-regulated in iPSC comparing to ES. They are mainly of two origins: (a) those retaining their methylation status from

somatic substrates and (b) those activated or repressed through the viral transduction in iPSC in the course of reprogramming. We identified and categorized 1367 up-regulated genes and 1113 down-regulated genes of Inherited Genes category focusing on genes found in more than 3 cell lines (Supplementary Tables 3 and 4). Inherited Genes includes half of genes with univalent (H3K4, trimethylation of lysine residue 4) modification status in ES cell genes with $p$-value less than 0.001 against Induced Genes category by chi-square independence test. Recent observation showed that reprogramming a somatic cell into a pluripotent state generates hundreds of aberrantly methylated loci, predominantly at CpG islands, and associated with genes (Meissner et al., 2008), thus this category of Inherited Genes including many demethylated loci is the most prone to reprogramming (Polo et al., 2010).

iPSC up-regulated group of Inherited Genes category (Supplementary Table 3) includes IGF2 parentally expressed and H19 maternally expressed imprinted genes, shared by 3 and 5 different iPSC lines, respectively, and several IGF-binding proteins: IGFBP7 (7 cells), IGFBP3 (6 cells), IGFBP5 (5 cells), IGFPB6 (3 cells). In the embryo, imprinted genes regulate growth and development. Postnatally, imprinted genes control behavior, which may also affect the flow of nutrients from the mother to the developing pup (Davis and Uthus, 2003; Haig, 2004). Increased activity of the IGF2 gene has been associated with many types of cancer (Chao and D'Amore, 2008).

### *Pathways in Inherited Genes category*

Following pathways (Huang et al., 2009) were detected in up- and down-regulated or either group in Inherited Genes category: Focal adhesion (4.28E-03, up-, down-), ECM-receptor interaction (7.71E-07, up-), p53 signaling pathway (1.97E-02, up-, down-), Pathways in cancer (2.83E-05, up-, down-), Melanoma (9.36E-04, down-), Lysosome (3.39E-03, up-), Apoptosis (5.47E-03, up-, down-), Prostate cancer (1.70E-02, down-), Bladder cancer (6.65E-03, up-), Small cell lung cancer (2.82E-02, down-), Colorectal cancer (4.09E-03, up-), and Tight junction (7.13E-03, up-, down-). Genes from these groups may affect carcinogenic nature of iPS cells.

### *Virus type correlation in down-regulated group of Inherited Genes*

Here we investigate correlations between virus-type and number of down regulated genes of Inherited category in each cell line. **Figure 3** represents down-regulated Inherited category genes shared by more than 3 cell lines and annotated by promoter CpG density and histone H3 modification status in ES cells. List of Inherited down-regulated genes is represented in Supplementary Table 4 (a–c).

We have following cell lines: 2 virus-free cell lines, such as p-iPS01 (249 genes), iPS-PDB-lox1 (242 genes); 8 retrovirus cell lines, such as, hiPS3_2 (673 genes), hiPS2_4 (478genes), hiPS1_8 (440 genes), hiPS20B1 (112 genes), Hips7 (519 genes), hiPSC2 (469 genes), rv-iPS01 (185 genes), KiPS4F2 (506 genes), and 3

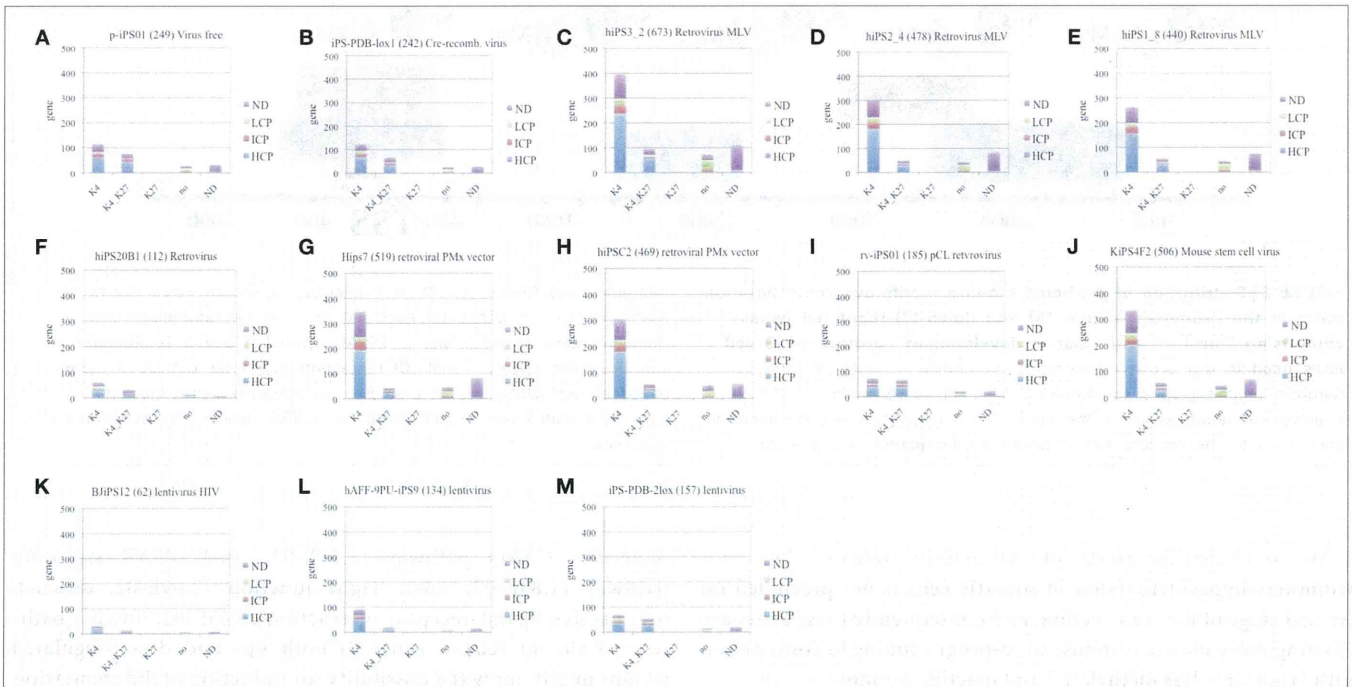


**FIGURE 3 | Classification of iPS cell down-regulated Inherited genes according to the histone H3 methylation status in ES cell and CpG promoter density. (A)** virus free **(B)** cre-recombinase excisable virus, **(C–J)** retrovirus, and **(K–M)** lentivirus. Figure legends specifications for methylation status of histone 3 (H3) in ES cell: "K4"—trimethylation of lysine on residue 4, "K4_K27"—trimethylation of lysine on residue 4 and residue 27, "K27"—trimethylation of lysine on residue 27, "no"-no methylation marks onH3, "ND"—undefined methylation status. Figure legends specifications for the promoter CpG density in ES cell: "NO", undefined CpG promoter density; LCP, Low CpG promoter density; ICP, intermediate CpG promoter density; HCP, high CpG promoter density.

lentivirus cell lines, such as BJiPS12 (62 genes), hAFF-4PU-iPS9 (134 genes), and iPS-PDB-2lox (157 genes).

The lowest number of down-regulated RRGs was found in following iPSC lines: BJiPS12 (62 genes), hiPS20B1 (112 genes), hAFF-4PU-iPS9 (134 genes), and iPS-PDB-2lox (157 genes), where last three cell lines are derived through the lentiviral transduction. Cre-recombinase excisable virus induced cell line iPS-PDB-lox1 (242 genes) and virus-free cell p-iPS01 (249 genes) had comparatively low number of down-regulated genes, but still higher than lentivirus induced cells BJiPS12 (62 genes), hAFF-4PU-iPS9 (134 genes), and iPS-PDB-2lox (157 genes) (**Figure 3**). While the transcription factors used for reprogramming can be excised by inducible gene expression once reprogramming is established (Pan et al., 2007; Jiang et al., 2011), residual sequences and chromosomal disruptions may still result in harmful alterations that could pose clinical risks. Retrovirus induced cell line rv-iPS01 (185 genes) produced in the same experiment (Kim et al., 2009) with straightforward protein delivery inducted cell line p-iPS01 (249 genes) show slightly lower fraction of down-regulated genes, while the opposite tendency is expected.

The rest of cells produced by retroviral induction show approximately same level of RRGs (510 genes, SD ± 82). It is known that the MLV retrovirus vector integrates near transcription start sites and CpG islands, while the HIV lentivirus vector integrates preferentially within active transcription units (Lewinski et al., 2006; Woltjen et al., 2009). Based on the results we assume that integration of retrovirus in the proximity of TSS of the dividing cells can significantly increase the probability of stochastic silencing, which results in larger fraction of genes down-regulated.

*T*-test showed statistical difference on genes with H3K4 methylation status with *p*-value less than 0.02, when retrovirus group and virus-free groups were compared. Therefore, we assume the higher plasticity of virus-free derived cells, as they are having less number of housekeeping and key developmental genes with H3K4 methylation status in ES cell.

ANOVA analysis ($p$-value $\leq 0.05$) for promoter density characteristics over virus-free, retrovirus, and lentivirus groups showed difference in distribution for ICP genes and genes with undefined CpG density ("ND"), which is also reflected in recent publications (Pan et al., 2007; Weber et al., 2007; Meissner et al., 2008).

Most of down-regulated groups enriched with genes having the univalent (H3K4) modification status in human ES-cells and HCP (**Figures 3E,J,H,G**) are those observed in iPSCs generated by retroviral or lentiviral transduction. Virus-free cell lines have higher fraction of genes with bivalent (H3K4K27) modification status in human ES cells, i.e., those, possibly exhibiting higher plasticity to reprogramming. These finding can help biologists to select the appropriate virus for the analysis.

Three cell lines of neonatal somatic cell origin, produced in the same experiment (Masaki et al., 2008) with different passage of length hiPS3_2 80 days (673 genes), hiPS2_4 102 days (478 genes), and hiPS1_8 180 days (440 genes) show gradual decrease in down-regulated gene number with the increase of passage number, which correlated with previous findings (Chin et al., 2010; Ohi et al., 2011).

## DISCUSSION

Genes showing statistical difference in expression between iPSCs and ESCs and shared by several iPSC lines can be considered as RRGs. RRGs were classified in two categories of Induced Genes and Inherited Genes depending on their expression status in somatic cells of origin. Induced Genes exhibit bivalent (H3K4K27) modification status in ES cells with the predominance of Intermediate and Low CpG density promoters and promoters with non-defined CpG density ("ND").

Faulty resetting of inactive yet "posed" state of bivalent domains is critical for the consequent differentiation of iPS cell in tissue (Kim et al., 2010). On the contrary, Inherited Genes category was enriched in univalent (H3K4) modification status in ES cell and showed preponderance for High CpG density promoter genes.

Part of our Inherited Genes category was identified as fibroblast-associated in prior studies (123 genes), while Induced Genes category associated with iPS-specific reprogramming network showed tiny overlap (8 genes) of our genes with resistant genes known from other studies (Chin et al., 2010; Gupta et al., 2010; Newman and Cooper, 2010; Lister et al., 2011; Ohi et al., 2011). The analysis of iPSC-differentiated cardiomyocytes beating clusters reported by Gupta et al. showed significant overlaps with our results in 111 genes (87 up-, 24 down-) with our Inherited Genes category and only 6 genes in Induced category (Supplementary Tables 1, 2, 3, 4 "overlap" columns). Ten genes (COMP1, DYNLT3, NME4, OXCT1, MGMT, PTGR1, MGC3207, CKLF, ZNF167, ZNF626) from our study were verified in functional analysis by qPCR to be over-expressed in somatic cell and continue their *up*-regulation in iPSC-derived cardiomyocyte beating clusters (Gupta et al., 2010). Seven out of 15 genes reported as differently expressed between iPS and ES cell lines over four laboratories in the study of Newman and Cooper also found in Inherited (6 genes) and Induced (1 gene) Gene categories in our study.

CSRP1 (cystein and glycine rich protein 1, 6 cell lines, neonate, and adult), COMT (catecol-O-methyltransferase, 5 cell lines, neonate and adult) in Inherited up-regulated group (Supplementary Table 3a) and C9orf64 (5 cell lines, neonate) Inherited down-regulated group (Supplementary Table 4a) are top somatic cell genes expressed in iPS cell (Ohi et al., 2011) also identified in our study. CAT (catalase) (Warren et al., 2010) fibroblast-associated gene is top shared gene (9/13 cell lines) in our Inherited up-regulated gene group (Supplementary Table 3a). Results of another functional analysis (Lister et al., 2011) overlapped with 5 genes of our list of Inherited Genes. This finding let us conclude about the higher plasticity of Induced Genes to reprogramming and consequent differentiation into somatic cell in comparison with Inherited somatic memory genes.

Transcriptional analysis of Induced Genes category revealed similar fraction of genes with predicted transcription factor binding motifs in up- and down-regulated groups in each cell line. Stronger binding affinity was observed in down-regulated groups of all cells suggesting that stochastic genes silencing should be the major problem to overcome for more successful reprogramming. Predicted NANOG binding motif in the proximity (−500 to +500) of TSS in the most cell lines leads us to the conclusion

about its ectopic activation. OCT4 and KLF4 are comparatively regular in binding allocation, while SOX2 and c-Myc did not show any consistency in our study. All predicted TFBS are distantly allocated, which is consistent with the recent experimental publication (Soufi et al., 2012).

Concerning the general characteristics of RRGs on the pathway level, Inherited Gene category included cancer and apoptosis-related pathways, such as focal adhesion, p53 signaling pathway, which also observed in the recent results (Soufi et al., 2012). This may affect unwanted tumorigenic propensity of iPS cells and further experimental verification of this issue is required. Induced Gene category was enriched in calcium-signaling pathway, cell adhesion, PPAR signaling, and tight junction. These pathways may contribute to the embryogenesis, development, and immune response, but biological implication of such differential expressions is yet to be elucidated. It is evident that substantial numbers of genes are differentially expressed due to various factors leading to differences between iPS and ES cells.

Pertaining to the virus type two conclusions can be drawn. First, virus-free and lentivirus-derived iPS cell lines have lower number of down-regulated Inherited Genes, while retrovirus insertion in the promoters of genes seems to provoke strong inhibitory effect. Remarkably, virus-free and lentivirus-derived iPS cell lines have larger number of genes with bivalent (H3K4K27) modification status in ES cells, i.e., under these conditions somatic cells exhibit higher susceptibility to reprogramming than those generated through the retrovirus transduction. Second conclusion is that passage length anti-correlates with number of RRGs.

For further improvement of iPSC technology several factors should be taken into the consideration. Basing on the status of gene in somatic cell of origin two natures of RRGs should be considered: Induced Genes active in somatic substrate and Inherited Genes inactive in somatic substrate. Tiny fraction of Induced Genes (bivalent H3K4K27, LCP/ICP) was identified up-/down-regulated in iPS-derived somatic cell (Gupta et al., 2010). This means that they might be activated during the first stage of reprogramming (Soufi et al., 2012), which make us suggest that longer passage can help to resolve this problem. More attention should be paid to the Inherited Genes category (univalent H3K4, HCP), retaining somatic cell transcriptional memory. They were abundantly found in iPS-derived somatic cell (Gupta et al., 2010) and overlap with RRGs from other studies (Chin et al., 2010; Newman and Cooper, 2010; Lister et al., 2011; Ohi et al., 2011). Active and demethylated High CpG density promoters might attract retrovirus insertion, which causes consequent silencing of the promoter through *de novo* methylation. Virus-free or miRNA-mediated (Ankye-Danso et al., 2011) reprogramming may be more plausible in the future. For the future work it is important to identify key transcription factors within Inherited Genes category to be able to reduce/block their activity. Donor age and developmental stage is important for the selection of somatic cell substrate. While heterogeneous tissue culture does not simply reflect the epigenetics status of the substrate cell, several reports indicate that somatic cell/progenitor cells can be epigenetically favored substrates for nuclear resetting (Aasen et al., 2008; Silva et al., 2008).

The influence of RRGs on the intactness of function after the consequent differentiation iPSCs in organs and tissues is extremely important for the validation and standardization of iPSC technology, and our results can be a help for this.

## ACKNOWLEDGMENTS

I wish to express my special gratitude to Prof. Hiroaki Kitano (Sony CSL) for his continuous support of this work and insightful discussions. I would like to thank Prof. Kazuhiro Sakurada (Sony CSL) and Prof. Mario Tokoro (Sony CSL) for the exchange of ideas on the future of iPS cell reprogramming technology. My special appreciations are addressed to Prof. Tomo Saric (Institute for Neurophysiology, University of Cologne) for the experimental verification of fibroblast-associated reprogramming resistant genes identified in both studies and critical reading of the manuscript. I am thankful to Prof. Edgar Wingender (Department of Bioinformatics, University Medical Center Göttingen) for critical reading and commenting on the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/Systems_Biology/10.3389/fphys.2013.00007/abstract

## REFERENCES

Aasen, T., Raya, A., Barrero, M. J., Garreta, E., Consiglio, A., Gonzalez, F., et al. (2008). Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes. *Nat. Biotechnol.* 26, 1276–1284.

Ankye-Danso, F., Trivedi, C. M., Juhr, D., Gupta, M., Cui, Z., Tian, Y., et al. (2011). Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell* 8, 376–388.

Aoi, T., Yae, K., Nakagawa, M., Ichisaka, T., Okita, K., Takahashi, K., et al. (2008). Generation of pluripotent stem cells from adult mouse liver and stomach cells. *Science* 321, 699–702.

Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucl. Acids Res.* 34, W369–W373.

Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122, 947–956.

Chao, W., and D'Amore, P. A. (2008). IGF2, epigenetic regulation and role in development and disease. *Cytokine Growth Factor Rev.* 19, 111–120.

Chin, M. H., Mason, M. J., Xie, W., Volinia, S., and Singer, M. (2009). Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* 5, 111–123.

Chin, M. H., Pellegrini, M., Plath, K., and Lowry, W. E. (2010). Molecular analyses of human induced pluripotent stem cells and embryonic stem cells. *Cell Stem Cell* 7, 263–269.

Davis, C. D., and Uthus, E. O. (2003). Dietary folate and selenium affect dimethylhydrazine-induced aberrant crypt formation, global DNA methylation and one-carbon metabolism in rats. *J. Nutr.* 133, 2907–2914.

Eckhardt, F. (2006). DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.* 12, 1378–1385.

Feng, Q., Lu, S. J., Klimanskaya, I., Gomes, I., Kim, D., Chung, Y., et al. (2009). Hemangioblastic derivatives from human induced pluripotent stem cells exhibit limited expansion and early senescence. *Stem Cells* 28, 704–712.

Goldman, J. E., Herbst, A., Schmidt, N. O., Aldenoff, J. B., Laurent, L.

C., Papapetrou, E. P., et al. (2011). A bioinformatic assay for pluripotency in human cells. *Nat. Methods* 8, 315–317.

Gupta, M. K., Illich, D. J., Gaarz, A., Matzkies, M., Hguemo, F., Pfannkuche, K., et al. (2010). Global transcriptional profiles of beating clusters derived from human induced pluripotent stem cells and embryonic stem cells are highly similar. *BMC Dev. Biol.* 10:98. doi: 10.1186/1471-213X-10-98

Gurdon, J. B., and Melton, D. A. (2008). Nuclear reprogramming in cells. *Science* 322, 1811–1815.

Haig, D. (2004). Genomic imprinting and kinship: how good is the evidence? *Annu. Rev. Genet.* 38, 553–585.

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nat. Protoc.* 4, 44–57.

Inoue, K., Kohda, T., Lee, J., Ogonuki, N., Mochida, K., Noguchi, Y., et al. (2002). Faithful expression of imprinted genes in cloned mice. *Science* 295, 297.

Jiang, J., Ding, G., Lin, J., Zhang, M., Shi, L., Lv, W., et al. (2011). Different developmental potential of pluripotent stem cells generated by different reprogramming strategies. *J. Mol. Cell Biol.* 3, 197–199.

Kel, A., Voss, N., Jauregui, R., Kel-Margoulis, O., and Wingender, E. (2006). Beyond microarrays: finding key transcription factors controlling signal transduction pathways. *BMC Bioinformatics* 7(Suppl 2):S13. doi: 10.1186/1471-2105-7-S2-S13

Kim, D., Kim, C. H., Moon, J. I., Chung, Y. G., Chang, M. Y., Han, B. S., et al. (2009). Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. *Cell Stem Cell* 4, 472–476.

Kim, K., Doi, A., Wen, B., Ng, K., and Zhao, R. (2010). Epigenetic memory in induced pluripotent stem cells. *Nature* 467, 285–290.

Lewinski, M. K., Yamashita, M., Emerman, M., Ciuffi, A., Marshall, H., Crawford, G., et al. (2006). Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathog.* 2:e60. doi: 10.1371/journal.ppat.0020060

Lister, R., Pelizzola, M., Kida, Y. S., Hawkins, R. D., Nery, J. R., Hon, G., et al. (2011). Hotspots of aberrant epigenomics reprogramming in human induced pluripotent stem cell. *Nature* 471, 68–73.

Lowry, W. E., Richter, L., Yachechko, R., Pyle, A. D., Tchieu, J., Clark, A. T., et al. (2008). Generation of human induced pluripotent stem cells from 17 dermal fibroblasts. *PNAS* 105, 2883–2888.

Maherali, N., Ahfeldt, T., Rigamonti, A., Utikal, J., and Cowan, C. (2008). A high-efficiency system for the generation and study of human induced pluripotent stem cells. *Cell Stem Cell* 3, 340–345.

Masaki, H., Ishikawa, T., Takahashi, S., Okumura, M., Sakai, N., Haga, M., et al. (2008). Heterogeneity of pluripotent marker gene expression in colonies generated in human iPS cell induction culture. *Stem Cell Res.* 1, 105–115.

Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454, 766–770.

Newman, A. M., and Cooper, J. B. (2010). Lab-specific gene expression signatures in pluripotent stem cells. *Cell Stem Cell* 7, 258–262.

Ogura, A., Inoue, K., Ogonuki, N., Lee, J., Kohda, T., and Ishino, F. (2002). Phenotypic effects of somatic cell cloning in the mouse. *Cloning Stem Cells* 4, 397–405.

Ohi, Y., Qin, H., Hong, C., Blouin, L., Polo, M. J., Guo, T., et al. (2011). Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. *Nat. Cell Biol.* 13, 541–550.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., et al. (2007). Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation inhuman embryonic stem cells. *Cell Stem Cell* 1, 299–312.

Polo, J. M., Liu, S., Figueroa, M. E., Kulatert, W., Eminli, S., Tan, K. Y., et al. (2010). Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat. Biotechnol.* 28, 848–855.

Polouliakh, N., Konno, M., Horton, P., and Nakai, K. (2005). Parameter landscape analysis for common motif discovery programs. *LNCS* 3318, 79–87.

Saxonov, S., Berg, P., and Brutlag, D. L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U.S.A.* 103, 1412–1417.

Silva, J., Barrandon, O., Nichols, J., Kawaguchi, J., Theunissen, T. W., and Smith, A. (2008). Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol.* 6:e253. doi: 10.1371/journal.pbio.0060253

Soldner, F., Hockemeyer, D., Beard, C., Gao, Q., Bell, G. W., Cook, E. G., et al., (2009). Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. *Cell* 136, 964–977.

Soufi, A., Donahue, G., and Zaret, K. S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* 151, 994–1004.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., et al. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131, 861–872.

Wakayama, T., Perry, A. C., Zuccotti, M., Johnson, K. R., and Yanagimachi, R. (1998). Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* 394, 369–374.

Warren, L., Manos, P. D., Ahfeldt, T., Loh, Y. H., Li, H., Lau, F., et al. (2010). Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified RNA. *Cell Stem Cell* 7, 618–630.

Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Pääbo, S., Rebhan, M., et al. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* 39, 457–466.

Wingender, E., Dietze, P., Karas, H., and Knuppel, R. (1996). TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucl. Acids Res.* 24, 238–241.

Woltjen, K., Michael, I. P., Mohseni, P., Desai, R., and Mileikovsky, M. (2009). piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* 458, 766–770.

Zhao, Y., Yin, X., Qin, H., Zhu, F., Liu, H., Yang, W., et al. (2008). Two supporting factors greatly improve the efficiency of human iPSC generation. *Cell Stem Cell* 3, 475–479.

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Received: 10 December 2012; paper pending published: 22 December 2012; accepted: 09 January 2013; published online: 30 January 2013.*

*Citation: Polouliakh N (2013) Reprogramming resistant genes: in-depth comparison of gene expressions among iPS, ES, and somatic cells. Front. Physio. 4:7. doi: 10.3389/fphys.2013.00007*

*This article was submitted to Frontiers in Systems Biology, a specialty of Frontiers in Physiology.*

# Minichromosome Maintenance 2 Bound with Retroviral Gp70 Is Localized to Cytoplasm and Enhances DNA-Damage-Induced Apoptosis

**Shinya Abe[1], Morito Kurata[1], Shiho Suzuki[1], Kouhei Yamamoto[1], Ken-ichi Aisaki[2], Jun Kanno[2], Masanobu Kitagawa[1]***

**1** Department of Comprehensive Pathology, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan, **2** Division of Cellular and Molecular Toxicology, National Institute of Health Sciences, Tokyo, Japan

## Abstract

The interaction of viral proteins with host-cellular proteins elicits the activation of cellular signal transduction pathways and possibly leads to viral pathogenesis as well as cellular biological events. Apoptotic signals induced by DNA-damage are remarkably up-regulated by Friend leukemia virus (FLV) exclusively in C3H hosts; however, the mechanisms underlying the apoptosis enhancement and host-specificity are unknown. Here, we show that C3H mouse-derived hematopoietic cells originally express higher levels of the minichromosome maintenance (MCM) 2 protein than BALB/c- or C57BL/6-deriverd cells, and undergo more frequent apoptosis following doxorubicin-induced DNA-damage in the presence of the FLV envelope protein gp70. Dual transfection with *gp70/Mcm2* reproduced doxorubicin-induced apoptosis even in BALB/c-derived 3T3 cells. Immunoprecipitation assays using various deletion mutants of MCM2 revealed that gp70 bound to the nuclear localization signal (NLS) 1 (amino acids 18–24) of MCM2, interfered with the function of NLS2 (amino acids 132–152), and suppressed the normal nuclear-import of MCM2. Cytoplasmic MCM2 reduced the activity of protein phosphatase 2A (PP2A) leading to the subsequent hyperphosphorylation of DNA-dependent protein kinase (DNA-PK). Phosphorylated DNA-PK exhibited elevated kinase activity to phosphorylate P53, thereby up-regulating *p53*-dependent apoptosis. An apoptosis-enhancing domain was identified in the C-terminal portion (amino acids 703–904) of MCM2. Furthermore, simultaneous treatment with FLV and doxorubicin extended the survival of SCID mice bearing 8047 leukemia cells expressing high levels of MCM2. Thus, depending on its subcellular localization, MCM2 plays different roles. It participates in DNA replication in the nucleus as shown previously, and enhances apoptosis in the cytoplasm.

**Citation:** Abe S, Kurata M, Suzuki S, Yamamoto K, Aisaki K-i, et al. (2012) Minichromosome Maintenance 2 Bound with Retroviral Gp70 Is Localized to Cytoplasm and Enhances DNA-Damage-Induced Apoptosis. PLoS ONE 7(6): e40129. doi:10.1371/journal.pone.0040129

**Editor:** Junji Yodoi, Institute for Virus Research, Laboratory of Infection and Prevention, Japan

**Received** March 2, 2012; **Accepted** June 1, 2012; **Published** June 29, 2012



**Funding:** This work was supported by a grant-in-aid (21590432) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan and by Health Sciences Research Grant H18-Kagaku-001 from the Ministry of Health, Labour and Welfare of Japan. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: masa.pth2@tmd.ac.jp

## Introduction

Because ionizing irradiation (IR) and chemical agents such as doxorubicin exhibit cell-killing activity by inducing double-strand breaks (DSBs) and *p53*-dependent apoptosis, they have been considered therapeutic tools against malignant tumors [1–5]. To protect normal cells from injury, tumor cell-specific induction of apoptosis would be one of the most important properties of anti-tumor therapeutics [6,7]. To regulate the *p53*-dependent apoptosis caused by DNA-damage, an understanding of upstream activators or regulators of P53 would be vital. These pathways partly involve the phosphatidylinositol 3-kinase (PI3K)-related protein kinase (PIKK) family of enzymes [8], including ataxia telangiectasia (ATM), ATM and Rad3-related (ATR), and DNA-dependent protein kinase (DNA-PK) [9–13].

Viral infections are known to modify cellular processes related to DNA-damage responses or DNA synthesis [14–16]. We have previously shown that Friend leukemia virus (FLV) infection markedly enhances the IR-induced apoptosis of hematopoietic cells in C3H mice via P53, ATM, and DNA-PK [17]. Mice infected with FLV and then treated with a low dose of total body irradiation (TBI) exhibit severe anemia. However, *p53* knockout mice, *Atm* knockout mice, and DNA-PK-deficient SCID mice with a C3H background do not exhibit this phenotype. A comparison of the apoptotic signals after FLV infection, TBI, or FLV+TBI treatment of these mice revealed that ATM is necessary for the general signal transduction of TBI-induced apoptosis [18], while DNA-PK plays a specific role in enhancing *p53*-dependent apoptosis following FLV infection [19,20].

The enhancement of *p53*-dependent apoptosis occurs almost exclusively in the C3H strain of mice [21]. In relation to this host-specific apoptosis-enhancement, we have previously demonstrated that the FLV-derived envelope protein gp70 enhances cellular apoptotic signaling in association with host-specific overexpressed proteins, including the minichromosome maintenance (MCM) 2 protein, resulting in the activation of DNA-PK, which phosphorylates P53 [22]. MCM2 is one of a set of 6 proteins (MCM complex; MCM2-7) that play essential roles in DNA replication [23]. The MCM complex associates with the origins of DNA

replication to form part of the pre-replicative complex (preRC) [24]. Activation of the MCM complex by cyclin-dependent kinases leads to the initiation of DNA synthesis and MCM proteins also act as a replicative helicase to unwind DNA at replication forks during DNA synthesis [25,26]. The MCM complex contains a nuclear localization signal (NLS) and a nuclear export signal (NES) [27]. The NLS is split between MCM2 and MCM3 and the NES is located in MCM3 adjacent to the NLS sequence. The transport of all MCM proteins is interdependent, suggesting that nuclear import requires the formation of the hexameric complex, which would result in the assembly of a complete NLS [28,29]. MCM proteins are expressed in cycling cells but are down-regulated and dissociated from the chromatin in quiescent cells [30]. Thus, detection of MCM proteins has emerged as a method for evaluating the proliferative state and growth fraction in dynamic cell populations. Indeed, elevated expression of several members of the MCM complex has been reported in various malignant tumors [31,32]. Furthermore, studies with human samples have indicated the utility of MCM2 as a proliferation marker, and a high level of MCM2 expression in malignant tumors has been associated with several clinicopathological parameters, such as advanced tumor grade, advanced stage, and poor prognosis [33–36]. Thus, MCM2 usually acts to support cellular proliferation. However, as described above, MCM2 enhances TBI-induced apoptosis in the presence of gp70. To determine importance of such contradictory functions of the MCM2 protein in the regulation of cellular dynamics, the molecular mechanisms underlying MCM2-induced apoptosis and MCM2-gp70 interaction need to be elucidated. An understanding of the overall functions of MCM2 would enable the molecular targeting of specific functions possibly to regulate cellular proliferation/apoptosis in a cell type-specific manner and develop a novel strategy to control tumor cell growth.

## Results

### Doxorubicin-induced Apoptosis of FLV-infected Cells Correlates with High Levels of *Mcm2 in Vivo*

In previous studies, TBI caused prominent apoptosis in the bone marrow cells of FLV-infected C3H mice, but not FLV-infected BALB/c and C57BL/6 mice [17]. From a therapeutic perspective, systemic distribution of the effects of DNA-damage would be more easily achieved by chemical agents than IR. Therefore, to determine whether DNA-damaging agents enhanced apoptosis to similar extents in FLV-infected mice of different strains, uninfected or FLV-infected BALB/c, C57BL/6, and C3H mice were intraperitoneally administered with a low dose of doxorubicin or PBS, and the apoptotic cell ratio was measured in the bone marrow and spleen. In FLV-infected BALB/c and C57BL/6 mice, the apoptotic cell ratios after treatment with doxorubicin were similar to the ratios in uninfected mice (Figure 1A, B). On the other hand, FLV-infected doxorubicin-treated C3H mice exhibited significantly higher ratios with uninfected mice (Figure 1C). Thus, we could generalize as to the effects of DNA-damage by IR and chemical agents on the enhancement of apoptosis by FLV-infection in hematopoietic organs.

Next, we examined the expression of *Mcm2* mRNA in the bone marrow and spleen of uninfected and FLV-infected BALB/c, C57BL/6, and C3H mice. *Mcm2* levels were significantly higher in the bone marrow cells of C3H mice than in BALB/c and C57BL/6 mice (Figure 1D). Spleen *Mcm2* levels were also higher in C3H mice than in BALB/c and C57BL/6 mice. Furthermore, in C3H mice, the spleen *Mcm2* levels were elevated by FLV-infection (Figure 1E). Similar trends were observed across all the inbred strains tested. These results suggest that doxorubicin treatment induces significant apoptosis in FLV-infected C3H mice in association with higher levels of *Mcm2*. Moreover, we performed a comparative GeneChip analysis using RNA isolates from mouse spleen and identified several genes that exhibited various expression patterns in the different mouse strains (Figure 1F-L). *Mcm2* expression was higher in C3H mice than in C57BL/6 mice, and *Mcm2* expression was elevated by FLV-infection (Figure 1G). Genes that exhibited expression patterns similar to that of *Mcm2* are listed in Table S1.

### Dual Transfection with *Mcm2/gp70* Enhances DNA-damage-induced Apoptosis in BALB/c-derived 3T3 Cells

To investigate whether apoptosis enhancement was related to the high levels of *Mcm2* in FLV-infected cells, we analyzed doxorubicin-induced apoptosis sensitivity in *Mcm2* and/or *gp70*-transfected 3T3 cells. First, the expression of *Mcm2* was analyzed in each mouse cell line. BALB/c-derived 3T3 cells and primary cultured BALB/c-fibroblasts expressed low levels of *Mcm2* compared to C3H-derived 8047 cells, 32D cells and primary cultured C3H-fibroblasts (Figure 2A).

Next, the viability and apoptotic cell ratios of 3T3 cells were evaluated after doxorubicin treatment. *Gp70* plus *Mcm2*-transfected 3T3 cells exhibited a significant decrease in viability and an increase in apoptotic cell ratio compared to control cells, whereas cells transfected with *gp70* or *Mcm2* exhibited no significant change in viability and apoptotic cell ratio (Figure 2B, C). Gp70 and/or MCM2 protein levels following *gp70*- and/or *Mcm2*-transfection were similar in all the experimental groups (Figure 2D). Next, we knocked down the expression of *Mcm2* in BaF3 and 32D cells using siRNA. The 32D cell line, with a high level of endogenous gp70 expression, was established from FLV-infected C3H mouse bone marrow [37] (Figure 2E). *Mcm2* knockdown significantly reduced *Mcm2* mRNA expression and apoptotic cell ratio of 32D cells treated with doxorubicin in contrast to the non-remarkable change in the apoptotic cell ratio of BaF3 cells (Figure 2F). These results suggest that the host-specific enhancement of DNA-damage-induced apoptosis is associated with the higher level of *Mcm2* expression in C3H-derived cells.

### Gp70 Directly Binds to the N-terminal Portion of MCM2

To examine the molecular interactions between MCM2 and gp70, immunoprecipitation experiments were performed. We generated plasmids encoding HA-tagged full-length MCM2 (MCM2-FL) and various deletion mutants: MCM2-ΔC, MCM2-ΔN, MCM2-N and MCM2-C (Figure 3A). Each of these plasmids was transfected into 3T3 cells along with *FLAG*-tagged *gp70*. Irrespective of doxorubicin treatment, gp70 interacted with MCM2-FL, MCM2-ΔC, and MCM2-N, but not with MCM2-ΔN or MCM2-C (Figure 3B, C). These results indicate that gp70 associates with the N-terminal portion of MCM2. Gp70 binding inhibited the formation of the MCM complex (Figure S1). As shown in Figure 3B and 3C, the size of MCM2-N was larger than the expected size. Generally, phosphorylated proteins are sometimes larger than their unphosphorylated counterparts [38,39]. Indeed, the N-terminal portion of MCM2 possesses many phosphorylation sites [40]. Therefore, the apparent molecular weight of MCM2-N may be higher than expected. Further, MCM2-C does not have as many phosphorylation sites [40]. As a result, MCM2-N may appear larger than MCM2-C.

We also generated plasmids encoding a FLAG-tagged gp70 deletion mutant (Figure S2A) and performed a similar pull-down assay after co-transfection with HA-tagged *Mcm2*-FL. MCM2 bound to the middle portion of gp70 (Figure S2B, C) and enhanced apoptosis in response to doxorubicin (Figure S2D, E).