Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Nat Biotechnol. Author manuscript; available in PMC 2016 May 01.

Published in final edited form as:

Nat Biotechnol. 2015 November ; 33(11): 1173–1181. doi:10.1038/nbt.3388.

A comparison of genetically matched cell lines reveals the equivalence of human iPSCs and ESCs

Jiho Choi[1,2,3,10], **Soohyun Lee**[4,10], **Kendell Clement**[5], **William Mallard**[5,6], **Guidantonio Malagoli Tagliazucchi**[4,8], **Hotae Lim**[9], **In Young Choi**[9], **Francesco Ferrari**[4], **Alex Tsankov**[5], **Ramona Pop**[2], **Gabsang Lee**[9], **John Rinn**[5,6,7], **Alexander Meissner**[2,5,6], **Peter J. Park**[4,*], and **Konrad Hochedlinger**[1,2,3,*]

[1]Massachusetts General Hospital, Department of Molecular Biology, Cancer Center and, Center for Regenerative Medicine, 185 Cambridge Street, Boston, MA 02114, USA

[2]Harvard Stem Cell Institute, 1350 Massachusetts Avenue, Cambridge, MA 02138, USA

[3]Howard Hughes Medical Institute and Department of Stem Cell and Regenerative Biology, 7 Divinity Avenue, Cambridge, MA 02138, USA

[4]Center for Biomedical Informatics, Harvard Medical School, Boston, MA, USA

[5]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[6]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA

[7]Beth Israel Deaconess Medical Center, Boston, MA 02115, USA

[8]Center for Genome Research, Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy

[9]Institute for Cell Engineering, Department of Neurology, The Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

The equivalence of human induced pluripotent stem cells (hiPSCs) and human embryonic stem cells (hESCs) remains controversial. Here we use genetically matched hESC and hiPSC lines to assess the contribution of cellular origin (hESC vs hiPSC), the Sendai virus (SeV) reprogramming method and, genetic background to transcriptional patterns while

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Correspondence: peter_park@harvard.edu (P.J.P.), khochedlinger@helix.mgh. harvard.edu (K.H.).

[10]These authors contributed equally to this work

AUTHOR CONTRIBUTIONS

J.C., S.L., P.J.P. and K.H. conceived the experiments, interpreted results and wrote the manuscript. J.C. generated all HUES2- and HUES3-derived *in vitro*-differentiated fibroblasts and iPSCs. A.M. and J.R. provided RNA-sequencing data from hESCs and hiPSCs generated with retroviral vectors. J.C. performed AP staining, immunostaining, lactate production and glucose uptake assays, Western blot, RT-PCR and qPCR analyses. S.L., W.M., G.M.T., F.F. and P.J.P. performed bioinformatics analysis of RNA-sequencing data. H.L, I.Y.C. and G.L. performed neural differentiation experiments and marker analyses differentiated cells. R.P. conducted the Scorecard assay, which was bioinformatically analyzed by A.T. K.C. performed bioinformatics analysis of RRBS data.

GEO accession code for RNA-sequencing data: GSE73211

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

controlling for cell-line clonality and sex. We find that transcriptional variation originating from genetic background dominates over variation due to cellular origin or SeV infection. Moreover, the 49 differentially expressed genes we detected between isogenic hESCs and hiPSCs neither predicted functional outcome nor distinguished an independently derived, larger set of unmatched hESC and hiPSC lines. We conclude that hESCs and hiPSCs are molecularly and functionally equivalent and cannot be distinguished by a consistent gene expression signature. Our data further imply that genetic background variation is a major confounding factor for transcriptional comparisons of pluripotent cell lines, explaining some of the previously observed expression differences between unmatched hESCs and hiPSCs.

The question of whether hiPSCs, derived from somatic cells by overexpression of the transcription factors Oct4, Klf4, Sox2 and c-Myc (OKSM)[1], are equivalent to hESCs, the gold standard of pluripotent cell lines, is becoming increasingly urgent as patient-specific hiPSCs are advanced toward clinical application[1-4]. Initial studies showed that hESC and hiPSC lines are fundamentally different at the transcriptional level, whereas subsequent work concluded that they are virtually indistinguishable when comparing larger sample sets[5-7]. More recent reports using refined gene expression analyses found small sets of differentially expressed genes (DEGs)[8-10]. However, the origins of these DEGs, their consistency across independent studies and their impact on the differentiation potential of hiPSC lines remain unclear. Transcriptional patterns are influenced by numerous biological and technical parameters that may confound results. The reprogramming method, including the choice of integrating versus non-integrating factor delivery systems, can alter gene expression in iPSCs[11-13]. Likewise, genetic background may influence transcriptional signatures in pluripotent cell lines since iPSCs derived from different individuals are reportedly more divergent than iPSCs derived from the same individual. The difference between the clonal origin of hiPSC lines, derived from single somatic cells, and the polyclonal origin of most hESC lines may also introduce transcriptional variation[14]. An additional consideration is the sex of cell lines and defects in X chromosome reactivation in female hiPSCs[17,18]. Some of these variables have been addressed in previous reports[11,12,15,16], but, to our knowledge, no comparative study of hESCs and hiPSCs has accounted for all of them.

We previously showed that comparing genetically matched mouse ESC and integration-free iPSC lines eliminates most of the transcriptional variation observed between unmatched cell lines[16]. Although we could not identify consistent transcriptional differences between mouse ESC and iPSC lines, we discovered a small group of transcripts that was aberrantly silenced in a subset of iPSC lines, which adversely affected their developmental potential. Here we extend our analyses to the human system and ask whether molecular differences can be identified in hiPSC lines relative to hESC lines that cannot be attributed to the SeV reprogramming method, genetic background, clonal origin or sex, and whether any such differences impact functional outcomes.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

RESULTS

Approach to generate isogenic hESCs and hiPSCs

To compare hESCs with genetically matched hiPSC lines devoid of viral integrations, we generated hiPSCs from *in vitro*-differentiated hESCs using a non-integrating Sendai virus (SeV)-based reprogramming system[19]; SeV is an RNA virus that is diluted from infected cells in a replication-dependent manner, leaving no genetic footprint behind (Fig. 1A,B). We chose two well-characterized hESC lines, HUES2 and HUES3[20], for these experiments. We selected male hESC lines because female iPSCs can exhibit defects in X chromosome reactivation[17,18], which might confound subsequent interpretations[9,21].

First, we subcloned each line in order to ensure genetic and epigenetic homogeneity of cells and to properly control for the clonal origin of hiPSCs (Fig. 1A). We differentiated one hESC subclone from each background by switching cells to serum-containing medium without basic fibroblast growth factor (bFGF), which is critical for the maintenance of hESCs, and sorting fibroblast-like cells based on $CD90^{+}$/$TRA\text{-}1\text{-}81^{-}$ expression (Fig. 1A,C). These fibroblast-like cells, which resemble primary human fibroblasts by morphological criteria (Fig. 1C), did not form Alkaline Phosphatase (AP)-positive colonies in hESC media, indicating successful differentiation and the absence of residual pluripotent cells in the culture (Supplementary Fig. 1A). Analysis of global gene expression by RNA-sequencing revealed that the fibroblast-like cells were highly similar to dermal fibroblasts but distinct from pluripotent stem cell lines (Supplementary Fig. 1B). Pluripotency-associated promoters, such as *POU5F1, LEFTY1, TDGF1,* and *SCNN1A,* were re-methylated and decreased in expression levels whereas fibroblast-specific promoters such as *TMEM173*, *EMILIN1, LMNA,* and *RIN2* were demethylated and regained expression in fibroblast-like cells (Fig. 1D). In a final step, the fibroblast-like cultures were reprogrammed into hiPSCs by infecting the cells with SeV vectors expressing *OKSM*, as previously reported[19] (Fig. 1A). Emerging colonies were isolated after ~3 weeks, expanded and confirmed to be positive for AP activity and endogenous OCT4 expression, indicating successful reprogramming (Fig. 1C). Moreover, we ensured loss of SeV expression in all lines, demonstrating reprogramming factor independent self-renewal (Supplementary Fig. 1C,D).

Genetic background drives transcriptional variation

First we studied whether the SeV reprogramming method affects global transcription. The parental hESC subclones were infected with GFP-expressing SeV (SeV-GFP) and passaged until GFP fluorescence was no longer detectable before analyzing cell lines by RNA-sequencing (Fig. 1A and 2A). We found a common set of 63 genes that was differentially expressed between three uninfected hESC subclones (hESC SCs) and three SeV-GFP infected hESC subclones (hESC GFPs) from each genetic background, which demonstrates that viral infection itself leads to subtle but statistically significant transcriptional changes that persist after viral loss (Fig. 2B). This 63-DEG set consistently separated hESC SC lines from hESC GFP lines (Fig. 2C). Gene Ontology terms significantly enriched among these 63 DEGs are related to transcription, DNA binding, and development (Supplementary Fig. 1E). Based on these observations, we decided to use expression data from hESC GFP lines as controls for all subsequent comparisons with SeV-generated hiPSC lines.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

A comparison of the transcriptional profiles of hESC subclones (hESC SCs and hESC GFPs), *in vitro*-differentiated fibroblasts and derivative hiPSCs by unsupervised clustering showed the largest differences between pluripotent cell lines and differentiated cell types, consistent with previous observations[5,15,22,23] (Supplementary Fig. 2A). Likewise, global methylation analysis of representative samples by reduced representation bisulfite sequencing (RRBS) separated pluripotent cells from *in vitro*-differentiated fibroblasts, indicating distinct epigenetic states (Supplementary Fig. 1F). Notably, we observed a clear segregation of all pluripotent samples into two transcriptionally related groups, irrespective of whether cell lines were infected with SeV or not (Fig. 2D, expanded from Supplementary Fig. 2A). This segregation could not be explained by the cellular origin of cell lines from embryos (hESCs) or somatic cells (hiPSCs) but instead correlated with the genetic background of each line. That is, HUES2-derived hESC subclones clustered with HUES2-derived hiPSC lines whereas HUES3-derived hESC subclones clustered with HUES3-derived hiPSC lines. Consistent with this finding, overall transcriptional variation between groups of genetically matched hESC and hiPSC lines was significantly lower than that between unmatched cell lines (Supplementary Fig. 2B). Moreover, transcriptional variation within groups of genetically matched hiPSC or hESC lines was similar, indicating that hiPSCs and hESCs are equally variable (Supplementary Fig. 2C). Of note, the number of promoters differentially methylated between unmatched pluripotent cell lines was approximately twice as high (2,610) as that between matched pluripotent cell lines (1,205), suggesting that genetic background also influences epigenetic patterns in hESCs and hiPSCs (Supplementary Fig. 2D). We conclude that genetic background is a major driver of transcriptional and epigenetic differences between pluripotent cell lines, whereas the SeV reprogramming method introduces more subtle yet stable transcriptional changes in hiPSCs.

Expression differences between matched hESCs and hiPSCs

Although genetic background accounted for most transcriptional differences among the analyzed pluripotent cell lines, we noticed that hESCs clustered with each other and separately from hiPSCs within a given background, suggesting subtle but consistent transcriptional differences that reflect distinct cellular origins (Fig. 2D). To identify any DEGs that distinguish hESC from hiPSC lines independent of SeV infection and genetic background, we compared transcriptional profiles of hiPSC lines with those of genetically matched hESC GFP lines. This analysis revealed that 52 and 91 genes were up- and down-regulated, respectively, in hiPSC lines derived from the HUES2 background, whereas 77 and 426 genes were up- and down-regulated in hiPSC lines derived from the HUES3 background, respectively. Forty-nine genes were commonly dysregulated in both genetic backgrounds (Fig. 3A). Considering the good depth of our RNA-seq data (~40 million mapped reads per sample on average) (Supplementary Fig. 2E), it is highly unlikely that this small number of DEGs was due to low sensitivity. As expected, the 49-DEG signature reliably separated our hiPSC lines from our hESC lines (Fig. 3B).

We did not detect any Gene Ontology term that was significantly enriched among the 49 DEGs. A comparison of our DEG set with 8 different protein interaction databases, including BIND, DIP, MINT and REACTOME INTERACTION using DAVID, also showed no significant enrichment (data not shown). Notably, 48 of 49 DEGs were downregulated in

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

hiPSCs relative to hESCs (Fig. 3C). This raised the possibility that the DEGs were silenced in fibroblast-like cells and were not properly reactivated in derivative hiPSCs. However, the expression levels of the DEGs in fibroblast-like cells did not show a consistent pattern, excluding incomplete reprogramming or the retention of 'epigenetic memory' (Fig. 3C).

We next asked whether the DEGs have functional consequences. We focused on two DEGs, *LDHA* and *SLC2A1* (also known as *GLUT1*), because of their strong basal expression in hESCs and reduced expression in all hiPSCs (Fig.3C,D). Both gene products are involved in energy metabolism; *LDHA* plays an important role in glycolysis by catalyzing the conversion of pyruvate to lactate[24,25], whereas *SLC2A1* facilitates glucose uptake in cells[26,27]. Accordingly, *LDHA* and *SLC2A1* are abundantly expressed in pluripotent cells, which produce energy through glycolysis[28] (Fig. 3C). Based on the down-regulation of these two genes in all examined hiPSC lines compared to hESC lines by RNA-seq and qPCR analyses (Fig. 3E), we hypothesized that hiPSC lines might be less glycolytic than hESC GFP lines. However, neither lactate production nor glucose uptake levels differed between isogenic hiPSC and hESC GFP lines (Fig. 3F). Further, there was no difference in LDHA protein levels despite the observed transcriptional differences (Fig. 3G). Thus, at least two of the 49 DEGs seem not to translate into functional differences, possibly owing to posttranscriptional compensatory mechanisms.

The low level of transcriptional differences between undifferentiated hESCs and hiPSCs does not exclude the existence of iPSC-specific aberrations that become detectable only after differentiation. We performed RNA-sequencing of fibroblast-like cells derived from 8 hESC subclones (2 hESC SC and 6 hESC GFP lines) and 6 hiPSC subclones using the same *in vitro* differentiation protocol as described above (Fig. 1A). Only two genes were consistently upregulated in hiPSC-derived fibroblast-like cells compared to hESC-derived fibroblast-like cells from both genetic backgrounds, and they did not overlap with the 49 DEGs between undifferentiated hESC and hiPSC lines (Supplementary Fig. 3A,B). However, HUES2-derived fibroblast-like cells tended to cluster together and apart from HUES3-derived fibroblast-like cells using PCA analysis (Supplementary Fig. 1B), which is consistent with the segregation of undifferentiated cells by genetic background. We infer that genetic background also drives transcriptional variation in differentiated cell populations, and that any transcriptional differences observed between undifferentiated hESC and hiPSC lines do not persist in differentiated fibroblast-like cells.

Dysregulation of genes in a subset of hiPSC lines

As most of the DEGs between undifferentiated hESC GFP and hiPSC lines produced low-abundance transcripts that were not obviously connected through a common biological process (Fig. 3C), we examined genes that were dysregulated in only a subset of hiPSC lines, which we refer to as inconsistently differentially expressed genes (iDEGs) (Supplementary Fig. 3C). We have previously shown that iDEGs between isogenic mouse ESCs and iPSCs could predict full developmental potential of subsets of iPSC lines[16]. Applying the same principle to our human data set, we found that 34 genes were upregulated, whereas 27 genes were downregulated in some of the HUES2-derived hiPSC lines when compared to genetically matched hESC GFP lines. Similarly, 9 genes were

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

upregulated and 32 genes were downregulated in some of the HUES3-derived hiPSC lines relative to matched hESC GFP controls (Supplementary Fig. 3C). Only eight iDEGs were dysregulated in both genetic backgrounds, and these were thus selected for further analysis (Fig. 4A and Supplementary Fig. 3C).

The iDEGs *IRX2* and *DPP10* have been linked to neural development or psychiatric disease[29-32] and IRX2 suppression reportedly impairs hESC differentiation into neural progenitors. Silencing of *IRX2* and *DPP10* in some of the hiPSC lines and none of the hESC lines (Fig. 4B) was confirmed by qPCR (Fig. 4C). However, the iDEGs did not affect the cells' potential to differentiate into neuroectodermal cells using a published protocol[33] (Fig. 4D), as determined by RNA expression analysis for *NESTIN, SOX1, PAX6*, and *FOXG1*, well- established markers of neuroectoderm differentiation from human pluripotent stem cells[34] (Fig. 4E). Consistent with this, PAX6 and SOX1 were equally expressed at the protein level during neural differentiation from hiPSC and hESC GFP lines (Fig. 4F and Supplementary Fig. 3D).

To determine whether hiPSCs exhibit biases in differentiation into other lineages, we evaluated their ability to generate ectodermal, endodermal and mesodermal derivatives by the Score card assay[6]. Briefly, hiPSC and hESC GFP lines from both genetic backgrounds were differentiated into embryoid bodies (EBs) before scoring for the expression of 77 developmental marker genes by qPCR. Hierarchical clustering of these data showed that all markers were expressed at similar levels in genetically matched cell lines (Fig. 4G). Thus, isogenic hESCs and hiPSCs appear to have equivalent potential to differentiate into cell types of the three germ layers.

Genetic background explains previous expression differences

We asked whether the 49 genes differentially expressed between our isogenic hESCs and hiPSCs are also dysregulated in hiPSC lines derived from primary somatic cells as well as in other published datasets. First, we reanalyzed a published set of unmatched hESC (n=18) and hiPSC (n=12) lines generated from primary fibroblasts using retroviral vectors, whose gene expression patterns were previously analyzed by microarrays[6]. Since many of the 49 DEGs were not covered in the available microarray data, we performed RNA-sequencing of these hESC and hiPSC lines, which offers increased sensitivity, especially for low-abundance transcripts[35,36]. However, unsupervised clustering was unable to separate these hESCs from hiPSCs (Fig. 5A). Although 3 DEGs (*RP11-1, MEG3, AL1327*) were identified between unmatched hESCs and hiPSCs, these were likely false positives based on permutation analysis. Indeed, supervised clustering of all samples with these 3 DEGs (data not shown) or an extended set of 16 DEGs using loosened criteria could not distinguish hESCs from hiPSCs (Fig. 5A and Supplementary Fig. 4A,D). Notably, our stringently defined 49-DEG signature was also unable to segregate the transcriptomes of this extended set of hESC and hiPSC lines (Fig. 5B).

Next, we determined the potential overlap between DEGs identified within our isogenic and unmatched hESC/hiPSC lines, and two previously reported sets of DEGs[8,10]. There was little to no overlap between DEGs discovered by independent laboratories (Fig. 5C and Supplementary Fig. 4B) and these DEGs could not distinguish hiPSC and hESC lines from

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the respective other data sets using supervised clustering (Supplementary Fig. 4C-I). Only two of our 49 DEGs (*MT1E, S100A14*) and two of our 8 iDEGs (*IRX2* and *DPP10*) overlapped with DEGs reported in ref. 10. Collectively, these data support the view that other parameters, such as reprogramming method, genetic background or sex, account for the majority of previously reported transcriptional differences between hESCs and hiPSCs.

In agreement with this conclusion, DEGs reported in ref. 10 distinguished our hESC and hiPSC cell lines by genetic background rather than cellular origin (Fig. 5D, left panel). In that study, multiple hiPSC lines generated from one man were compared to male and female hESC lines of different genetic backgrounds. Conversely, genes that distinguish our HUES2-derived and HUES3-derived pluripotent cell lines were able to separate the hESCs and hiPSCs in ref. 10 (Fig. 5D, right panel; Fig. 5E; Supplementary Fig. 4J,K).

In further support of the notion that genetic background profoundly influences transcriptional patterns in pluripotent cell lines, we found that DEGs that distinguish our HUES2-derived and HUES3-derived cell lines account for more transcriptional variation among our 30 unmatched hESCs and hiPSCs than do all genes or DEGs that distinguish our isogenic hESCs and hiPSCs or SeV-infected and uninfected hESCs (Fig. 5F). Taken together, these meta-analyses suggest that the main transcriptional differences between genetically unrelated hESC and hiPSC lines are primarily driven by genetic background rather than cellular origin or reprogramming method.

DISCUSSION

Here we show that isogenic male hESC and hiPSC lines are transcriptionally highly similar to one another, suggesting that genetic background variability and possibly sex differences account for most of the previously reported gene expression differences between hESCs and hiPSCs. This conclusion is particularly relevant to studies where only a limited number of hESC lines or a single iPSC donor individual was used, as imbalances in genetic background and sex may further inflate transcriptional differences[9,10,37,38]. Our finding that a previously reported set of DEGs between 4 hiPSC lines derived from a single individual and 4 hESC lines separated our hESC and hiPSC lines by genetic background rather than cellular origin supports this conclusion (Fig. 5D).

Our study reveals that a commonly used non- integrating reprogramming method can subtly but stably alter transcriptional patterns in iPSCs (Fig. 2B,C). However, the transcriptional signature introduced by SeV infection (63 DEGs) did not separate hESCs from hiPSCs previously generated with retroviral or episomal vectors, suggesting that each reprogramming system may introduce unique transcriptional alterations into iPSCs (Supplementary Fig. 4F,I). Whereas the molecular mechanisms of this observation remain to be elucidated, our findings highlight the importance of controlling for the method of iPSC induction when studying transcriptional patterns in iPSCs. Indeed, a recent comparison of hiPSCs generated with different OKSM delivery systems showed that hiPSCs derived with integrating vectors (e.g., retroviral transgenes) more often exhibit expression, methylation and differentiation defects compared to hiPSCs produced with non-integrating approaches[39].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

We identified only 49 DEGs that could distinguish hESCs and hiPSCs and 8 iDEGs that were dysregulated in a subset of hiPSC lines (Fig. 3C and 4A). This small number of genes contrasts with previous studies, which identified much larger numbers of DEGs when comparing unmatched hESCs and hiPSCs using a similar cutoff[5-7,10,13,22,23]. Of note, we found no evidence that two of the tested DEGs (*LDHA* and *SLC2A1*) and two of the tested iDEGs (*IRX2* and *DPP10*) predict functional outcome, i.e., energy production or differentiation potential into neural cells or EBs. These results therefore suggest that hESC and hiPSC lines are equivalent after accounting for genetic background differences. We surmise that the remaining DEGs we detected between isogenic hESCs and hiPSCs might represent transcriptional noise from weakly expressed genes. In support of this notion, the vast majority of the 49 DEGs was expressed at relatively low levels in our hESCs and hiPSCs and showed no overlap with previously reported gene expression signatures. However, we cannot exclude that the lack of an obvious phenotype with the abovementioned assays could be due to insufficient expression of the analyzed genes in undifferentiated hiPSCs or compensation by posttranscriptional mechanisms, as appears to be the case with LDHA (Fig. 3G). Alternatively, our metabolic and *in vitro*-differentiation assays may not have been sensitive enough to detect functional differences. Another possibility is that hiPSCs are distinguished from hESCs by epigenetic or genetic differences that do not manifest in the pluripotent state. However, our finding that fibroblast-like cells derived from all examined hESC and hiPSC lines show no discernable transcriptional differences argues against this explanation (Supplementary Fig. 3A,B). The fact that isogenic hESC and hiPSC lines exhibit equivalent differentiation potentials using either a directed or spontaneous differentiation paradigm further supports this interpretation (Fig. 4D-G). Critically, hiPSCs were derived from *in vitro*- differentiated fibroblasts in this study and, we can therefore not rule out that hiPSCs produced from primary cells accrue additional aberrations that cannot be recapitulated with our *in vitro* differentiation approach.

Our results may have implications for the use of iPSC technology in disease modeling approaches, where hiPSC lines from healthy individuals are usually compared to hiPSC lines from affected individuals. Because of the apparent influence of genetic background on gene expression patterns in both undifferentiated and differentiated cells, it will be critical to study a sufficient number of hiPSC lines to detect robust phenotypes; this is particularly relevant in complex diseases where the causal mutation(s) are not known. When studying monogenic diseases, it may be necessary to introduce mutations into wild- type hESCs or rescue mutations in patient-derived hiPSCs, as different backgrounds may mask subtle transcriptional differences[40].

METHODS

Cell culture

hESC lines and hiPSC lines were cultured with mouse embryonic fibroblasts (MEFs, Globalstem) pre-plated at 12-15,000 cells/cm^2. Medium containing DMEM/F12, 20% knockout serum replacement, 1mM L-glutamine, 100 uM MEM non-essential amino acids, and 0.1 mM beta-mercaptoethanol was used. 10 ng/ml of FGF-2 was added after sterile

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

filtration and cells were fed daily and passaged weekly using 6 U/mL dispase or mechanically.

hiPSC generation

hESC lines were cultured in fibroblast medium without FGF-2 containing DMEM, 10% FBS, 1 mM L-glutamine, 100 uM MEM non-essential amino acids, and 0.1 mM beta-mercaptoethanol, for a week. Cells were passaged three times using 0.25% trypsin and then sorted for $hThy1^{+}$/$hTRA\text{-}1\text{-}81^{-}$ populations. Sorted fibroblast-like cells were plated, passaged one more time, and then reprogrammed by using CytoTune®-iPS Sendai Reprogramming Kit (Invitrogen) following manufacturer's instructions.

RNA-sequencing

Undifferentiated hESC/hiPSC cells were sorted for $hTRA\text{-}1\text{-}81^{+}$ to control for the homogeneity of cells before RNA extraction. The quality and quantity of total input mRNA was determined on an Agilent BioAnalyzer 2100 using Agilent RNA 6000 Nano kit. One microgram of total RNA from each sample was then used as input for library preparation using Illumina TruSeq RNA Sample Prep Kit, following manufacturer's instructions. Each paired-end library was prepared with an adaptor with unique index sequence. The size profile and quantity of resulting libraries were than determined on the BioAnalyzer 2100 with Agilent High Sensitivity DNA kit. These libraries were then pooled together at equal molar concentration and sequenced on an Illumina HiSeq 2000. All hESC and hiPSC samples for RNA-Seq analysis were prepared on the same day by the same person, and then sequenced simultaneously on the same run (except for hiPSC lines 1, 2 and 3; this did not affect the clustering). All fibroblasts samples were prepared and sequenced in the same manner as the pluripotent samples but on different days.

RNA-seq reads were mapped using Bowtie 0.12.7 allowing up to 2 mismatches, to the library of human transcriptome sequences obtained from ENSEMBL (GRCh37.67) reference chromosomes, then entries with identical gene symbols were merged. The transcriptome includes both protein-coding genes and non-coding genes such as lincRNAs. EMSAR was used to quantify the expression levels in TPM (transcripts per million) and to infer read counts for individual genes. Differentially expressed genes were identified using edgeR 3.4.2 and confirmed using DESeq 1.8.3.

Methylation analysis

Methylation of individual CpGs was derived by observing bisulfite conversion of unmethylated cytosines in RRBS reads when compared to the reference genome. Methylation maps of individual CpGs show the average methylation value obtained by dividing the number of reads on which the CpG was methylated by the total times the CpG was covered by a read. Promoters were defined as 1 Kb up- and downstream of Refseq gene transcription start sites. Methylation values of individual CpGs in promoters were pooled in a weighted manner (i.e. proportional to the number of reads covering that CpG).

To count differentially-methylated promoters that supported variance due to cellular origin or genetic background, within-sample methylation difference was compared to the between-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

sample methylation difference for each promoter in sets based on cellular origin (hESC/hiPSC) and cell background (HUES2/HUES3). The promoter was assigned to the set with the lesser methylation difference, such that promoters in the hESC/hiPSC set showed greater methylation difference between hESCs and hiPSCs and lesser metylation difference between HUES2 and HUES3.

Global methylation clustering was performed by first pooling individual CpG methylation levels into 1 Kb non-overlapping tiles using weighted averages as with promoters, and then using Pearson's correlation to compute distance between samples. Ward's method was used for hierarchical clustering analysis. Analyses were performed using R and Perl.

Immunostaining

Immunostaining was performed using the following antibodies: α-hTRA-1-81 (330704, BioLegend), Streptavidin APC (17-4317-82, eBioscience) α-hCD90 (328118, BioLegend), α-Sendai viral protein (PD029, MBL International), and α- OCT4 (ASK-3006, Applied StemCell), α-PAX6 (Cat. no. PAX6, DSHB), and α- SOX1 (Cat. no. 4194, Cell Signaling).

Lactate production assay

Lactate production assay was done according to Zhong *et al*[41]. Lactate concentration was determined with the Lactate Assay Kit (BioVision). O.D. was measured at 570nm, 30 min. after addition of substrate.

Glucose uptake assay

The glucose uptake assay was done according to Sebastián *et al.*[42]. Cells were grown under normal conditions for 24 hr and 100 mM 2-NBDG (Invitrogen) was added to the media for 2 hr. Fluorescence was measured in a FACSCalibur Analyzer (BD).

Neural differentiation

Neural induction was performed as previously reported[33]. Briefly, cells were dissociated to single cells using Accutase and plated on gelatin for 10 minutes to remove MEFs. Non-adherent cells were collected and plated on Geltrex-treated dishes at a density of 150-200k cells per well of a 24-well plate in the presence of MEF-conditioned hESC media containing 10 ng/ml of FGF-2 (Life Tech) and 10 uM of Y-27632 (Tocris). Neural differentiation was initiated when cells were confluent using KSR media containing 820 ml of Knockout DMEM (Life Tech), 150 ml Knockout Serum Replacement (Life Tech), 1 mM L-glutamine (Life Tech), 100 uM MEM non-essential amino acids (Life Tech), and 0.1 mM beta-mercaptoethanol (Life Tech) to inhibit SMAD signaling, 100 nM of LDN-193189 (Cat. no. ab142186, Abcam) and 5 uM of SB431542 (Cat. No. 13031, Cayman Chemical) were added on Days 0 through 9. Cells were fed daily, and N2 media (Life Tech) was added in increasing 25% increments every other day starting on Day 4 (100% N2 on Day 10).

Western blot analysis

For Western blot analysis of PAX6, 10 ug of whole cell lysates was loaded to 4-20% gradient SDS-PAGE gels and then transferred to nitrocellulose membranes (BIO-RAD) by using Trans-Blot® Turbo™ Transfer System (BIO-RAD). Blocked membranes were

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

incubated with antibodies against PAX6 (Cat. no. 5790, Abcam) or GAPDH (Cat. no. 2118, Cell Signaling), respectively. For Western blot analysis of LDHA, undifferentiated hESC/hiPSC cells were sorted for hTRA- 1-81^{+} in order to control for the homogeneity of the cells, and then the rest of the procedure ensued as above. LDHA (Cat. no. 2012S, Cell Signaling), β-ACTIN (Cat. no. MA5-15739-HRP, Thermo Scientific).

RNA extraction and qPCR

Total RNA was extracted from differentiating hESC/hiPSC lines using the TRIzol Reagent (Life Tech), and 0.51 ug of RNA was reverse transcribed by High Capacity cDNA Reverse Transcription Kit RT2 first strand kit (ABIQiagen). Primer sequences are provided below. qRT-PCR mixtures were prepared with SYBR Green PCR Master Mix Universal (Applied BiosystemsKapabiosystem) and reactions were done with the Eppendorf Realplex2.

EB scorecard assay

EB differentiation was performed as described previously[6]. On day 7, EBs were lysed and total RNA was extracted before analyzing differentiation markers using qPCR.

Primer sequences

GAPDH	Forward	AGG TCG GAG TCA ACG
	Reverse	GTG ATG GCA TGG ACT
SOX1	Forward	GCG GAA AGC GTT TTC
	Reverse	TAA TCT GAC TTC TCC
NESTIN	Forward	GAA ACA GCC ATA GAG
	Reverse	TGG TTT TCC AGA GTC TTC
PAX6	Forward	CTT TGC TTG GGA AAT CCG
	Reverse	AGC CAG GTT GCG AAG
FOXG1	Forward	CCC TCC CAT TTC TGT
	Reverse	CTG GCG GCT CTT AGA
OTX2	Forward	AAG CAC TGT TTG CCA
	Reverse	CAG GAA GAG GAG GTG

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank members of the Hochedlinger and Park lab for productive discussions and a critical reading of the manuscript. We are grateful to K. Folze-Donahue, M. Weglarz and L. Prickett at the MGH/HSCI flow cytometry core for their constant assistance and support. We are also thankful to the members of the Tufts Genomics Core for

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

performing RNA-sequencing. Work in the Lee lab was supported by grants from the Robertson Investigator Award of the New York Stem Cell Foundation and from the Maryland Stem Cell Research Fund (TEDCO). Parts of this work were supported by HHMI, MGH startup funds, the Gerald and Darlene Jordan Endowed Chair for Regenerative Medicine (to K.H.) and a pilot grant from the NIH (P01GM099117 to K.H.). J.C. was supported by the Vranos Family Graduate Research Fellowship in Developmental & Regenerative Biology.

References

1. Takahashi KK, Yamanaka SS. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006; 126:663–676. [PubMed: 16904174]
2. Takahashi K, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2007; 131:861–872. [PubMed: 18035408]
3. Park I-H, et al. Disease-specific induced pluripotent stem cells. Cell. 2008; 134:877–886. [PubMed: 18691744]
4. Yu J, et al. Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. Science. 2007; 318:1917–1920. [PubMed: 18029452]
5. Chin MH, et al. Induced Pluripotent Stem Cells and Embryonic Stem Cells Are Distinguished by Gene Expression Signatures. Cell Stem Cell. 2009; 5:111–123. [PubMed: 19570518]
6. Bock C, et al. Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. Cell. 2011; 144:439–452. [PubMed: 21295703]
7. Chin MH, Pellegrini M, Plath K, Lowry WE. Molecular Analyses of Human Induced Pluripotent Stem Cells and Embryonic Stem Cells. Cell Stem Cell. 2010; 7:263–269. [PubMed: 20682452]
8. Ruiz S, et al. Identification of a specific reprogramming-associated epigenetic signature in human induced pluripotent stem cells. Proc Natl Acad Sci U.S.A. 2012; 109:16196–16201. [PubMed: 22991473]
9. Teichroeb JH, Betts DH, Vaziri H. Suppression of the Imprinted Gene NNAT and X-Chromosome Gene Activation in Isogenic Human iPS Cells. PLoS ONE. 2011; 6:e23436. [PubMed: 22022350]
10. Phanstiel DH, et al. Proteomic and phosphoproteomic comparison of human ES and iPS cells. Nat Methods. 2011; 8:821–827. [PubMed: 21983960]
11. Soldner F, et al. Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. Cell. 2009; 136:964–977. [PubMed: 19269371]
12. Stadtfeld M, et al. Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and facilitates generation of all-iPS cell mice from terminally differentiated B cells. Nat Genet. 2012:1–10.10.1038/ng.1110
13. Newman AM, Cooper JB. Lab-Specific Gene Expression Signatures in Pluripotent Stem Cells. Cell Stem Cell. 2010; 7:258–262. [PubMed: 20682451]
14. Humpherys D, et al. Epigenetic instability in ES cells and cloned mice. Science. 2001; 293:95–97. [PubMed: 11441181]
15. Rouhani F, et al. Genetic background drives transcriptional variation in human induced pluripotent stem cells. PLoS Genet. 2014; 10:e1004432. [PubMed: 24901476]
16. Stadtfeld M, et al. Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. Nature. 2010; 465:175–181. [PubMed: 20418860]
17. Tchieu J, et al. Female Human iPSCs Retain an Inactive X Chromosome. Cell Stem Cell. 2010; 7:329–342. [PubMed: 20727844]
18. Anguera MC, et al. Molecular signatures of human induced pluripotent stem cells highlight sex differences and cancer genes. Cell Stem Cell. 2012; 11:75–90. [PubMed: 22770242]
19. Fusaki N, Ban H, Nishiyama A, Saeki K, Hasegawa M. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. Proc Jpn Acad, Ser B, Phys Biol Sci. 2009; 85:348–362.
20. Cowan CA, et al. Derivation of embryonic stem-cell lines from human blastocysts. N Engl J Med. 2004; 350:1353–1356. [PubMed: 14999088]
21. Mallon BS, et al. Comparison of the molecular profiles of human embryonic and induced pluripotent stem cells of isogenic origin. Stem Cell Research. 2014; 12:376–386. [PubMed: 24374290]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

22. Guenther MG, et al. Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. Cell Stem Cell. 2010; 7:249–257. [PubMed: 20682450]
23. Maherali N, et al. A high-efficiency system for the generation and study of human induced pluripotent stem cells. Cell Stem Cell. 2008; 3:340–345. [PubMed: 18786420]
24. Everse J, Kaplan NO. Lactate dehydrogenases: structure and function. Adv Enzymol Relat Areas Mol Biol. 1973; 37:61–133. [PubMed: 4144036]
25. Fantin VR, St-Pierre J, Leder P. Attenuation of LDH-A expression uncovers a link between glycolysis, mitochondrial physiology, and tumor maintenance. Cancer Cell. 2006; 9:425–434. [PubMed: 16766262]
26. Mueckler M, et al. Sequence and structure of a human glucose transporter. Science. 1985; 229:941–945. [PubMed: 3839598]
27. Young CD, et al. Modulation of glucose transporter 1 (GLUT1) expression levels alters mouse mammary tumor cell growth in vitro and in vivo. PLoS ONE. 2011; 6:e23205. [PubMed: 21826239]
28. Folmes CDL, et al. Somatic oxidative bioenergetics transitions into pluripotency-dependent glycolysis to facilitate nuclear reprogramming. Cell Metab. 2011; 14:264–271. [PubMed: 21803296]
29. Cohen DR, Cheng CW, Cheng SH, Hui CC. Expression of two novel mouse Iroquois homeobox genes during neurogenesis. Mech Dev. 2000; 91:317–321. [PubMed: 10704856]
30. Matsumoto K, et al. The prepattern transcription factor Irx2, a target of the FGF8/MAP kinase cascade, is involved in cerebellum formation. Nat Neurosci. 2004; 7:605–612. [PubMed: 15133517]
31. Girirajan S, et al. Refinement and discovery of new hotspots of copy- number variation associated with autism spectrum disorder. Am J Hum Genet. 2013; 92:221–237. [PubMed: 23375656]
32. Marshall CR, et al. Structural Variation of Chromosomes in Autism Spectrum Disorder. The American Journal of Human Genetics. 2008; 82:477–488. [PubMed: 18252227]
33. Chambers SM, et al. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nature Biotechnology. 2009; 27:275–280.
34. Zhang X, et al. Pax6 Is a Human Neuroectoderm Cell Fate Determinant. Cell Stem Cell. 2010; 7:90–100. [PubMed: 20621053]
35. Wang C, et al. The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. Nature Biotechnology. 2014; 32:926–932.
36. Zhao S, Fung-Leung W-P, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS ONE. 2014; 9:e78644. [PubMed: 24454679]
37. Loewer S, et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat Genet. 2010; 42:1113–1117. [PubMed: 21057500]
38. Abyzov A, et al. Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. Nature. 2012; 492:438–442. [PubMed: 23160490]
39. Koyanagi-Aoi M, et al. Differentiation-defective phenotypes revealed by large-scale analyses of human pluripotent stem cells. Proc Natl Acad Sci U.S.A. 201310.1073/pnas.1319061110
40. Soldner F, et al. Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell. 2011; 146:318–331. [PubMed: 21757228]
41. Zhong L, et al. The Histone Deacetylase Sirt6 Regulates Glucose Homeostasis via Hif1a. Cell. 2010; 140:280–293. [PubMed: 20141841]
42. Sebastián C, et al. The Histone Deacetylase SIRT6 Is a Tumor Suppressor that Controls Cancer Metabolism. Cell. 2012; 151:1185–1199. [PubMed: 23217706]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

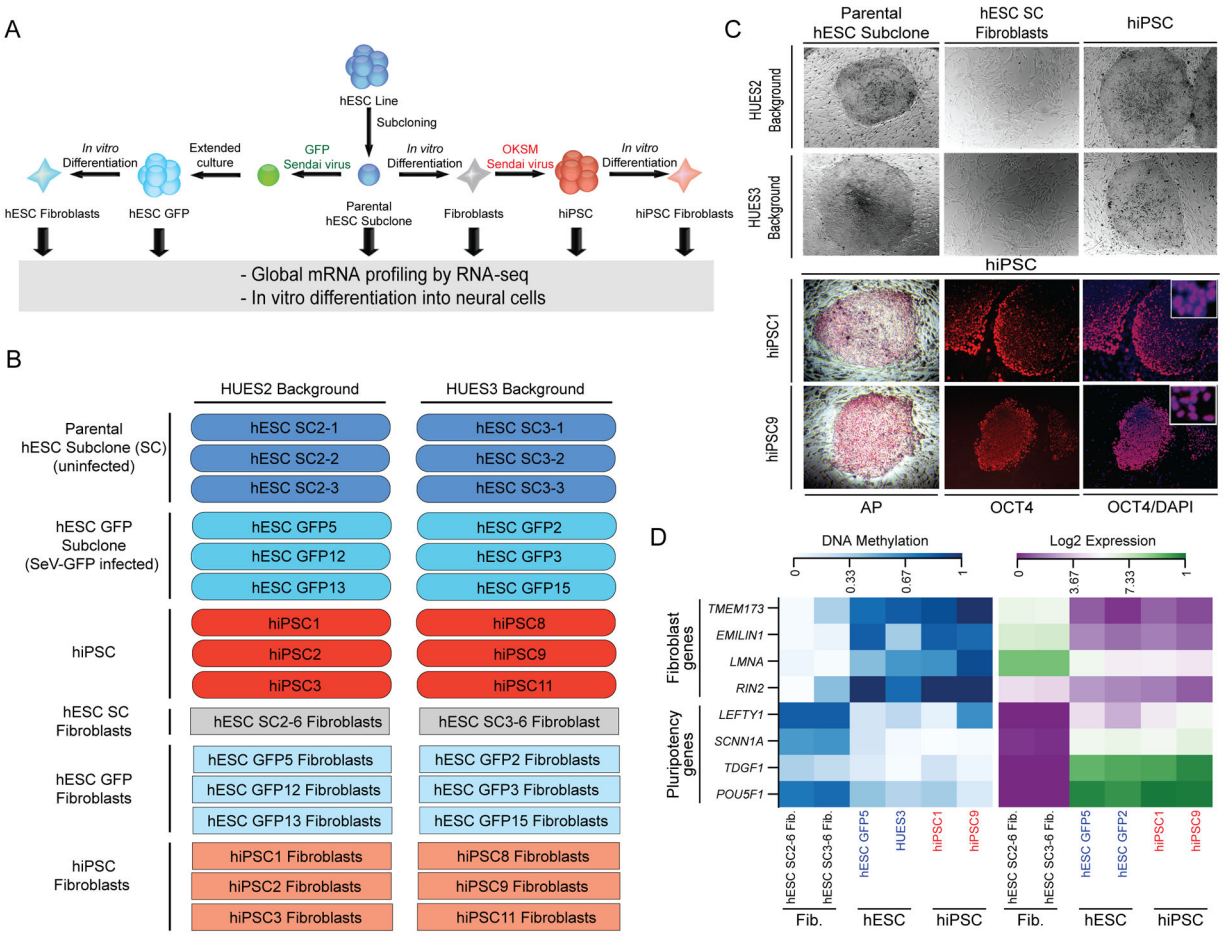

Figure 1. Generation of genetically matched hESCs and hiPSCs

(A) Schematic for the generation of genetically matched hESC and hiPSC lines. **(B)** Overview of HUES2 and HUES3 derivatives used for RNA-sequencing. **(C)** Top panel shows bright images of hESC subclones, *in vitro*-differentiated fibroblasts, whereas the bottom panel shows hiPSC lines stained for alkaline phosphatase activity or OCT4 expression. Co-staining with DAPI confirmed nuclear expression of OCT4 (inset). **(D)** Heatmaps depicting DNA methylation (left) and gene expression (right) levels of key fibroblast-associated and pluripotency-associated genes in isogenic hESCs, *in vitro*-differentiated fibroblasts and derivative hiPSCs.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

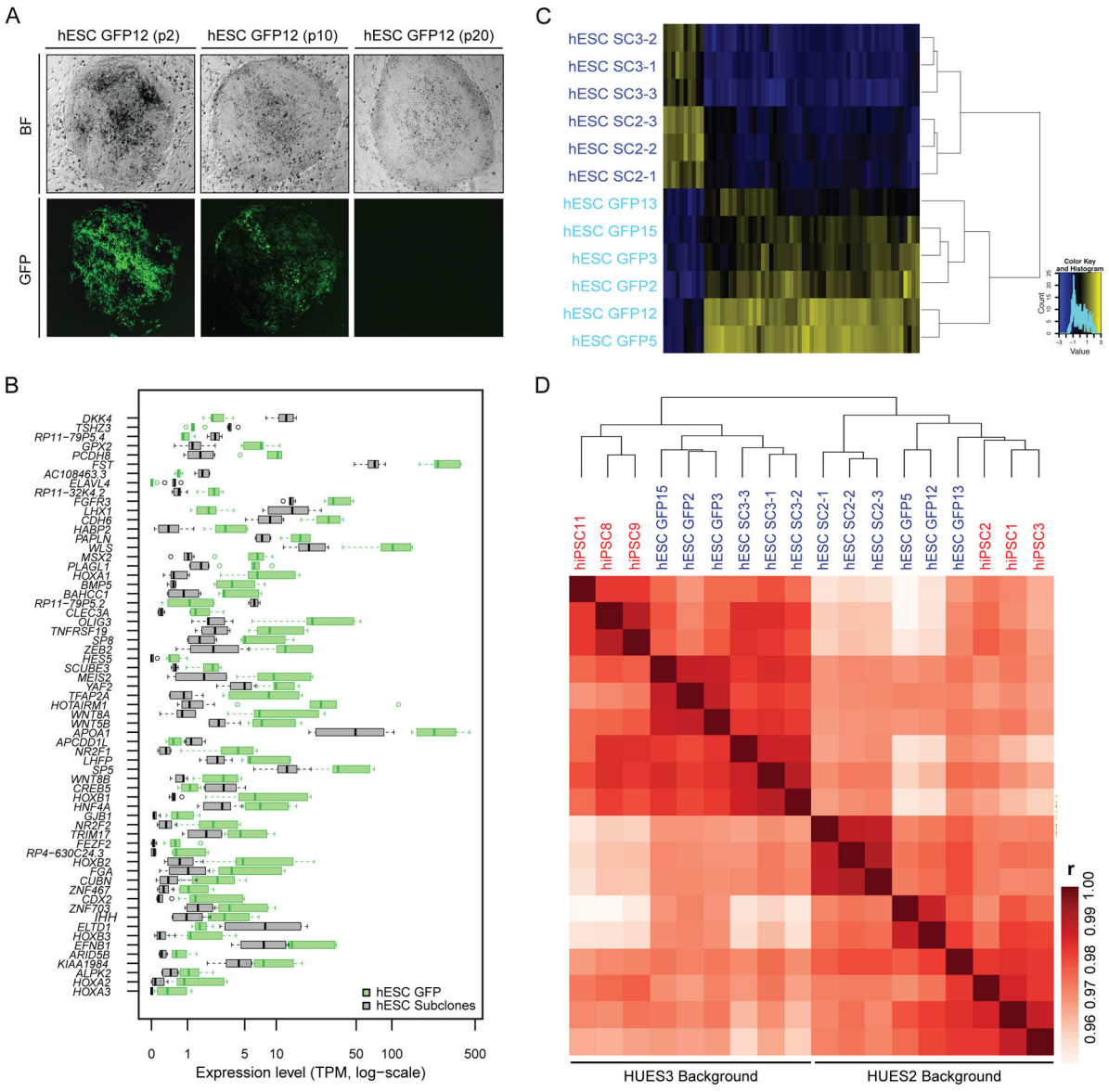

Figure 2. Influence of viral infection and genetic background on transcriptional patterns in isogenic hESCs and hiPSCs

(A) Representative bright field (top) and fluorescence (bottom) images of the hESC GFP12 line at passage 2, 10, and 20 after SeV-GFP infection. **(B)** Expression levels of 63 genes that were identified to be significantly different between 3 biological replicates of hESC GFP and 3 biological replicates of hESC SC lines within each of the two genetic backgrounds (FDR<0.01 and fold change >2 or <0.5; see details in the Methods section). Green and grey boxes indicate the expression ranges for each differentially regulated gene in 6 hESC GFP and 6 hESC SC lines, respectively. TPM; transcripts per million. **(C)** Heatmap and dendrogram separating all isogenic hESC lines based on the 63 DEGs shown from Fig. 2B. hESC SC lines, dark blue; hESC GFP lines, light blue. **(D)** Heatmap and dendrogram for all

isogenic hESC and hiPSC lines based on pairwise Pearson correlation (r) of global gene expression levels (log-scaled). hiPSC lines, red; hESC lines, blue.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

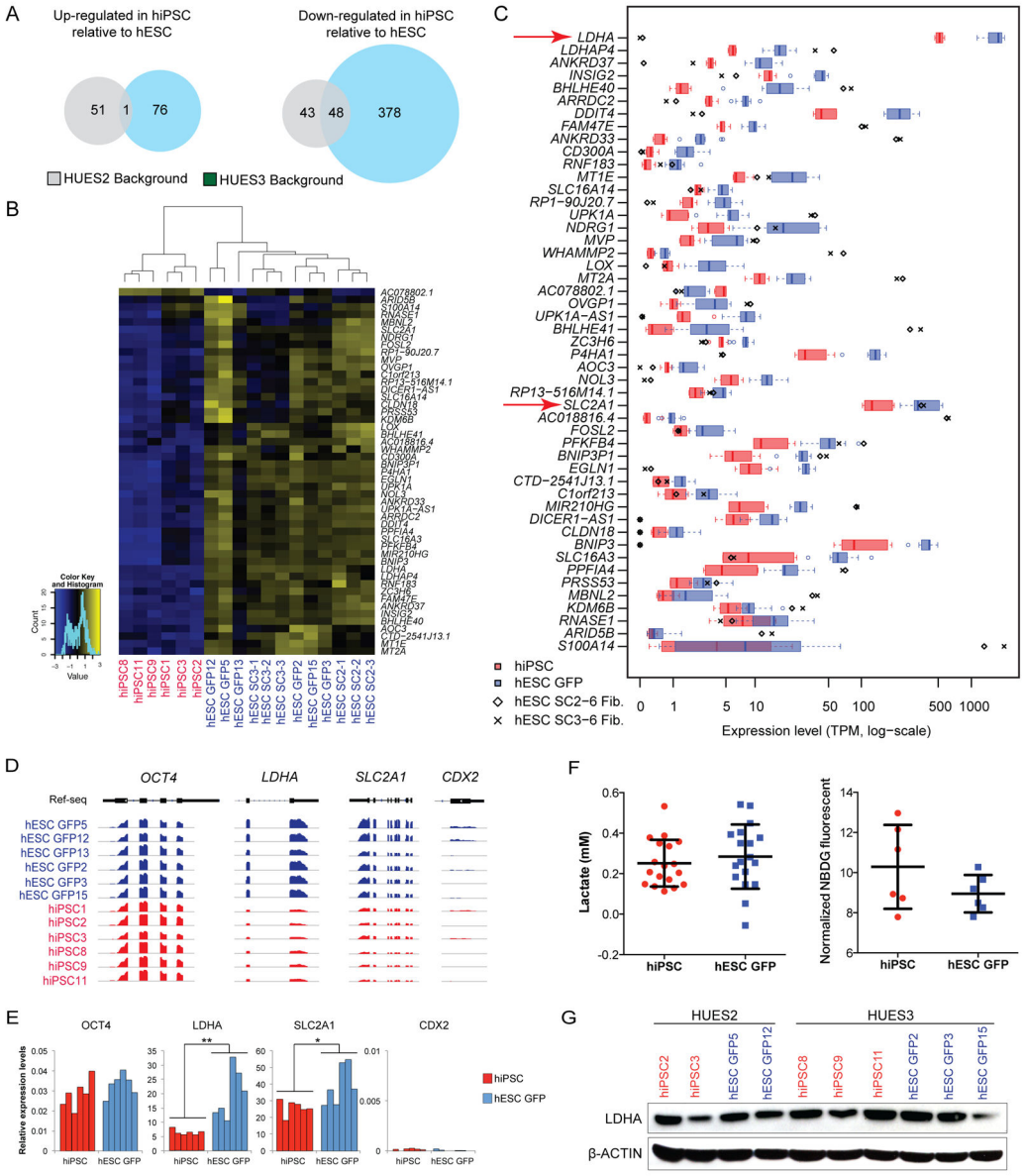

Figure 3. Differentially expressed genes (DEGs) between isogenic hESC and hiPSC lines

(a) Venn diagram showing the number of genes consistently up- or down- regulated in 3 biological replicate hiPSC lines when compared to 3 biological replicate hESC GFP lines from the same genetic background. (FDR<0.01 and fold change <2 or <1/2, see details in the Methods section). **(B)** Heatmap and dendrogram for all isogenic hESC and hiPSC lines based on the 49 differentially expressed genes (DEGs) that were common between the HUES2 and HUES3 backgrounds, using hierarchical clustering based on row-scaled expression levels. hiPSC lines, red; hESC lines, blue. **(C)** Box plot of 6 hESC GFP lines, 6 hiPSC lines, and parental fibroblasts for the 49 DEGs. Red and blue boxes indicate the expression ranges of each gene in hiPSC and hESC GFP lines, respectively. Diamonds and crosses indicate the expression levels of each gene in parental fibroblasts derived from HUES2 and HUES3 backgrounds, respectively. Genes are ordered by Student's t-test p-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

value between the 6 hiPSC and 6 hESC GFP lines. Red arrows depict genes discussed in main text. **(D)** RNA-seq read density of hESC GFP and hiPSC lines for *OCT4, LDHA, SLC2A1,* and *CDX2*. **(E)** Expression levels of *OCT4, LDHA, SLC2A1,* and *CDX2* by qPCR in hESC GFP and hiPSC lines, normalized to *ACTB* (n=6). Student's t-test *, $p<0.05$; **, $p<0.01$. Mean ± s.d. **(F)** Lactate production levels (left) and glucose uptake levels (right) of hESC GFP (blue) and hiPSC lines (red). Shown are data from three biological replicates for lactate assay (6 technical replicates) and from six biological replicates for glucose uptake assay ($p>0.05$ for both assays). Mean ± s.d. **(G)** Representative Western blot for LDHA levels in hESC GFP (blue) and hiPSC (red) lines.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

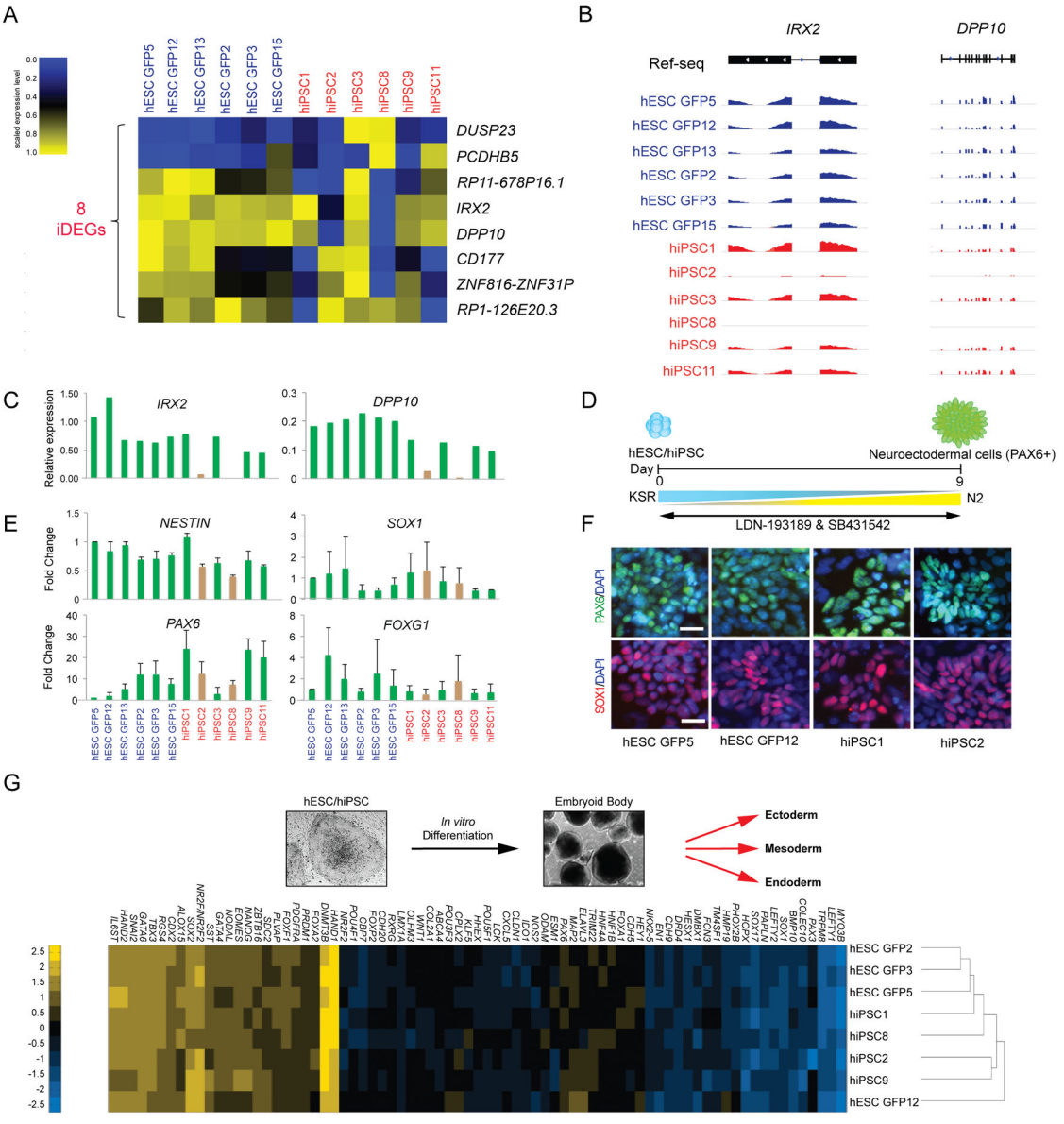

Figure 4. Dysregulation of genes in a subset of hiPSC lines

(A) Heatmap of the 8 inconsistently differentially expressed genes (iDEGs) for all isogenic hESCs and hiPSC lines (as defined in Supplementary Fig. 3C) within each of the two genetic backgrounds at FDR<0.01 and fold change <2 or <0.5. hiPSC lines, red; hESC lines, blue. **(B)** Genome browser images of *IRX2* and *DPP10* RNA-seq reads in hESC GFP and hiPSC lines. **(C)** Expression levels of *IRX2* and *DPP10* by qPCR in each hESC GFP and hiPSC line, normalized to *ACTB*. Brown bars indicate hiPSC lines that have undergone aberrant silencing of *IRX2* and *DPP10*. **(D)** Schematic for neural induction using the combination of SB431542, an ALK inhibitor, and LDN-193189, a BMP inhibitor. **(E)** Fold change of the neural markers *NESTIN, SOX1, PAX6*, and *FOXG1* by qPCR in hESC GFP and hiPSC lines relative to the hESC GFP5 line. Brown bars indicate the hiPSC lines that have undergone transcriptional silencing of *IRX2* and *DPP10*. Results are shown from three

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

independent experiments. Mean ± s.d. **(F)** Immunofluorescence staining of PAX6 (green) and SOX1 (red) indicates neural differentiation at day 6 in hESC GFP and hiPSC lines. DAPI (blue). (G) Scorecard assay of embryoid bodies (EBs) derived from isogenic hESC and hiPSC lines. Heatmap and dendrogram for these EBs based on the expression levels of the indicated developmental genes.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

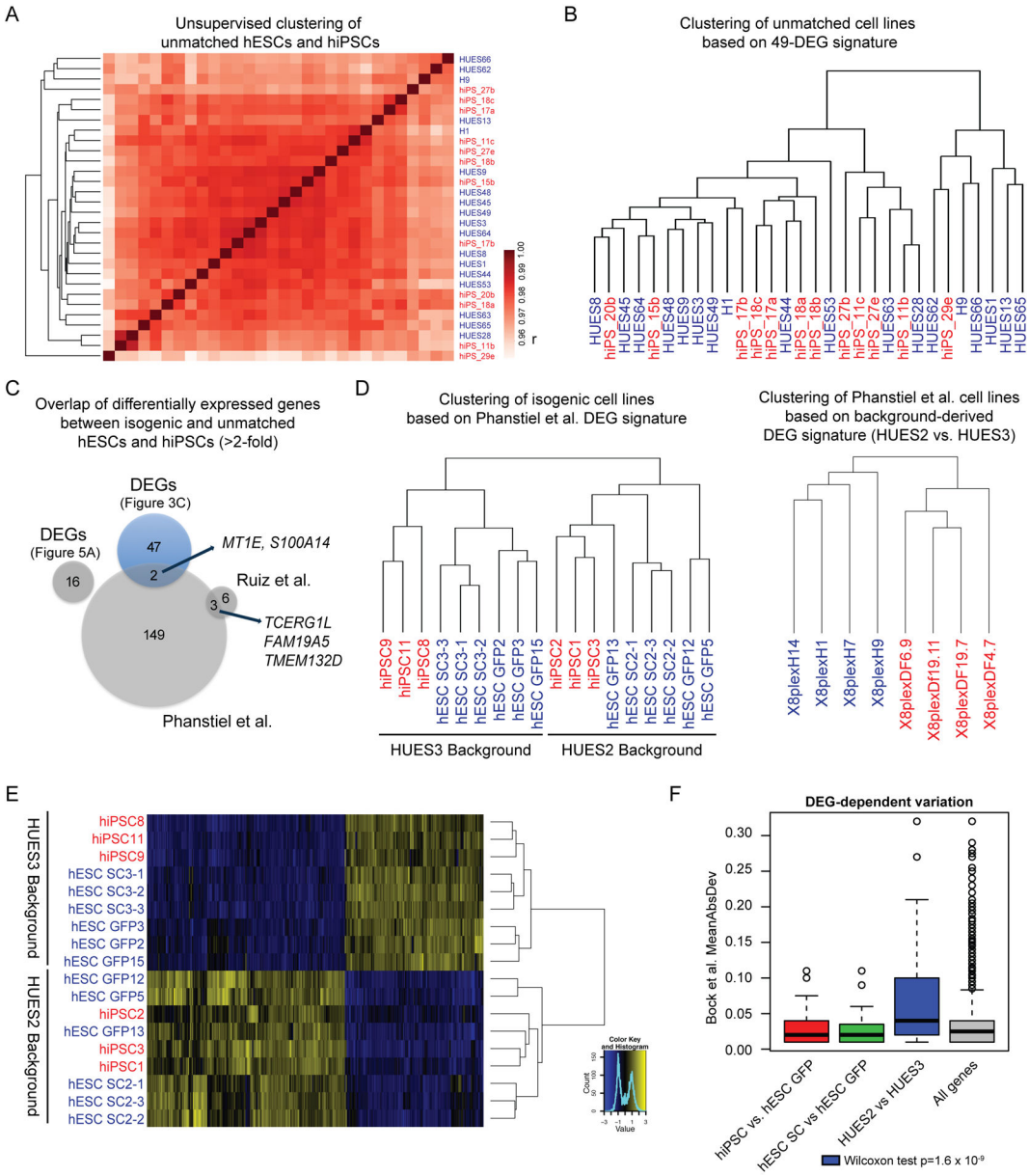

Figure 5. Analysis of previously reported gene expression differences and role of genetic background

(A) Dendrogram and heatmap for unmatched hESC (blue) and hiPSC (red) lines based on pairwise Pearson correlation (r) on global gene expression levels (log- scaled). **(B)** Dendrogram based on the 49 DEGs identified using isogenic lines in Fig. 3C for all unmatched hESC (blue) and hiPSC (red) lines. Note the lack of clustering. **(D)** Venn diagram of differentially expressed genes between hESCs and hiPSCs identified in this study and previous reports utilizing unmatched hESCs and hiPSCs. Overlapping genes between DEGs from independent reprogramming studies are indicated by arrows. **(D)** Dendrogram of isogenic hESC and hiPSC lines using the differentially expressed genes identified by Phanstiel *et al.*[10] (left panel) and dendrogram of hESC and hiPSC lines from Phanstiel *et al.* using HUES2 vs. HUES3-specific DEGs as discussed in the main text (right). hiPSC lines,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

red; hESC lines, blue. **(E)** Dendrogram and heatmap of isogenic hESC (blue) and hiPSC (red) lines based on HUES2 vs. HUES3-specific DEGs. **(F)** Transcriptional variation of different gene sets across unmatched hESC and hiPSC lines reported by Bock et al. Boxplots show mean absolute deviation (MAD) among hiPSCs and hESCs when considering indicated DEG sets. Note that HUES2 vs HUES3-specific DEGs show the greatest variation. A one-tailed Wilcoxon ranksum test was performed between each set of DEGs and all genes.