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

Jiho Choi1–3,10, Soohyun Lee4,10, William Mallard5,6, Kendell Clement2,5,6, Guidantonio Malagoli Tagliazucchi4,7, Hotae Lim8, In Young Choi8, Francesco Ferrari8, Alexander M Tsankov2,5,6, Ramona Pop2,5,6, Gabsang Lee8, John L Rinn5,6,9, Alexander Meissner2,5,6, Peter J Park4 & Konrad Hochedlinger1–3

The equivalence of human induced pluripotent stem cells (hiPSCs) and human embryonic stem cells (hESCs) remains controversial. Here we use genetically matched hESCs and hiPSCs to assess the contribution of cellular origin (hESC vs. hiPSC), the Sendai virus (SeV) reprogramming method and genetic background to transcriptional and DNA methylation patterns while controlling for cell line clonality and sex. We find that transcriptional and epigenetic variation originating from genetic background dominates over variation due to cellular origin or SeV infection. Moreover, the 49 differentially expressed genes we detect between genetically matched hESCs and hiPSCs neither predict functional outcome nor distinguish 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 support that genetic background variation is a major confounding factor for transcriptional and epigenetic comparisons of pluripotent cell lines, explaining some of the previously observed differences between genetically 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 advancing toward clinical application2–4. 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 sets5–7. 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 nonintegrating factor delivery systems and culture conditions, can alter gene expression in iPSCs11–13. Likewise, genetic background may influence transcriptional signatures in pluripotent cell lines as iPSCs derived from different individuals are reportedly more divergent than iPSCs derived from the same individual14. The difference between the clonal origin of hiPSC lines, derived from single somatic cells, and the polyclonal origin of hESC lines may also introduce transcriptional variation15. An additional consideration is the sex of cell lines and defects in X-chromosome reactivation in female hiPSCs16,17. Some of these variables have been addressed in previous reports11,12,14,18, 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 ESCs and integration-free iPSCs eliminates most of the transcriptional variation observed between unmatched cell lines19. 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 affect functional outcomes.

RESULTS

Approach to generate genetically matched hESCs and hiPSCs

To compare hESCs with genetically matched hiPSCs devoid of viral integrations, we generated hiPSCs from in vitro-differentiated hESCs using a nonintegrating SeV-based reprogramming system19. 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 (ref. 20), for these experiments. We selected male hESC lines because...
female iPSCs can exhibit defects in X-chromosome reactivation\textsuperscript{16,17}, which might confound subsequent interpretations\textsuperscript{9,21}.

First, we subcloned each HUES line to ensure genetic and epigenetic homogeneity of cells and to properly control for the clonal origin of hiPSCs (Fig. 1a). We will refer to these HUES derivatives as parental hESC clones. We differeniated one parental hESC clone 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\textsuperscript{+}/TRA-1-81\textsuperscript{−} (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 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 remethylated 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 infection with SeV vectors expressing OKSM, as previously reported\textsuperscript{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).

### Background drives transcriptional and epigenetic variation

First, we studied whether the SeV reprogramming method affects global transcription. The parental hESC lines were infected with GFP-expressing SeV (SeV-GFP), subcloned and passaged until GFP fluorescence was no longer detectable before analyzing cell lines by RNA-seq (Figs. 1a and 2a). We found a common set of 63 genes that was differentially expressed between three uninfected parental hESC and three SeV-GFP-infected hESC clones from each genetic background, which demonstrates

---

**Figure 1** Generation of genetically matched hESCs and hiPSCs. (a) Schematic for the generation of genetically matched hESC and hiPSC lines. RRBS, reduced representation bisulfite sequencing. (b) Overview of HUES2 and HUES3 derivatives used for RNA-seq. (c) Top panel shows bright images of parental hESC clones, in vitro-differentiated fibroblasts and derivative hiPSC lines, whereas the bottom panel shows hiPSC lines stained for alkaline phosphatase (AP) activity or OCT4 expression. Co-staining with DAPI confirmed nuclear expression of OCT4 (inset). Scale bars, 100 μm. (d) Heatmaps depicting DNA methylation (left) and gene expression (log\textsubscript{2} TPM; right) levels of key fibroblast-associated and pluripotency-associated genes in genetically matched hESCs, in vitro-differentiated fibroblasts and derivative hiPSCs. TPM, transcripts per million; Fib., fibroblasts.
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 parental hESC lines from hESC GFP lines (Supplementary Fig. 1e). Gene Ontology terms significantly enriched among these 63 DEGs are related to transcription, DNA binding and development (Supplementary Fig. 1f). 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.

A comparison of the transcriptional profiles of parental hESCs, hESC GFPs, in vitro-differentiated fibroblasts and derivative hiPSCs by unsupervised clustering showed the largest differences between pluripotent and differentiated cell types, consistent with previous observations5,14,22,23 (Supplementary Fig. 2a). Likewise, global

Figure 2 Influence of viral infection and genetic background on transcriptional and epigenetic patterns in hESCs and hiPSCs. (a) Representative brightfield (top) and fluorescence (bottom) images of the hESC2.6 G12 line at passage 2, 10 and 20 after SeV-GFP infection. Scale bars, 100 µm. (b) Expression levels of 63 genes that were identified to be significantly different between three biological replicates of hESC GFP and three biological replicates of parental hESC lines within each of the two genetic backgrounds (FDR < 0.01 and fold change >2 or <0.5). Green and gray boxes indicate the expression ranges for each differentially regulated gene in hESC GFP and parental hESC lines, respectively. TPM, transcripts per million. (c) Heatmap and dendrogram for all hESC and hiPSC lines based on pairwise Pearson correlation (r) of global gene expression levels (log scale). hiPSC lines, red; hESC lines, blue. (d) Dendrogram showing relation of indicated hESC and hiPSC lines based on global DNA methylation patterns. Note that cell lines separate by background rather than cell type. hiPSC lines, red; hESC lines, blue. (e) Number of differentially methylated promoters between genetically matched hESC GFP and hiPSC lines or between HUES2-derived and HUES3-derived pluripotent stem cell lines.
methylolation analysis of representative samples by reduced representation bisulfite sequencing separated pluripotent cells from in vitro-differentiated fibroblasts, indicating distinct epigenetic states (Supplementary Fig. 2b).

Unsupervised clustering indicated a clear segregation of all pluripotent samples into two transcriptionally related groups, irrespective of whether cell lines had been infected with SeV or not (Fig. 2c, 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 clones clustered with HUES2-derived hiPSC lines, whereas HUES3-derived hESC clones 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 hESC and hiPSC lines (Supplementary Fig. 2c). Moreover, transcriptional variation within matched hiPSC lines was similar to that within matched hESC lines, indicating that hiPSC lines are not more variable than hESC lines (Supplementary Fig. 2d). To determine whether genetic background also influences epigenetic patterns in pluripotent stem cells, we analyzed additional HUES2-derived cell lines (one hESC clone and one hiPSC clone) and HUES3-derived cell lines (two hESC clones and one hiPSC clone) by reduced representative bisulfite sequencing. Similar to gene expression analysis, unsupervised clustering of Cpg methylation values separated hESC and hiPSC lines by background rather than cell type (Fig. 2d). Moreover, the number of promoters differentially methylated between unmatched pluripotent cell lines was approximately six times higher than that between matched pluripotent cell lines, emphasizing the epigenetic diversity among cells of different genetic background (Fig. 2e).

Alltogether, these data show that genetic background is a major driver of both transcriptional and epigenetic differences between pluripotent cell lines, whereas the SeV reprogramming method introduces more subtle yet stable transcriptional changes in hESCs and 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. 2c). 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 downregulated, respectively, in hiPSC lines derived from the HUES2 background, whereas 77 and 426 genes were up- and downregulated 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 eight 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 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, which excludes 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. 3d). Both gene products are involved in energy metabolism; LDHA plays an important role in glycolysis by catalyzing the conversion of pyruvate to lactate24,25, whereas SLC2A1 facilitates glucose uptake in cells26,27. Of note, human pluripotent stem cells produce energy through glycolysis28. Based on the downregulation 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 genetically matched hiPSC and hESC GFP lines (Fig. 3f). Further, there was no difference in LDHA protein levels despite the observed transcriptional differences (Fig. 3g and Supplementary Fig. 3). Thus, at least 2 of the 49 DEGs seem not to translate into functional differences, possibly owing to post-transcriptional compensatory mechanisms.

The low number of differentially expressed genes between undifferentiated hESCs and hiPSCs does not exclude the existence of iPSC-specific aberrations that become detectable only after differentiation. We performed RNA-seq of fibroblast-like cells derived from two parental hESC, six hESC GFP and six hiPSC lines using the same in vitro differentiation protocol as described above (Fig. 1a). Only two genes were differentially expressed 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 identified between undifferentiated hESC and hiPSC lines (Supplementary Fig. 4a,b). However, HUES2-derived fibroblast-like cells tended to cluster together and apart from HUES3-derived fibroblast-like cells as shown in the principal component 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. 4c). We have previously shown that iDEGs between genetically matched mouse ESCs and iPSCs could predict the full developmental potential of subsets of iPSC lines18. 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 upregulated and 32 genes were downregulated in some of the HUES3-derived hiPSC lines relative to matched hESC GFP controls (Supplementary Fig. 4c). Only eight iDEGs were dysregulated in both genetic backgrounds, and these were thus selected for further analysis (Fig. 4a and Supplementary Fig. 4c).
Figure 3 Genes differentially expressed between genetically matched hESC and hiPSC lines. (a) Venn diagram showing the number of genes consistently up- or downregulated in three biological replicates of hiPSC lines when compared to three biological replicates of hESC lines. Genes are ordered by Student’s t-test P-value between the six hiPSC and six hESC 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. Error bars, mean ± s.d. (f) Lactate production (left) and glucose uptake (right) in hESC GFP (blue) and hiPSC lines (red). Shown are data from three biological replicates for lactate assay (six technical replicates) and from six biological replicates for glucose uptake assay (P > 0.05 for both assays). Error bars, mean ± s.d. (g) Representative western blot for LDHA levels in hESC GFP (blue) and hiPSC (red) lines. Full blot is available in Supplementary Figure 3.
The iDEGs IRX2 and DPP10 have been linked to neural development and psychiatric disease, 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 cell’s potential to differentiate into neuroectodermal cells using a published protocol, as determined by RNA expression analysis for NESTIN, SOX1, PAX6, and FOXG1, well-established markers of neuroectoderm differentiation from human pluripotent stem cells (Fig. 4d). 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. 4d). 

To determine whether hiPSCs exhibit biases in differentiation into other lineages, we evaluated their ability to generate ectodermal, endodermal and mesodermal derivatives by the ScoreCard assay. Briefly, hiPSC and hESC GFP lines from both genetic backgrounds were differentiated into embryoid bodies 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, genetically matched hESCs and hiPSCs cannot be discriminated on the basis of their ScoreCard 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 genetically matched hESCs and hiPSCs are also dysregulated in hiPSC lines derived from primary somatic cells as well as in other published data sets. 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. As many of the 49 DEGs were not covered in the available microarray data, we performed RNA-seq of these...
hESCs and hiPSCs, which offers increased sensitivity, especially for low-abundance transcripts. However, unsupervised clustering was unable to separate these hESCs from hiPSCs (Fig. 5a). Although three 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. 5a,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 matched and unmatched hESC/hiPSC lines, and two previously reported sets of DEGs. There was little to no overlap between DEGs discovered by independent laboratories (Fig. 5c and Supplementary Fig. 5b), and these DEGs could not distinguish hiPSC and hESC lines from the respective other data sets using supervised clustering (Supplementary Fig. 5c–i). Only 2 of our 49 DEGs (MT1E, S100A14) and 2 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 generated from one woman. Regardless of cell type were able to separate the hESCs and hiPSCs in ref. 10 (Fig. 5d, right panel; Fig. 5e and Supplementary Fig. 5f). 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 rank-sum test was performed between each set of DEGs and all genes.

**DISCUSSION**

Here we show that genetically matched, male hESC and hiPSC lines are transcriptionally and epigenetically 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. However, as in the current study, we found little to no overlap between DEGs discovered by independent laboratories (Fig. 5c and Supplementary Fig. 5b), and these DEGs could not distinguish hiPSC and hESC lines from the respective other data sets using supervised clustering (Supplementary Fig. 5c–i). Only 2 of our 49 DEGs (MT1E, S100A14) and 2 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.
methylations between hESCs and hiPSCs. This conclusion is particularly relevant to studies using (i) a limited number of hESC or hiPSC lines, (ii) hiPSC lines derived from a single donor individual and (iii) hESC and hiPSC lines of opposite sex⁴⁰, as these variables may further inflate transcriptional and epigenetic differences. Although the comparison of larger sets of hESC and hiPSC lines may control for these variables (Fig. 5a), it also increases transcriptional noise, which can obscure subtle, but meaningful molecular differences¹⁸.

We identified 49 DEGs that could distinguish genetically matched hESCs and hiPSCs and 8 iDEGs that were dysregulated in a subset of hiPSC lines across two different genetic backgrounds (Figs. 3c and 4a). This small number of genes contrasts with previous studies, which identified much larger sets of DEGs when comparing unmatched hESCs and hiPSCs using a similar cutoffs⁴⁵–⁷,10,13,22,23. 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, that is, energy production or differentiation potential into neural cells or embryoid bodies. These results therefore suggest that hESC and hiPSC lines are equivalent after accounting for genetic background differences. We surmise that the remaining DEGs we identified between genetically matched hESCs and hiPSCs might represent transcriptional noise.

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 the possibility that the lack of an obvious phenotype with the above-mentioned assays could be traced to insufficient expression of the analyzed genes in undifferentiated hiPSCs or compensation by post-transcriptional 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 are not manifested 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. 4a,b). The fact that genetically matched 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 aberrations that cannot be recapitulated with our in vitro differentiation approach.

Our study further reveals that a commonly used nonintegrating reprogramming method can subtly but stably alter transcriptional patterns in hESCs (Fig. 2b and Supplementary Fig. 1c). Notably, 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. 5f,i). Whereas the molecular mechanisms of this observation remain to be elucidated, our findings highlight the importance of controlling not only for genetic background but also for the method of reprogramming when comparing transcriptional patterns between hESCs and hiPSCs. Corroborating this notion, a recent study showed that hiPSC lines generated with integrating viruses exhibit dramatic differences in expression, methylation and differentiation potential compared to hiPSC lines generated with nonintegrating reprogramming systems⁴¹.

Our results have implications for the use of iPSC technology in disease modeling approaches, which often compare hiPSC lines derived from healthy and affected individuals. Because of the apparent influence of genetic background on gene expression patterns in both undifferentiated and differentiated cells, it may be helpful to study hiPSC lines from closely related individuals in order 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 advantageous to introduce mutations into wild-type hESCs or rescue mutations in patient-derived hiPSCs in order to minimize variation introduced by multiple genetic backgrounds⁴².

**METHODS**

Methods and any associated references are available in the online version of the paper.

**Accession codes.** GEO: GSE73211.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

**ACKNOWLEDGMENTS**

We thank members of the Hochedlinger and Park laboratories for productive discussions and a critical reading of the manuscript. We also thank M. Stadfeld for his helpful discussions and D. Melton for his generous donation of HUES2 and HUES3 lines. We are grateful to K. Fols-Donahue, M. Weglarz and L. Prickett at the Massachusetts General Hospital (MGH)/Harvard Stem Cell Institute (HSCI) flow cytometry core for their constant assistance and support. We are also thankful to the members of the Tufts Genomics Core for performing RNA-seq. Work in the Lee laboratory 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). A.M. and J.L.R. are supported by US National Institutes of Health (NIH) grant P01GM099117. A.M. is a New York Stem Cell Foundation Robertson Investigator. Parts of this work were supported by the Howard Hughes Medical Institution (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.

**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 hiPSCs. A.M. and J.L.R. provided RNA-seq data from hESCs and hiPSCs generated with retroviral vectors. J.C. performed AP staining, immunostaining, lactate production and glucose uptake assays, western blot analysis, RT-PCR and qPCR analyses. S.L., W.M., G.M.T., F.F. and P.J.P. performed immunostaining, lactate production and glucose uptake assays, western blot analysis, RT-PCR and qPCR analyses. S.L., W.M., G.M.T., F.F. and P.J.P. performed bioinformatics analysis of RNA-seq data. H.L., I.Y.C. and G.L. performed neural differentiation experiments and marker analyses of differentiated cells. R.P. conducted the ScoreCard assay, which was bioinformatically analyzed by A.M.T. and K.C. performed bioinformatics analysis of reduced representation bisulfite sequencing data.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676 (2006).
2. Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131, 861–872 (2007).
3. Park, I.-H. et al. Disease-specific induced pluripotent stem cells. Cell 134, 877–886 (2008).
4. Yu, J. et al. Induced pluripotent stem cell lines derived from human somatic cells. Science 318, 1917–1920 (2007).
5. Chiu, M.H. et al. Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. Cell Stem Cell 5, 111–123 (2009).
6. Bock, C. et al. Reference Maps of human ES and iPS cell variants enable high-throughput characterization of pluripotent cell lines. Cell 144, 439–452 (2011).
7. Chin, M.H., Pellegrini, M., Plath, K. & Lowry, W.E. Molecular analyses of human induced pluripotent stem cells and embryonic stem cells. Cell Stem Cell 7, 263–269 (2010).

8. Ruiz, S. et al. Identification of a specific reprogramming-associated epigenetic signature in human induced pluripotent stem cells. Proc. Natl. Acad. Sci. USA 109, 16196–16201 (2012).

9. Terchoeb, J.H., Betts, D.H. & Vaziri, H. Suppression of the imprinted gene NNAT and X-chromosome gene activation in isogenic human iPS cells. PLoS One 6, e23436 (2011).

10. Phansriel, D.H. et al. Proteomic and phosphoproteomic comparison of human ES and iPS cells. Nat. Methods 8, 821–827 (2011).

11. Soldner, F. et al. Parkinson’s disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. Cell 136, 964–977 (2009).

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. 44, 398–405, S1–S2 (2012).

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

14. Rouhani, F. et al. Genetic background drives transcriptional variation in human induced pluripotent stem cells. PLoS Genet. 10, e1004432 (2014).

15. Humphreys, D. et al. Epigenetic instability in ES cells and cloned mice. Science 293, 95–97 (2001).

16. Tchieu, J. et al. Female human iPScs retain an inactivate X chromosome. Cell Stem Cell 7, 329–342 (2010).

17. Anguera, M.C. et al. Molecular signatures of human induced pluripotent stem cells highlight sex differences and cancer genes. Cell Stem Cell 11, 75–90 (2012).

18. Stadtfeld, M. et al. Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. Nature 465, 175–181 (2010).

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. 85, 348–362 (2009).

20. Cowan, C.A. et al. Derivation of embryonic stem-cell lines from human blastocysts. N. Engl. J. Med. 350, 1353–1356 (2004).

21. Mallon, B.S. et al. Comparison of the molecular profiles of human embryonic and induced pluripotent stem cells of genetically matched origin. Stem Cell Res. (Amst.) 12, 376–386 (2014).

22. Guenther, M.G. et al. Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. Cell Stem Cell 7, 249–257 (2010).

23. Maherali, N. et al. A high-efficiency system for the generation and study of human induced pluripotent stem cells. Cell Stem Cell 3, 340–345 (2008).

24. Everse, J. & Kaplan, N.O. Lactate dehydrogenases: structure and function. Adv. Enzymol. 37, 61–133 (1973).

25. Fantin, V.R., St-Pierre, J. & Leder, P. Attenuation of LDH-A expression uncovers a link between glycolysis, mitochondrial physiology, and tumor maintenance. Cancer Cell 9, 425–434 (2006).

26. Mueckler, M. et al. Sequence and structure of a human glucose transporter. Science 229, 941–945 (1985).

27. Young, C.D. et al. Modulation of glucose transporter 1 expression levels alters mouse mammary tumor cell growth in vitro and in vivo. PLoS One 6, e23205 (2011).

28. Zhou, W. et al. HIF1α induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/ESt transition. EMBO J. 31, 2103–2113 (2012).

29. Cohen, D.R., Cheng, C.W., Cheng, S.H. & Hui, C.C. Expression of two novel mouse Iroquois homeobox genes during neurogenesis. Mech. Dev. 91, 317–321 (2000).

30. Matsumoto, K. et al. The prepattern transcription factor Inx2, a target of the FGF8/MAP kinase cascade, is involved in cerebellum formation. Nat. Neurosci. 7, 605–612 (2004).

31. Girirajan, S. et al. Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. Am. J. Hum. Genet. 92, 221–237 (2013).

32. Marshall, C.R. et al. Structural variation of chromosomes in autism spectrum disorder. Am. J. Hum. Genet. 82, 477–488 (2008).

33. Zhang, Y. et al. Functional genomic screen of human stem cell differentiation reveals pathways involved in neurodevelopment and neurodegeneration. Proc. Natl. Acad. Sci. USA 110, 12361–12366 (2013).

34. Chambers, S.M. et al. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat. Biotechnol. 27, 275–280 (2009).

35. Zhang, X. et al. Pax6 is a human neuroectoderm cell fate determinant. Cell Stem Cell 7, 90–100 (2010).

36. Tsankov, A.M. et al. A qPCR ScoreCard quantifies the differentiation potential of human pluripotent stem cells. Nat. Biotechnol. doi:10.1038/nbt.3387 (2015).

37. Wang, C. et al. The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. Nat. Biotechnol. 32, 926–932 (2014).

38. 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 9, e78644 (2014).

39. Louwer, S. et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat. Genet. 42, 1113–1117 (2010).

40. Abyzov, A. et al. Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. Nature 492, 438–442 (2012).

41. Koyanagi-Aoi, M. et al. Identification of a specific reprogramming-associated epigenetic signature in human induced pluripotent stem cells. Proc. Natl. Acad. Sci. USA 110, 20569–20574 (2013).

42. Soldner, F. et al. Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell 146, 318–331 (2011).
ONLINE METHODS

Cell culture. hESC and hiPSC cultures were maintained in mouse embryonic fibroblasts (MEFs, Globalstem) pre-plated at 12,000–15,000 cells/cm². Medium containing DMEM/F12, 20% knockout serum replacement, 1 mM L-glutamine, 100 µM MEM nonessential amino acids, and 0.1 mM beta-mercaptoethanol was used. 10 ng/ml of FGF-2 was added after sterile 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 µM MEM nonessential amino acids, and 0.1 mM beta-mercaptoethanol, for a week. Cells were passaged three times using 0.25% trypsin and then sorted for hTERT 1/HTRA1 -1 to 8. 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-seq. Undifferentiated hESC hiPSC cells were sorted for hTERT -1 to 8 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 adapter with unique index sequence.

The size profile and quantity of resulting libraries were 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 (ref. 43) allowing up to two mismatches, same manner as the pluripotent samples but on different days. RNA-seq reads were mapped using Bowtie 0.12.7 (ref. 43) allowing up to two mismatches.

The size profile and quantity of resulting libraries were determined on the BioAnalyzer 2100 using 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 (ref. 43) allowing up to two mismatches, same manner as the pluripotent samples but on different days. RNA-seq reads were mapped using Bowtie 0.12.7 (ref. 43) allowing up to two mismatches.

To measure MAD (mean absolute deviation) for each hESC GFP or hiPSC sample, first a reference set of samples was determined. Then, for each gene, the gene expression level (log(TPM+1)) was scaled by the mean across all hESC GFP and hiPSC samples and the absolute deviation was computed between a target sample and the average of the reference samples. For each target sample, the average absolute deviation over all genes with TPM > 1.5 was computed and this serves as the MAD of the target hESC GFP or hiPSC sample.

The mean and s.d. across target samples were obtained and plotted in Supplementary Figure 2b. In Supplementary Figure 2a, all hiPSCs are target samples and either the set of genetically matched hESC GFPs or the set of genetically unmatched hESC GFPs were used as reference. For Supplementary Figure 2c, either all hiPSCs or hESC GFPs were targets and either the set of all hESC GFPs or the set of all hiPSCs were used as reference. Changing technical details such as using median absolute deviation or mean squared deviation instead of mean absolute deviation, or scaling gene expression level versus not scaling, did not alter the conclusion.

Clustering of samples using various DEGs. Unsupervised hierarchical clustering was performed using the complete linkage method, after z-score transformation of log(TPM+1) across samples for each gene. Genes with missing or ambiguously matching IDs between published lists and our gene lists were excluded. The row scaling for Figure 3a was based on (x-min(x))/(max(x)-min(x)) normalization instead of z-score, for best visualization.

Functional enrichment analysis. Enrichment for Gene Ontology (GO) terms was tested using DAVID version 6.7 (refs. 49,50), with all the nonzero genes used as input for DEG analysis as the background set. We included all the default databases such as GO, pathway databases and protein domain databases. Additionally, we included eight protein-protein interaction databases to test for enrichment.

DEG ranking analysis. For each gene, a P-value was computed using the edger package (as described above) against the null hypothesis that there is no difference in expression of that gene between HUES2 and HUES3 hESC GFP lines. Likewise, another P-value was computed between HUES2 and HUES3 hiPSC lines. The two sets of P-values were each adjusted for multiple-testing using the Benjamini-Hochberg method (FDR). The ability to distinguish between HUES2 and HUES3 backgrounds is computed as the sum of the −log10 of the adjusted P-values. We sorted genes by this score and placed Phanstiel’s DEGs10 in that sorted list.

DEG permutation analysis. We observed that Phanstiel’s DEGs can separate our genetically matched samples by genetic background using unsupervised clustering. To test whether Phanstiel’s DEGs do this better than a random set of DEGs, we generated 10,000 size- and expression-level-matched random gene set and computed P-values as follows: (i) size- and expression-level-matched random gene set. The mean expression level across all 12 hESC and hiPSCs were computed for each gene. Then, genes with nonzero mean expression were divided into k quantile groups according to the mean expression level. Adding the zero-expressed gene group, we obtained k+1 groups of genes based on expression-level. From each group, we randomly chose the same number of genes as the number of Phanstiel DEGs in that group. We used k = 5 and k = 1 (ii) P-values for separation by genetic background. For a given gene set, we computed the score of separation as

\[ w = I_{\text{separated}} \cdot D(S_{\text{HUES2}}, S_{\text{HUES3}}) \]

where \( I_{\text{separated}} = 1 \) if the gene set completely separates HUES2 and HUES3 samples(\( S_{\text{HUES2}} \) and \( S_{\text{HUES3}} \), respectively). If not, \( D(X,Y) \) is the Dunn index53 between two groups X and Y. We also computed an alternative score based on Silhouette Width. Both Dunn index and Silhouette width22 represent how well two groups are separated and how compact each group is. To compute

Read coverage plots were generated using the Integrative Genome Browser48, after rescaling each track by a factor proportional to the total number of uniquely mapped reads, for best visualization.

Comparison of distance between cell lines using mean absolute deviation. To measure MAD (mean absolute deviation) for each hESC GFP or hiPSC sample, first a reference set of samples was determined. Then, for each gene, the gene expression level (log(TPM+1)) was scaled by the mean across all hESC GFP and hiPSC samples and the absolute deviation was computed between a target sample and the average of the reference samples. For each target sample, the average absolute deviation over all genes with TPM > 1.5 was computed and this serves as the MAD of the target hESC GFP or hiPSC sample.

The mean and s.d. across target samples were obtained and plotted in Supplementary Figure 2b. In Supplementary Figure 2a, all hiPSCs are target samples and either the set of genetically matched hESC GFPs or the set of genetically unmatched hESC GFPs were used as reference. For Supplementary Figure 2c, either all hiPSCs or hESC GFPs were targets and either the set of all hESC GFPs or the set of all hiPSCs were used as reference. Changing technical details such as using median absolute deviation or mean squared deviation instead of mean absolute deviation, or scaling gene expression level versus not scaling, did not alter the conclusion.

Clustering of samples using various DEGs. Unsupervised hierarchical clustering was performed using the complete linkage method, after z-score transformation of log(TPM+1) across samples for each gene. Genes with missing or ambiguously matching IDs between published lists and our gene lists were excluded. The row scaling for Figure 3a was based on (x-min(x))/(max(x)-min(x)) normalization instead of z-score, for best visualization.

Functional enrichment analysis. Enrichment for Gene Ontology (GO) terms was tested using DAVID version 6.7 (refs. 49,50), with all the nonzero genes used as input for DEG analysis as the background set. We included all the default databases such as GO, pathway databases and protein domain databases. Additionally, we included eight protein-protein interaction databases to test for enrichment.

DEG ranking analysis. For each gene, a P-value was computed using the edger package (as described above) against the null hypothesis that there is no difference in expression of that gene between HUES2 and HUES3 hESC GFP lines. Likewise, another P-value was computed between HUES2 and HUES3 hiPSC lines. The two sets of P-values were each adjusted for multiple-testing using the Benjamini-Hochberg method (FDR). The ability to distinguish between HUES2 and HUES3 backgrounds is computed as the sum of the −log10 of the adjusted P-values. We sorted genes by this score and placed Phanstiel’s DEGs in that sorted list.

DEG permutation analysis. We observed that Phanstiel’s DEGs can separate our genetically matched samples by genetic background using unsupervised clustering. To test whether Phanstiel’s DEGs do this better than a random set of DEGs, we generated 10,000 size- and expression-level-matched random gene set and computed P-values as follows: (i) size- and expression-level-matched random gene set. The mean expression level across all 12 hESC and hiPSCs were computed for each gene. Then, genes with nonzero mean expression were divided into k quantile groups according to the mean expression level. Adding the zero-expressed gene group, we obtained k+1 groups of genes based on expression-level. From each group, we randomly chose the same number of genes as the number of Phanstiel DEGs in that group. We used k = 5 and k = 1. (ii) P-values for separation by genetic background. For a given gene set, we computed the score of separation as

\[ w = I_{\text{separated}} \cdot D(S_{\text{HUES2}}, S_{\text{HUES3}}) \]

where \( I_{\text{separated}} = 1 \) if the gene set completely separates HUES2 and HUES3 samples(\( S_{\text{HUES2}} \) and \( S_{\text{HUES3}} \), respectively). If not, \( D(X,Y) \) is the Dunn index between two groups X and Y. We also computed an alternative score based on Silhouette Width. Both Dunn index and Silhouette width22 represent how well two groups are separated and how compact each group is. To compute
Methylation analysis. Methylation of individual CpGs was derived by observing bisulfite conversion of unmethylated cytosines in reduced representation bisulfite sequencing reads when compared to the reference genome. Methylation levels of individual CpGs were obtained by dividing the number of reads on which the CpG was methylated by the total number of reads covering that CpG.

Promoters were defined as 1 kb up- and downstream of RefSeq gene transcription start sites. Promoter methylation levels were determined by pooling methylation values of individual CpGs in a weighted manner (i.e., proportional to the number of reads covering that CpG). Differentially methylated promoters were defined using a two-sample t-test, with a cutoff of $P \leq 0.05$ and an absolute change in methylation of at least 10% between groups.

Global methylation clustering was performed using Pearson’s correlation distances between samples had been first obtained on z-transformed log(TPM+1) values.

RNA extraction and qPCR. Total RNA was extracted from differentiating hESC/iPSC lines using the TRIzol Reagent (Life Tech), and 0.51 μg 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 (Kapabiosystem) and reactions were done with the Eppendorf Realplex2.

Embryoid body ScoreCard assay. Embryoid body differentiation was performed, as described previously. On day 7, embryoid bodies were lysed and total RNA was extracted before analyzing differentiation markers using qPCR.

Primer sequences.

GAPDH: Forward AGG TCG GAG TCA ACG GAT TTG
Reverse GTG ATG GCA TGG ACT GTG GT
SOX1: Forward GCG GAA AGC GTT TTC TTG
Reverse TAA TCT GAC TTC TCC TCC C
NESTIN: Forward GAA ACA GCC ATA GAG GGC AAA
Reverse TGG TTT TTC AGA GTG TTC AGT GA
PA6X: Forward CTT TGC TTG GGA AAT CCG AG
Reverse AGC CAG GTT GCG AAG AAC TC
FOXG1: Forward CCC TAT TTC TGG AGC TCT
Reverse CTG CGG GCT ATT AGA GAT
OTX2: Forward AAG CAC TGT TTG CCA AGA CC
Reverse CAG GAA GAG GTG GAC AA

43. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25 (2009).
44. Flicek, P. et al. Ensembl 2014. Nucleic acids research. 42, D749–755 (2014).
45. Lee, S., Seo, C.H., Alver, B.H., Lee, S. & Park, P.J. EMSAR: estimation of transcript abundance from RNA-seq data by mappability-based segmentation and reclustering. BMC bioinformatics 16, 278 (2010).
46. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140 (2010).
47. Anders, S. & Huber, W. Differential expression analysis for sequence count data. Genome Biol. 11, R106 (2010).
48. Thovoldsdtrø, H., Robinson, J.T. & Mesrovi, J.P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. 14, 178–192 (2013).
49. Huang, W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1–13 (2009).
50. Dunn, J.C. Well-separated clusters and optimal fuzzy partitions. J. Cybern. 4, 95–104 (1974).
51. Rousseau, P.J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65 (1987).
52. Boyle, P. et al. Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. Genome Biol. 13, R92 (2012).
53. Zhong, L. et al. The histone deacetylase Sirt6 regulates glucose homeostasis via Hif1α. Cell 140, 280–293 (2010).
54. Sebastián, C. et al. The histone deacetylase SIRT6 is a tumor suppressor that controls cancer metabolism. Cell 151, 1185–1199 (2012).