Supplementary Information

Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines

Christoph Bock*, Evangelos Kiskinis*, Griet Verstappen*, Hongcang Gu, Gabriella Boulting, Zachary D Smith, Michael Ziller, Andreas Gnirke, Kevin C Eggan, Alexander Meissner

Supplementary Figure Legends

Supplementary Figure 1 (related to Figure 1)

(A) Representative bright field images and immunostainings of ES cell lines included in the current study. The scale bars correspond to 125μm.

(B) Pie charts illustrating the RRBS coverage at gene promoters, CpG islands and putative enhancers. Coverage is measured as the number of individual observations (i.e. aligned and quality-controlled sequencing reads) at CpGs within each region of a given type. Data are shown for one representative human ES cell line (H1).

(C) UCSC Genome Browser screenshot illustrating RRBS coverage at the SNAI1 gene locus. The promoter region of SNAI1 (violet) exhibits the highest density of CpGs (black) and also the highest RRBS coverage (blue). Additional RRBS coverage is visible at a downstream CpG island (green) and at an upstream regulatory element (orange). Most CpG-rich regions are unmethylated (light blue), while CpG-poor regions tend to be methylated (dark blue). Each blue dot corresponds to a single CpG that is covered by RRBS. Some epigenetic variation can be seen between H1 and H7, but overall the promoter region is unmethylated in all four ES cell lines.

(D) Pairwise correlation coefficients and scatterplots comparing DNA methylation (left) and gene expression (right) between biological replicates of three ES cell lines (HUES1, passage 28 and 29; HUES8, passage 29 and 30; H1, passage 37 and 38). In addition, the DNA methylation comparison includes two biological replicates of H1 that were grown at the University of Wisconsin (passage 25) and at Cellular Dynamics (passage 32), respectively. High similarity was observed for all pairwise comparisons. However, two types of differences between pairs of ES cell lines are visible from these diagrams: (i) Small but dense point clouds located in the bottom left close to the x-axis or y-axis (DNA methylation only). These points correspond to X-chromosome associated differences which distinguish female ES cell lines with widespread X-inactivation from male ES cell lines. (ii) Off-diagonal points scattered throughout the diagram. Most of these differences are located on the autosomes and constitute epigenetic or transcriptional differences between the ES cell lines.

Supplementary Figure 2 (related to Figure 3)

(A) Joint hierarchical clustering and heatmaps of human ES cell lines, iPS cell lines and fibroblasts. The clustering was performed as in Figure 1A. In addition, the "MEG3" column lists the expression status of the *MEG3* non-coding RNA: "+" stands for *MEG3* being expressed in the respective cell line (*MEG3* expression level $\geq$ 0.1) and "-" indicates that *MEG3* is not expressed (*MEG3* expression level < 0.1).

(B) Boxplots of the cell-line specific deviation from the ES-cell reference averaged across all genes, for the following cell lines: (i) those ES cell lines in which the *MEG3* non-coding RNA was expressed, (ii) those cell lines in which *MEG3* was not expressed and (iii) six primary fibroblast cell lines.

(C) UCSC Genome Browser screenshot illustrating how transgene silencing gives rise to spurious hypermethylation at the endogenous loci of the reprogramming factors. Due to the way in which RRBS reads are aligned to the genome, most viral transgene reads are placed at the endogenous loci of *OCT4*, *SOX2* and *KLF4*. This phe-

nomenon is illustrated for *KLF4*: In ES cells the *KLF4* gene is largely unmethylated (green box), while it appears partially methylated in iPS cells, but only at those exons that are part of the transgene (red box), never at introns that are not part of the transgene (blue box). Furthermore, incomplete transgene silencing in hiPS 27e (yellow box) is correlated with substantially lower DNA methylation levels in transgenic *KLF4*.

Supplementary Figure 3 (related to Figure 5)

(A) Immunostainings and gene expression data for five lineage marker genes. Four representative ES cell lines were subjected to non-directed EB differentiation. After two days, the EBs were plated onto matrigel and allowed to differentiate for another five days. After a total of seven days of EB differentiation, immunostainings were performed for marker genes of the three germ layers. The figure shows representative pictures of the undifferentiated ES cells, the EBs at day 7 and the immunostainings. The gene expression levels were obtained for 16-day EBs using the nCounter system (Supplementary Table 6). Scale bars correspond to 500μm (4e), 125μm (1a-g, 2a-b, 2d-f, 3a-b, 3e-g, 4a-b, 4f-g), 62.5μm (2g, 3d, 4c) and 31.25μm (2c, 3c, 4d).

(B) Mean lineage scorecard values for four ES cell lines (HUES1, HUES8, H1, H9) that were differentiated under conditions that favored ectoderm differentiation (blue) and mesoderm differentiation (red).

(C) Number of AFP-positive cells (determined by FACS) and mRNA expression levels (determined using the nCounter system) in day 16 EBs for hiPS 17a and hiPS 27e.

Supplementary Figure 4 (related to Experimental Procedures)

(A) Outline of the algorithm for calculating the deviation scorecard based on genome-wide DNA methylation and/or gene expression data.

(B) Illustration of the minimum threshold for DNA methylation differences in heterogeneous cell populations. Several ways for meeting the deviation threshold of 20 percentage points are shown: (i) all cells exhibit DNA methylation levels that are increased (decreased) by 20 percentage points; (ii) a subset of 20% of all cells exhibit DNA methylation levels that are increased (decreased) by 100 percentage points, while the remaining 80% do not show any difference; (iii) any combination as illustrated in the diagram.

(C) Outline of the algorithm for calculating the lineage scorecard based on marker gene expression in differentiating EBs.

(D) Saturation plot estimating the scorecard performance for DNA methylation (left) or gene expression (right) assays with reduced genomic coverage. Based on the data of all 20 ES cell lines (or random subsets of size 10, 5 and 1), all genes were ranked according to the average deviation from the ES-cell reference. Next, the top 1%, 5%, 10%, up to 90% most ES-cell variable genes were selected and the percentage of iPS cell-line specific deviations was calculated that would have been detected if only these genes were monitored for deviations.

(E) Correlation plot estimating the lineage scorecard performance without EB differentiation. Gene expression profiles were obtained for ES and iPS cell lines using the nCounter system and processed in the same way as the gene expression profiles from the day 16 EBs, giving rise to a lineage scorecard that is exclusively based on gene expression profiles of ES/iPS cell lines maintained under normal growth conditions. The scatterplots visualize the correlation between lineage scorecard estimates calculated from day 16 EBs (x-axis) and lineage scorecard estimates calculated from the pluripotent state (y-axis).

Supplementary Tables

All Supplementary Tables are provided as Excel files.

Supplementary Table 1 (related to Figure 1). Summary of cell lines and genomic experiments

Supplementary Table 2 (related to Figure 1). Combined dataset of gene-based DNA methylation and gene expression levels

Supplementary Table 3 (related to Figure 2). DNA methylation and gene expression levels, thresholds of the reference corridor and measures of gene-specific variation

Supplementary Table 4 (related to Figure 3). Validation of previously reported iPS gene signatures

Supplementary Table 5 (related to Figure 4). Deviation scorecard summarizing cell-line specific deviations from the ES-cell reference

Supplementary Table 6 (related to Figure 5). Configuration, raw data, results and reproducibility of the lineage scorecard

Supplementary Table 7 (related to Figure 6). Lineage scorecard predictions and differentiation efficiency into motor neurons

Supplementary Experimental Procedures

ES/iPS culture conditions:
The pluripotent cell lines were grown in human ES media consisting of KO-DMEM (Invitrogen), 10% KOSR (Invitrogen), 10% plasmanate (Talecris), 1% glutamax or L-glutamin, non-essential amino acids, penicillin/streptomycin, 0.1% 2-mercaptoethanol and 10-20ng/ml bFGF. All pluripotent cells were grown on a monolayer of irradiated CF1-MEFs (GlobalStem) and passaged using trypsin (0.05%) or dispase (Invitrogen). Before collection of DNA and RNA for analysis, ES and iPS cells were either isolated using trypsin (0.05%) or dispase, or plated on matrigel (BD Biosciences) for one passage and fed with human ES media conditioned in CF1-MEFs for 24h.

Differentiation protocols
A total of five ES/iPS cell differentiation protocols were used in the current study:

(i) *Non-directed EB differentiation*. Undifferentiated cells were harvested using dispase or trypsin and plated in suspension in low-adherence plates in the presence of human ES cell culture media without bFGF and plasmanate. Cell aggregates (EBs) were allowed to grow for a total of 16 days, refreshing media every 48h.

(ii) *Monocyte/macrophage differentiation*. Undifferentiated cells were treated with multiple recombinant proteins following a published protocol for hematopoietic differentiation (Grigoriadis et al., 2010). Briefly, feeder-depleted pluripotent cells were grown as small aggregates in suspension in 6-well low attachment plates (Corning) in StemPro-34 medium (Invitrogen) containing penicillin/streptomycin, glutamine (2mM), monothioglycerol (0.0004M), ascorbic acid (50µg/ml) (Sigma-Aldrich) and BMP4 (10ng/ml) (R&D Systems) for 24h. To induce primitive steak/mesoderm formation, EBs were washed and cultured further in the StemPro-34 differentiation medium, supplemented with human recombinant bFGF (5ng/ml) (Millipore) for another 3 days. At day 4, EBs were harvested again and cultured in the differentiation medium described above, additionally containing hVEGF (10ng/ml) (PeproTech), hbFGF (1ng/ml), hIL-6 (10ng/ml) (PeproTech), hIL-3 (40 ng/mL) (PeproTech), hIL-11 (5ng/mL) (PeproTech), and human recombinant SCF (100ng/mL) (PeproTech) for another 4 days to induce hematopoietic specification. From day 8 onwards, cells were further cultured in StemPro-34 medium, containing hVEGF (10ng/ml), human erythropoietin (4U/ml) (Cell Sciences), human thrombopoietin (50 ng/ml) (Cell Sciences), and human stem cell factor, hIL-6, hIL-11, and hIL-3 to promote hematopoietic cell maturation and expansion.

(iii) *Mesoderm differentiation.* Undifferentiated cells were treated with Activin A and BMP4 according to a published protocol that fosters mesoderm differentiation (Laflamme et al., 2007). Briefly, cells were harvested by incubation with collagenase IV (Invitrogen) and plated onto a Matrigel-coated cell culture dish. To induce mesoderm differentiation, cells were cultured in RPMI-B27 medium (Invitrogen) supplemented with human recombinant Activin A (100ng/ml) (R&D Systems) for 24h. Human recombinant BMP4 (10ng/ml) was added to the medium for four days, after which cells were fed further with supplement-free RBMI-B27 medium.

(iv) *Ectoderm differentiation*. Undifferentiated cells were harvested by incubation with collagenase IV (Invitrogen) and plated onto a Matrigel-coated cell culture dish. Cells were grown in KO-DMEM (Invitrogen) medium, containing knockout serum replacement (Invitrogen), supplemented with Noggin (500ng/ml) (R&D Systems) and SB431542 (10µM) (Tocris).

(v) *Motor neuron differentiation*. Undifferentiated cells were differentiated following a published protocol (Di Giorgio et al., 2008), as described in more detail by Boulting *et al.* (submitted).

Clonal bisulfite sequencing
Genomic DNA was isolated using PureLink genomic DNA mini kit (Invitrogen), DNA was bisulfite-converted using the EpiTect kit (Qiagen), and 50 ng of bisulfite converted DNA was PCR-amplified. Primer sequences were *CD14* forward 5'-AGTTGTGGTTGAGGTTTAGGTT-3' and reverse 5'-ACCACAAAACTTACACTTTCCA-3'. Amplicons were gel-purified and subcloned using TOPO TA cloning

kit (Invitrogen). Clones were randomly selected for sequencing, and the sequencing data were processed using the BiQ Analyzer software (Bock et al., 2005).

Quantitative RT-PCR analysis

Total RNA was isolated using RNeasy kit (Qiagen) according to manufacturer's recommendation followed by cDNA synthesis using standard protocols. Briefly, cDNA was synthesized using Superscript II Reverse Transcriptase (Invitrogen) and Random Hexamers (Invitrogen) with 500 ng of total RNA input. SYBR Green PCR master mix (Applied Biosystems) was used for qPCR analysis, which was done on a StepOnePlus real time PCR system (Applied Biosystems). PCR conditions were as follow: 94°C initial denaturation for 5min, 94°C 15s, 60°C 15s, 72°C 30s for 40 cycles, and 72°C for 10min. Primer sequences were: *CD14* forward 5'-ACGCCAGAACCTTGTGAGC-3' and reverse 5'-GCATGGATCTCCACCTCTACTG-3'; *CD33* forward 5'-TCTTCTCCTGGTTGTCAGCT-3' and reverse 5'-GAGGCAGAGACAAAGAGCG-3' (Garnache-Ottou et al., 2005); *CD64* forward 5'-GTGTCATGCGTGGAAGGATA-3' and reverse 5'-GCACTGGAGCTGGAAATAGC-3' (Li et al., 2010); and GAPDH forward 5'-ACCCACTCCTCCACCTTTGAC-3' and reverse 5'-ACCCTGTTGCTGTAGCCAAATT-3'. Relative quantification was calculated using the comparative threshold cycle (delta delta Ct) method.

Immunocytochemistry and FACS analysis

Immunostainings were performed using the following primary antibodies: AFP (Dako), NESTIN (Chemicon), OCT4 (Santa Cruz Biotechnology), alpha-SMA (Sigma), SSEA3 (Biolegend), SSEA4 (Chemicon), TRA-1-60 (Chemicon), TRA-1-81 (Chemicon), beta III Tubulin (Abcam), VEGFRII (Abcam). For FACS analysis, EBs were trypsin-dissociated to single cells, washed with PBS, fixed overnight with 4% paraformaldehyde and permeabilized with 0.5% PBS-Tween for 20min to 1h. Cells (~500k) were then blocked in 0.1% PBS-Tween supplemented with 10% donkey serum for 1h, and incubated with primary antibody (AFP:1:300) (DakoCytomation) overnight and secondary for 1h, washed and re-suspended in 1ml PBS with 0.1% donkey serum. Samples were analyzed using BD Biosystems LSRII analyzer. For FACS analysis, EBs were trypsin-dissociated to single cells, washed with PBS, fixed overnight with 4% paraformaldehyde and permeabilized with 0.5% PBS-Tween for 20min to 1h. Cells (~500k) were then blocked in 0.1% PBS-Tween supplemented with 10% donkey serum for 1h, and incubated with primary antibody (AFP:1:300) (DakoCtomation) overnight and secondary for 1h, washed and re-suspended in 1ml PBS with 0.1% donkey serum. Samples were analyzed using BD Biosystems LSRII analyzer.

Deviation scorecard calculation

The deviation scorecard summarizes which and how many genes in a cell line of interest deviate from the ES-cell reference. The reference is being constituted by the 20 low-passage ES cell lines – or by the 19 remaining ES cell lines when calculating the deviation scorecard for a cell line that is normally part of the reference. The algorithm for calculating the deviation scorecard (outlined in Supplementary Figure 4A) is the same for DNA methylation and gene expression data, with the only exception that the microarray data require an additional normalization step. From a statistical point of view, the deviation scorecard is based on non-parametric outlier detection using Tukey's outlier filter (Tukey, 1977). All genes for which the DNA methylation or gene expression value of the cell line of interest fall outside of the center quartiles by more than 1.5 times the interquartile range are considered suspected outliers and flagged as such. Next, the magnitude of the change is considered and only genes for which the deviation from the ES cell reference is sufficiently large to be considered biologically meaningful are ultimately reported as outliers. For the current study we used thresholds of at least 20 percentage points for DNA methylation and at least twofold for gene expression, consistent with prior work (Bock et al., 2010) and further justified in Supplementary Figure 4B. To account for the fact that deviations may be more or less concerning depending on which genes are affected, we assembled two lists of genes that need to be monitored particularly closely for DNA methylation defects, namely lineage marker genes and cancer genes. Devia-

tions at these genes are specifically highlighted in the extended version of the deviation scorecard (Supplementary Table 5). Finally, we note that we have also evaluated alternative strategies for flagging outliers, including a parametric approach that was based on moderated *t*-tests. Overall we found that the Tukey's outlier filter gave the most relevant results, and it has the additional advantage that it can be intuitively visualized by "reference corridor" boxplots (Figures 1C and 4A).

Lineage scorecard calculation

The lineage scorecard quantifies the differentiation propensity of a cell line of interest relative to a reference constituted by 19 low-passage ES cell lines. The algorithm for calculating the lineage scorecard (outlined in Supplementary Figure 5C) uses a combination of moderated *t*-tests (Smyth, 2004) and gene set enrichment analysis performed on *t*-scores (Nam and Kim, 2008; Subramanian et al., 2005). To provide a biological basis for quantifying lineage-specific differentiation propensities, we curated several sets of marker genes for each of the three germ layers (ectoderm, mesoderm, endoderm) as well as for the neural and hematopoietic lineages (Supplementary Table 6A). Next, Bioconductor's limma package was used to perform moderated *t*-tests comparing the gene expression in the EBs obtained for the cell line of interest to the EBs obtained for the ES cell reference, and the mean *t*-scores were calculated across all genes that contribute to a relevant gene set. High mean *t*-scores indicate increased expression of the gene set's genes in the tested EBs and are considered indicative of a high differentiation propensity for the corresponding lineage. In contrast, low mean *t*-scores indicate decreased expression of relevant genes and are considered indicative of a low differentiation propensity for the corresponding lineage. To increase the robustness of the analysis, the mean *t*-scores were averaged over all gene sets assigned to a given lineage. The lineage scorecard diagrams (Figure 5B and D) list these "means of gene-set mean *t*-scores" as quantitative indicators of cell-line specific differentiation propensities. The lineage scorecard analyses and validations were performed using custom R scripts (http://www.r-project.org/). Finally, we note that the motor neuron differentiation efficiencies that were experimentally derived by Boulting *et al.* provided a genuine test set for determining the predictive power of the lineage scorecard: The bioinformatic algorithms of the lineage scorecard had already been finalized before the first comparisons between the two datasets were made, and no aspects of the scorecard were retrospectively optimized to improve the fit.

Bioinformatic analysis and data access

In addition to method-specific data normalization and the calculation of the scorecard (described above), bioinformatic analyses were conducted as follows:

(i) *Hierarchical clustering (Figures 1 and 3, Supplementary Figures 1 and 2)*. DNA methylation levels were calculated as the coverage-weighted average over all CpGs in the promoter regions of Ensembl-annotated transcripts; gene expression levels were calculated for each Ensembl gene by averaging over all associated probes on the microarray. Prior to hierarchical clustering the two datasets were separately normalized to zero mean and unit variance in order to give equal weight to both datasets. The heatmaps show a representative selection of 250 genes. Hierarchical clustering was performed in R (http://www.r-project.org/), using a Euclidean distance function and the average-linkage method.

(ii) *Annotation clustering and promoter characteristics (Figure 2D)*. Identification of common characteristics among the most variable genes was performed using DAVID (Huang et al., 2007) and EpiGRAPH (Bock et al., 2009) with default parameters and based on Ensembl gene annotations (promoters were defined as the -5kb to +1kb sequence window surrounding the transcription start site).

(iii) *Classification of ES vs. iPS cell lines (Figure 3D)*. To validate the previously reported iPS gene signatures, the mean DNA methylation or expression level over all genes in a given signature was calculated from the current dataset. Logistic regression was used for selecting the most discriminatory threshold, and the predictiveness of each signature was evaluated by leave-one-out cross-validation. To derive new classifiers, support vector machines were trained on the DNA methylation data, the gene expression data, or the combination of both datasets.

Each classification was based on 7500 randomly selected attributes, which was the maximum number of attributes that were computationally feasible in a single analysis. The predictiveness of all classifiers was evaluated by leave-one-out cross-validation, and the average performance over 100 classifications with random attribute sets are reported in Figure 3D. Note that none of these classifications used feature selection. It is likely that supervised or unsupervised feature selection could increase the prediction accuracy, but in the absence of a second validation dataset it is unclear whether such an improvement reflects a genuine increase in predictiveness or overfitting to the current dataset. All predictions were performed using the Weka software (Frank et al., 2004)

(iv) Linear models of epigenetic memory. Two alternative linear models were constructed for both DNA methylation and gene expression. The first model regresses the iPS-cell specific mean DNA methylation (or gene expression) levels of each gene on the ES-cell specific mean DNA methylation (or gene expression) levels. The second model regresses the iPS-cell specific mean DNA methylation (or gene expression) levels of each gene on the ES-cell specific and the fibroblast-specific mean DNA methylation (or gene expression) levels. Both models were compared by an analysis of variance (ANOVA). All calculations were performed in R (http://www.r-project.org/).

Technical notes

1. Are all three scorecard assays required? Our data support the importance of all three assays: (i) Epigenetic defects at genes such as *CD14* (which interferes with macrophage differentiation in HUES8) and *GRM1* (which may affect motor neuron function in hiPS 17a) are more readily identified by DNA methylation profiling than by gene expression profiling because both genes show little to no detectable expression in pluripotent cell lines. (ii) Gene expression data of pluripotent cell lines provide an efficient way of identifying partially reprogrammed cell lines (Mikkelsen et al., 2008; Sridharan et al., 2009), although this was not a major concern in the current study. (iii) Only the quantitative differentiation assay could accurately predict which cell lines could most efficiently differentiate into motor neurons, but because this assay covers only 500 selected marker genes it is likely to miss relevant cell-line specific defects at other genes.

2. Is the deviation scorecard compatible with other genomic assays? The current deviation scorecard uses two assays that were readily available to us, namely Affymetrix high-throughput microarrays and the RRBS protocol, which performed well in a recent technology comparison (Bock et al., 2010). But there is no reason why the lineage scorecard could not be combined with other assays for measuring DNA methylation and gene expression. To facilitate the use of other assays, we investigated whether it is possible to reduce the number of genes for which DNA methylation and gene expression are measured. This appears to be a reasonable strategy given our observation that epigenetic and transcriptional variation is common only at a small number of genes (Figure 2A). Based on our dataset, we would detect almost 90% of the DNA methylation deviations in iPS cell lines by monitoring only the 10% most variable genes in ES cells (Supplementary Figure 4D). Focusing on the ~3,000 most variable genes (plus maybe another ~1,000 manually selected genes that should be monitored even for rare defects) brings the number of promoter regions well within the range commercial epigenotyping assays (Bibikova et al., 2009), which are widely available through microarray core facilities. In contrast, for gene expression it is not possible to focus on a small number of most variable genes in ES cells while still capturing the lion's share of iPS-specific deviations (Supplementary Figure 4D). However, this is less likely to constitute a practical limitation because commercial gene expression microarrays are widely available, easy-to-use and relatively cost-efficient.

3. Is 16-day EB differentiation essential for obtaining good results with the lineage scorecard? We explored several ways to speed up the quantitative differentiation assay, which in its basic version uses a 16-day EB differentiation step. In order to minimize the time-to-results and cell culture costs, it would be ideal if our assay were sensitive enough to estimate differentiation propensities using RNA isolated directly from the undifferentiated pluripotent cell lines, for example by measuring low levels of spontaneous differentiation in otherwise self-renewing cell cultures. We initially tested this approach based on the gene expression microarray data, but could

not pick up any informative signal. This could either be due to the relatively low dynamic range of microarrays compared to the nCounter assay that was used for the lineage scorecard (Geiss et al., 2008) or due to the fact that only a single microarray replicate was available for most cell lines. However, when we profiled RNA from the ES and iPS cell lines under self-renewing conditions in two biological replicates using the nCounter assay, we were able to construct as scorecard that was moderately correlated with EB-based lineage scorecard (Supplementary Figure 4E). These data suggest that it may be possible to derive an accurate lineage scorecard based only on RNA from the pluripotent cell lines themselves. However, we observed a substantially reduced dynamic range for these predictions (Supplementary Figure 4E), indicating that abandoning the EB differentiation may decrease the robustness of the lineage scorecard. As an alternative, we also evaluated whether the duration of the EB assay could be reduced from 16 days to 7 days. In a preliminary comparison we observed excellent agreement (Pearson's $r>0.9$) between the two assays on four representative iPS cell lines. While further optimization and validation are warranted, these results indicate that it should be possible to reduce the duration of the differentiation assay without jeopardizing its accuracy.

4. How reproducible are the DNA methylation and gene expression measurements? For DNA methylation, we have recently confirmed that almost perfect technical reproducibility (Pearson's $r=0.98$) can be achieved between two researchers conducting the RRBS protocol on the same DNA (H. Gu, unpublished data). High technical reproducibility is also a well-established fact for gene expression microarrays (Irizarry et al., 2005). Furthermore, when we collected two biological replicates of the same cell line one passage apart, we observed high correlations for both DNA methylation (Pearson's $r>0.981$) and gene expression (Pearson's $r>0.986$). In contrast, the similarity across cell lines was lower than between biological replicates of the same cell line, for DNA methylation ($0.965<r<0.978$) as well as for gene expression ($0.973<r<0.987$). Although low in absolute terms, these differences provide further confirmation that human ES cell lines exhibit DNA methylation and gene expression alterations that are cell-line specific and stable over a limited number of passages (Supplementary Figure 1D). We also compared the DNA methylation profiles obtained for H1 cells grown in our labs to those of H1 cells that were grown elsewhere (Harris et al., 2010). Correlations were again high (Pearson's $r>0.965$), confirming that potential handling differences did not have a strong effect on DNA methylation or gene expression (Supplementary Figure 1D).

5. Is it necessary to recalibrate the assays? Differentiation of pluripotent cells into EB is generally robust, but small differences in the experimental parameters such as physical handling, media renewal and plasticware could still have an effect on the outcome. To test the relevance of this effect for the lineage scorecard, we performed a systematic comparison in which several biological replicates of one cell line (hiPS 17b) were cultured for two passages by two different investigators in two different labs, who also performed the EB differentiation separately and independently. The correlation of lineage scorecard predictions between biological replicates done by the same researcher in the same lab was high (Pearson's $r=0.82$), although not quite in the same range as the correlations observed for the RRBS and microarray analyses. This result is not unexpected given the higher complexity of the assay, and we were able to successfully compensate for the somewhat higher variability by performing at least 2-3 biological replicates of the quantitative differentiation assay for each cell line. When we compared the correlation of lineage scorecard predictions between two researchers in two different labs, we observed a correlation that was further reduced but could still be considered reproducible (Pearson's $r=0.59$). Based on these data we suggest that researchers who use the lineage scorecard for the first time should profile a few high-quality ES or iPS cell lines grown in their labs to recalibrate the assay. In contrast, such a calibration does not appear to be necessary for the DNA methylation and gene expression data that give rise to the deviation scorecard.

6. Do the scorecard measurements apply to all instances of an ES cell line? We generally observed high reproducibility over a low to moderate number of passages (5 to 15 passages) when the cells are being handled by the same researcher. However, we also found that lab stocks of high-passage ES cell lines that had been maintained by different people for several dozen passages were as distinct from the corresponding low-passage freezes maintained under standardized culture conditions as two different cell lines were from each other. This observa-

tion, while anecdotal, suggests that the properties that are measured by the scorecard can substantially change over time, especially when the cell lines are being cultured under varying conditions. We would therefore like to underline that all scorecard measurements reported in the current paper refer to the specific freezes of ES and iPS cell lines that were used for this study, and they are unlikely to generalize to other instances of ES and iPS cell lines that are cultured elsewhere. The safest and most informative way to make conclusions about specific instances of pluripotent cell lines will be to perform the scorecard assay in the same cells (plus/minus a few passages) that will be used for the application of interest. We also recommend repeating the scorecard assays regularly (e.g., every 10 passages, as is common practice for karyotyping) to detect deteriorations in cell line quality early on.

7. How to adapt the lineage scorecard for predicting other differentiation protocols? The current paper shows that the lineage scorecard accurately predicts the efficiency of directed differentiation into motor neurons using an EB-based protocol (Figure 6,). Additional validation data suggest that the lineage scorecard also detects the differentiation bias induced by culture conditions that favor ectoderm or mesoderm differentiation (Supplementary Figure 3B). From a bioinformatic perspective, it is easy to extend the lineage scorecard to other lineages, simply by collecting gene sets that are relevant for the lineage of interest and including them as a separate gene set group in the calculation of the lineage scorecard (Supplementary Figure 4C). In many cases this could be sufficient to predict the propensity for directed differentiation into cell types other than motor neurons. However, it is possible that certain cell types and differentiation protocols need specific adaptations of the scorecard. For example, both the current lineage scorecard and the motor neuron differentiation protocol are EB-based, while there are many directed differentiation protocols that do not go through an EB phase. To adapt the scorecard to a new type of differentiation protocols, one would conduct the initial step of the directed differentiation protocol (in the absence of further differentiation clues) for a number of high-quality pluripotent cell lines and recalibrate the lineage scorecard relative to this reference.

Supplementary References

Bibikova, M., Le, J., Barnes, B., Saedinia-Melnyk, S., Zhou, L., Shen, R., and Gunderson, K.L. (2009). Genome-wide DNA methylation profiling using Infinium assay. Epigenomics *1*, 177-200.

Bock, C., Halachev, K., Büch, J., and Lengauer, T. (2009). EpiGRAPH: User-friendly software for statistical analysis and prediction of (epi-) genomic data. Genome biology *10*, R14.

Bock, C., Reither, S., Mikeska, T., Paulsen, M., Walter, J., and Lengauer, T. (2005). BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. Bioinformatics *21*, 4067-4068.

Bock, C., Tomazou, E.M., Brinkman, A.B., Muller, F., Simmer, F., Gu, H., Jager, N., Gnirke, A., Stunnenberg, H.G., and Meissner, A. (2010). Quantitative comparison of genome-wide DNA methylation mapping technologies. Nat Biotechnol.

Di Giorgio, F.P., Boulting, G.L., Bobrowicz, S., and Eggan, K.C. (2008). Human embryonic stem cell-derived motor neurons are sensitive to the toxic effect of glial cells carrying an ALS-causing mutation. Cell Stem Cell *3*, 637-648.

Frank, E., Hall, M., Trigg, L., Holmes, G., and Witten, I.H. (2004). Data mining in bioinformatics using Weka. Bioinformatics *20*, 2479-2481.

Garnache-Ottou, F., Chaperot, L., Biichle, S., Ferrand, C., Remy-Martin, J.P., Deconinck, E., de Tailly, P.D., Bulabois, B., Poulet, J., Kuhlein, E., *et al.* (2005). Expression of the myeloid-associated marker CD33 is not an exclusive factor for leukemic plasmacytoid dendritic cells. Blood *105*, 1256-1264.

Geiss, G.K., Bumgarner, R.E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D.L., Fell, H.P., Ferree, S., George, R.D., Grogan, T., *et al.* (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs. Nature Biotechnology *26*, 317-325.

Grigoriadis, A.E., Kennedy, M., Bozec, A., Brunton, F., Stenbeck, G., Park, I.H., Wagner, E.F., and Keller, G.M. (2010). Directed differentiation of hematopoietic precursors and functional osteoclasts from human ES and iPS cells. Blood *115*, 2769-2776.

Harris, R.A., Wang, T., Coarfa, C., Nagarajan, R.P., Hong, C., Downey, S.L., Johnson, B.E., Fouse, S.D., Delaney, A., Zhao, Y., *et al.* (2010). Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nat Biotechnol.

Huang, D.W., Sherman, B.T., Tan, Q., Kir, J., Liu, D., Bryant, D., Guo, Y., Stephens, R., Baseler, M.W., Lane, H.C., *et al.* (2007). DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Research *35*, W169-175.

Irizarry, R.A., Warren, D., Spencer, F., Kim, I.F., Biswal, S., Frank, B.C., Gabrielson, E., Garcia, J.G., Geoghegan, J., Germino, G., *et al.* (2005). Multiple-laboratory comparison of microarray platforms. Nature Methods *2*, 345-350.

Laflamme, M.A., Chen, K.Y., Naumova, A.V., Muskheli, V., Fugate, J.A., Dupras, S.K., Reinecke, H., Xu, C., Hassanipour, M., Police, S., *et al.* (2007). Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. Nat Biotechnol *25*, 1015-1024.

Li, Y., Lee, P.Y., Kellner, E.S., Paulus, M., Switanek, J., Xu, Y., Zhuang, H., Sobel, E.S., Segal, M.S., Satoh, M., *et al.* (2010). Monocyte surface expression of Fcgamma receptor RI (CD64), a biomarker reflecting type-I interferon levels in systemic lupus erythematosus. Arthritis Res Ther *12*, R90.

Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature *454*, 49-55.

Nam, D., and Kim, S.Y. (2008). Gene-set approach for expression pattern analysis. Brief Bioinform *9*, 189-197.

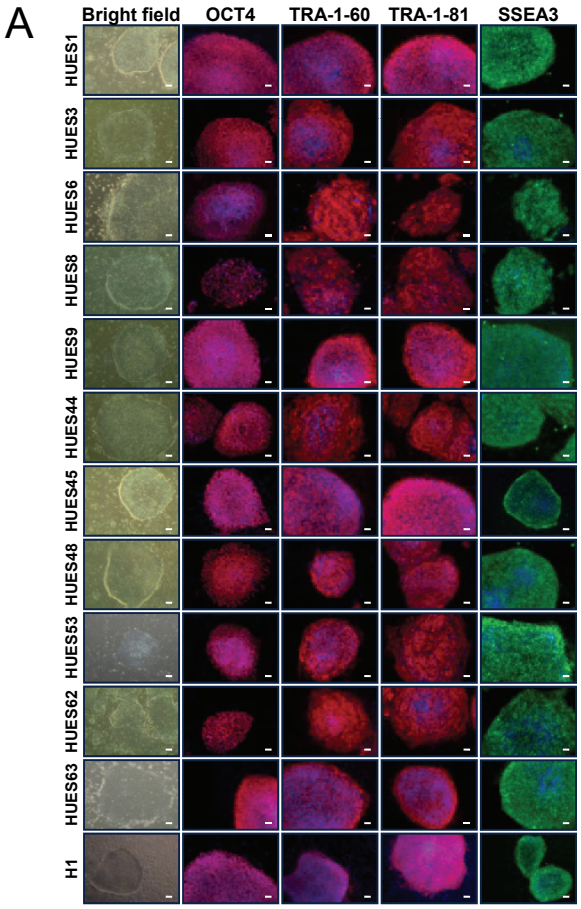
Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol *3*, Article3.

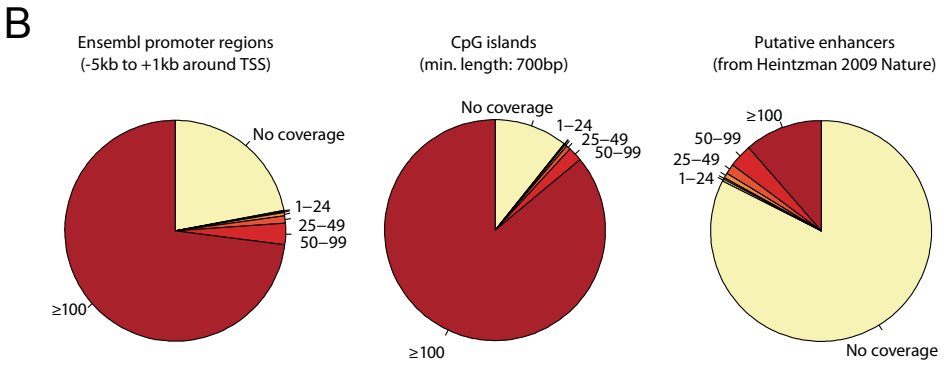
Sridharan, R., Tchieu, J., Mason, M.J., Yachechko, R., Kuoy, E., Horvath, S., Zhou, Q., and Plath, K. (2009). Role of the murine reprogramming factors in the induction of pluripotency. Cell *136*, 364-377.

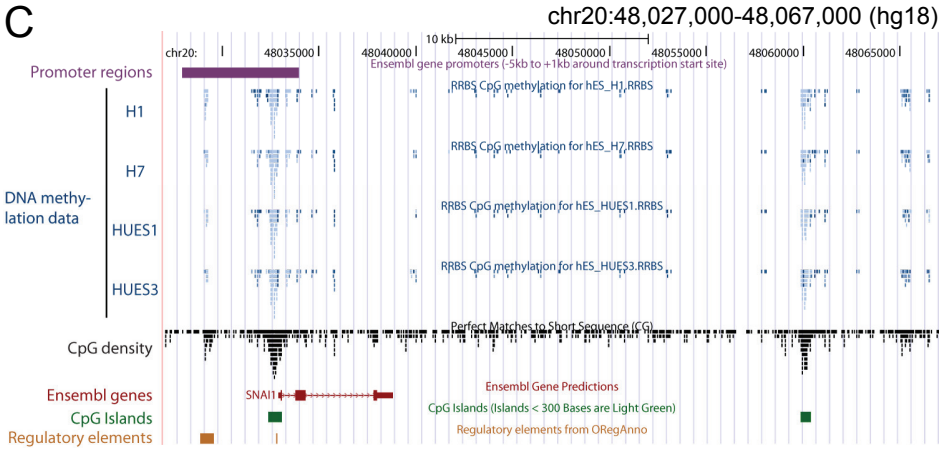
Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., *et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.

Tukey, J.W. (1977). Exploratory data analysis (Reading, Mass., Addison-Wesley Pub. Co.).

Supplementary Figure 1 (related to Figure 1)



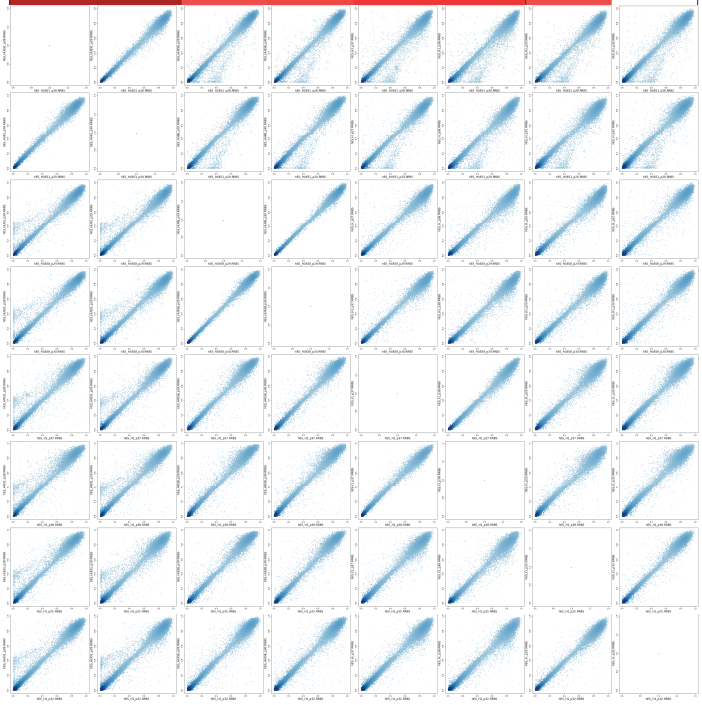




D

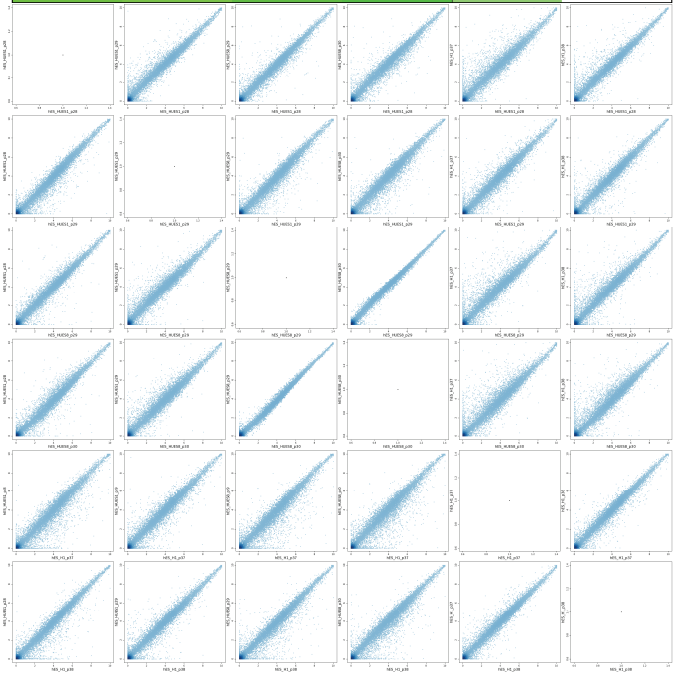
DNA methylation

	HUES1 (p28)	HUES1 (p29)	HUES8 (p29)	HUES8 (p30)	H1 (p37)	H1 (p38)	H1 (p25)	H1 (p32)
HUES1 (p28)		0.986	0.974	0.973	0.965	0.965	0.968	0.966
HUES1 (p29)	0.986		0.974	0.973	0.967	0.967	0.966	0.965
HUES8 (p29)	0.974	0.974		0.989	0.978	0.978	0.979	0.977
HUES8 (p30)	0.973	0.973	0.989		0.977	0.977	0.978	0.976
H1 (p37)	0.965	0.967	0.978	0.977		0.981	0.974	0.974
H1 (p38)	0.965	0.967	0.978	0.977	0.981		0.975	0.974
H1 (p25)	0.968	0.966	0.979	0.978	0.974	0.975		0.977
H1 (p32)	0.966	0.965	0.977	0.976	0.974	0.974	0.977	

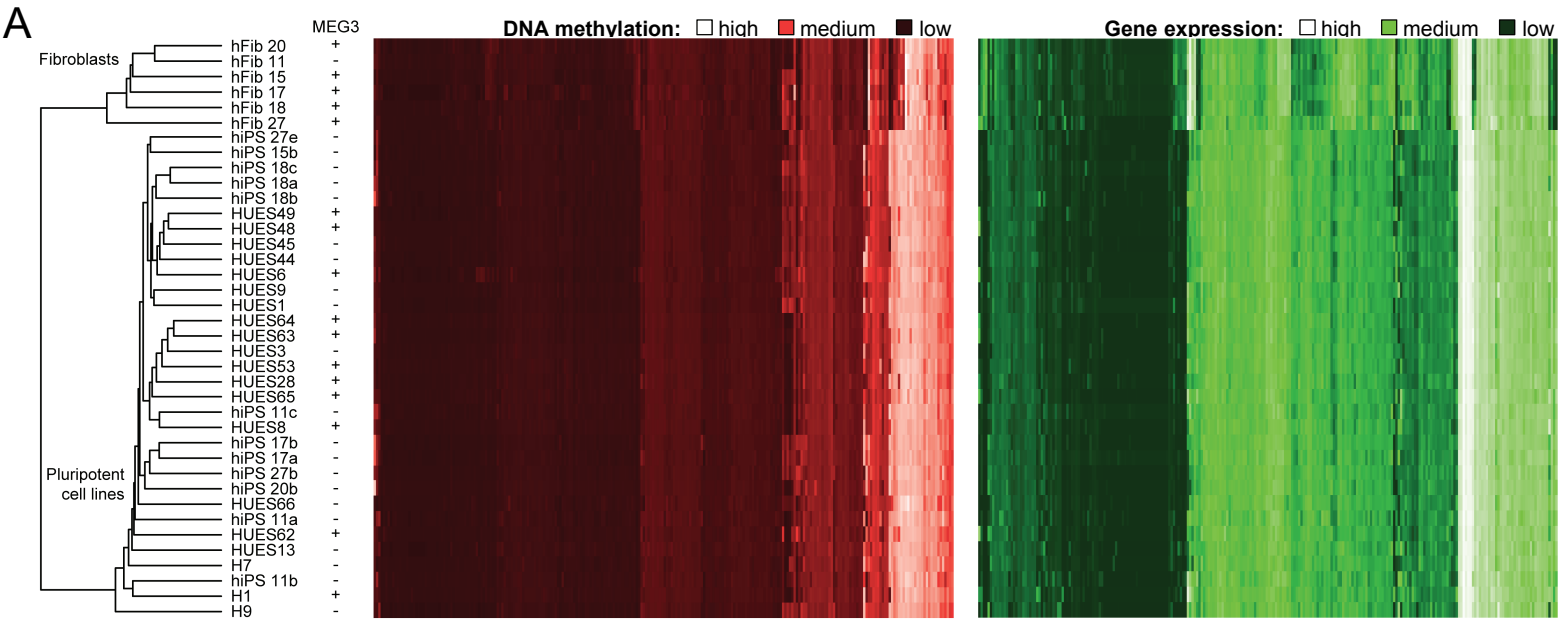


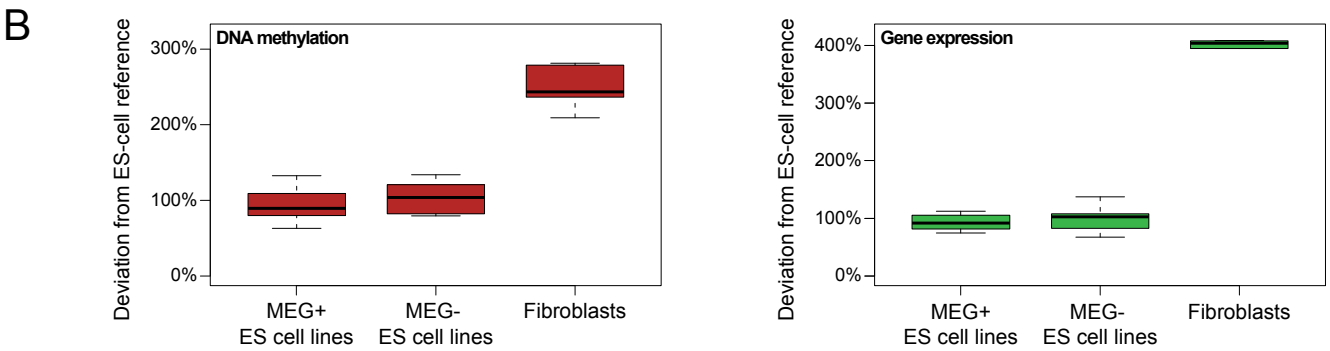
Gene expression

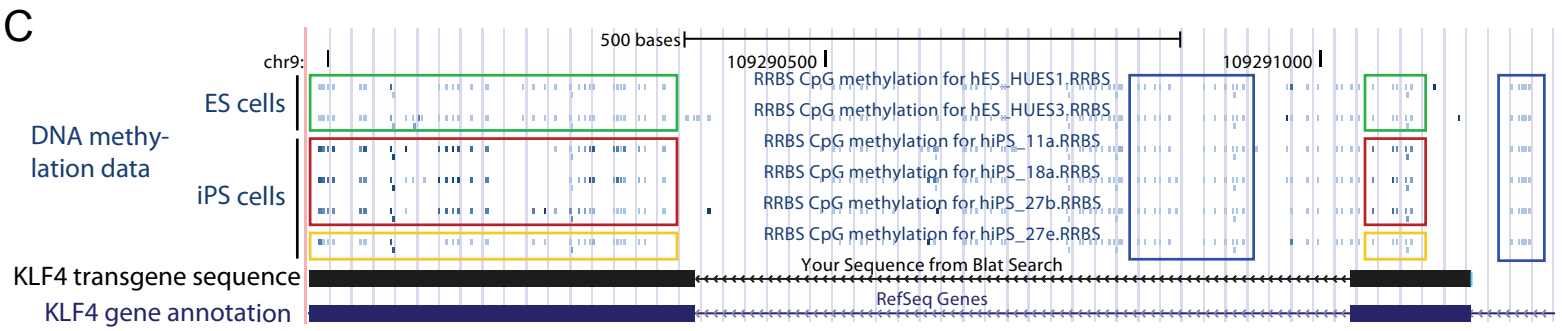
	HUES1 (p28)	HUES1 (p29)	HUES8 (p29)	HUES8 (p30)	H1 (p37)	H1 (p38)
HUES1 (p28)		0.986	0.987	0.984	0.975	0.982
HUES1 (p29)	0.986		0.983	0.983	0.983	0.985
HUES8 (p29)	0.987	0.983		0.997	0.974	0.979
HUES8 (p30)	0.984	0.983	0.997		0.973	0.977
H1 (p37)	0.975	0.983	0.974	0.973		0.989
H1 (p38)	0.982	0.985	0.979	0.977	0.989	



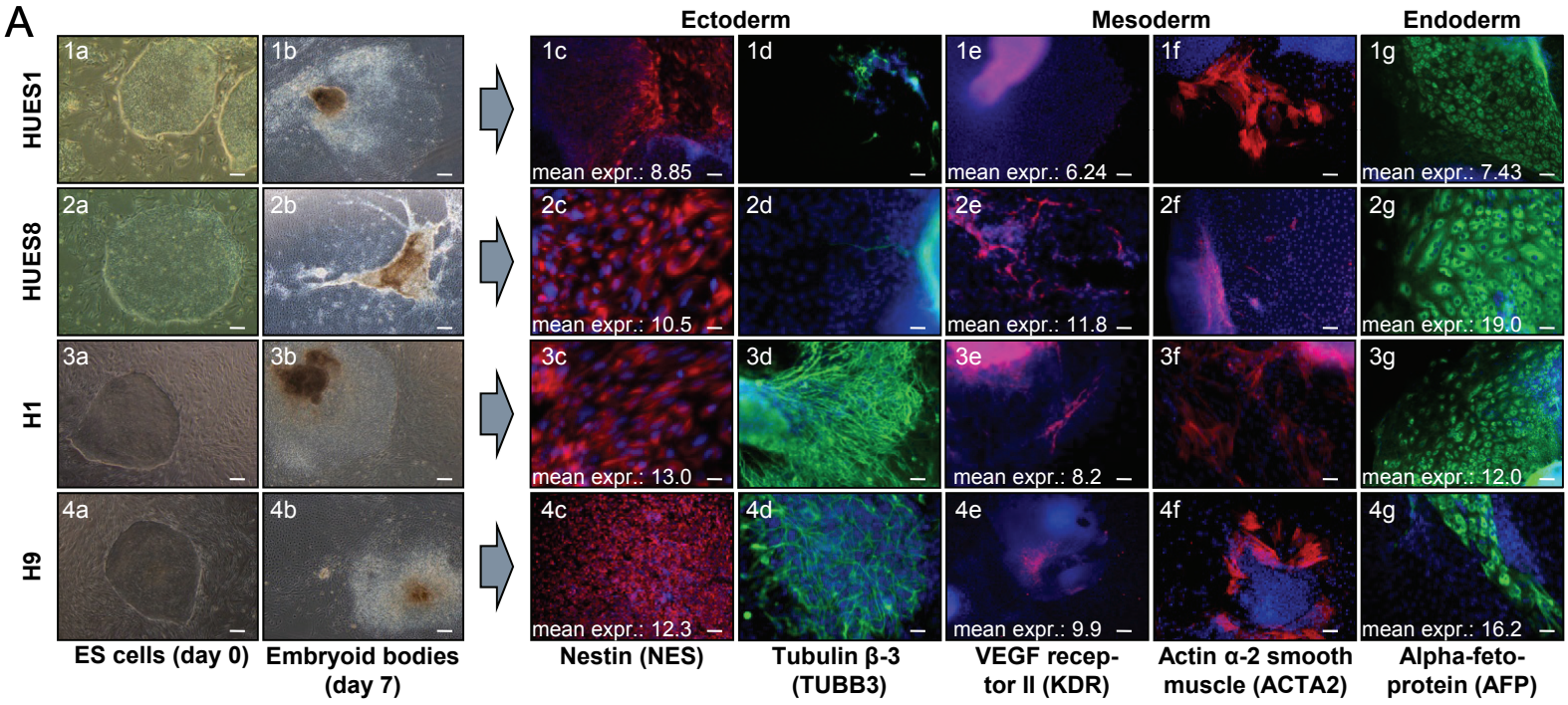
Supplementary Figure 2 (related to Figure 3)

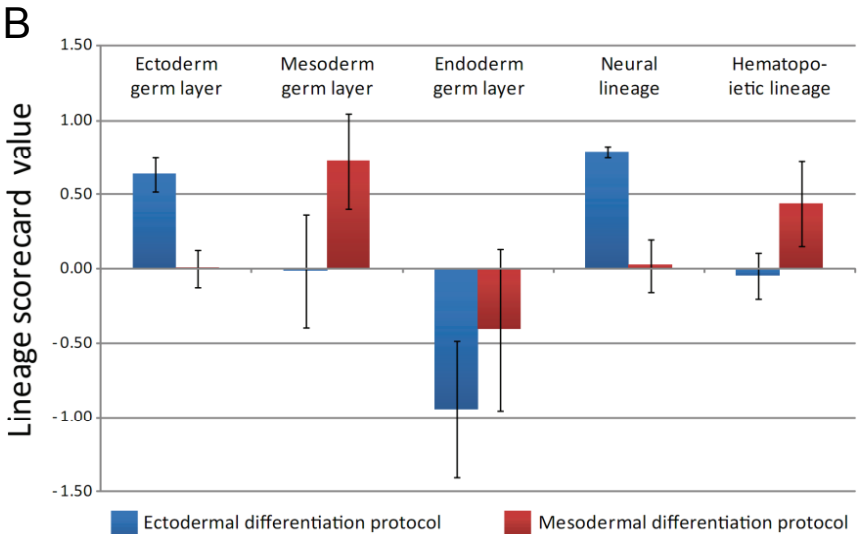


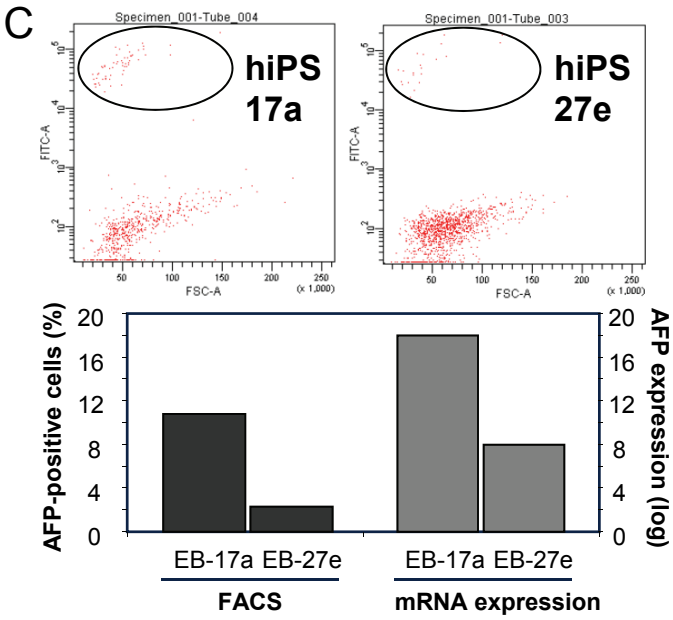




Supplementary Figure 3 (related to Figure 5)







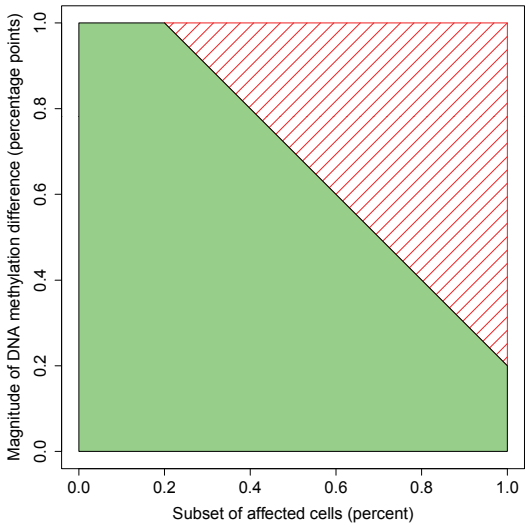
Supplementary Figure 4 (related to Experimental Procedures)

A

Algorithm for calculating the deviation scorecard

1. **Data Import**: Load DNA methylation data (as measured by reduced-representation bisulfite sequencing) and gene expression data (as measured by Affymetrix microarrays) for the following samples: (i) the cell lines of interest, and (ii) a reference set of high-quality pluripotent cell lines
2. **Data Normalization**: Perform gcRMA normalization of the Affymetrix microarrays and scale all gene expression values to a target interval ranging from 0 to 10
3. **Gene Mapping**: Load Ensembl gene annotations and calculate – for each gene – the DNA methylation level (averaging over all CpGs in a promoter region) and the gene expression level (averaging over alternative transcripts). A weighting scheme corrects for differential sequencing coverage between samples
4. **Reference Comparison**: Calculate the expected range of DNA methylation and expression values based on the reference set of high-quality pluripotent cell lines and flag cell lines of interest as outliers if their value for a specific gene falls outside of the center quartiles by more than 1.5 times the interquartile range (Tukey's outlier filter)
5. **Relevance Filter**: Discard minor outliers with a DNA methylation difference of less than 20 percentage points or an expression change of less than twofold
6. **Gene Sets**: Load gene sets containing relevant genes for the application of interest, such as lineage marker genes or cancer genes
7. **Outlier Summary**: List the number of deviations for each cell line, together with a list of affected genes (Figure 4B and Supplementary Table 5)

B

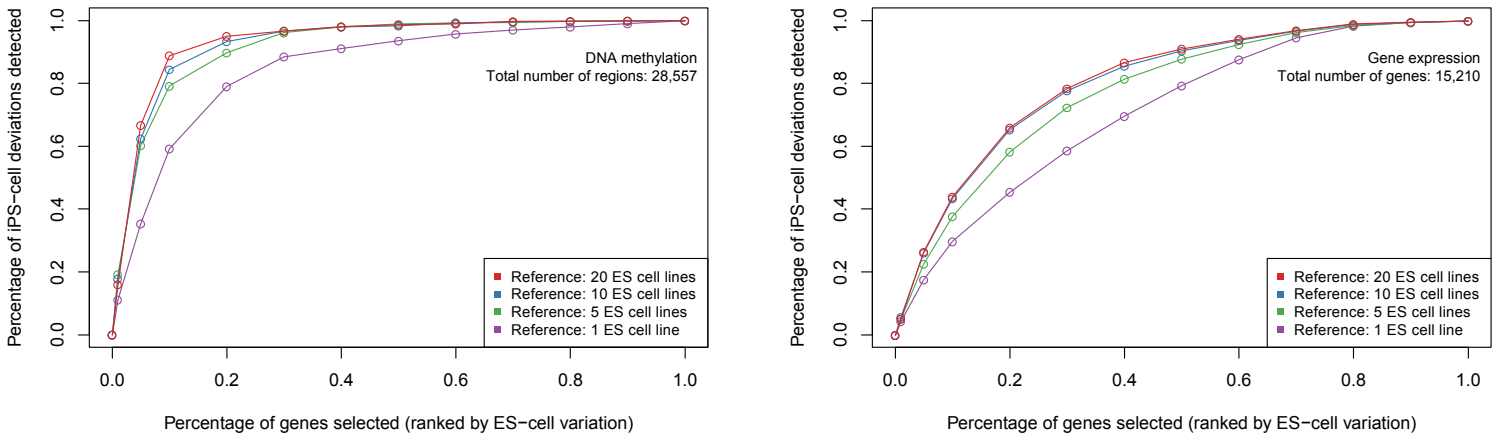


C

Algorithm for calculating the lineage scorecard

1. **Data Import**: Load absolute transcript counts (as measured by the nCounter system) for 500 marker genes in the following samples: (i) embryoid bodies (EB) derived from the cell lines of interest, and (ii) a reference set of EBs derived from high-quality pluripotent cell lines
2. **Assay Normalization**: Use positive spike-in controls to calculate an assay normalization factor and rescale the data accordingly (separately for each experiment)
3. **Sample Normalization**: Perform variance stabilization and normalization across all experiments (using Bioconductor's VSN package)
4. **Reference Comparison**: Perform moderated *t*-tests for each marker gene, comparing all EB replicates for the cell line of interest against the reference set of high-quality EBs (using Bioconductor's limma package)
5. **Gene Sets**: Load gene sets that are characteristic for the cellular lineage or germ layer of interest (gene sets can be obtained from Gene Ontology, MolSigDB or from manual curation efforts)
6. **Enrichment Analysis**: For each gene set calculate the mean *t*-scores of all marker genes that belong to the gene set
7. **Scorecard Calculation**: For each cell line, the mean of the *t*-scores for all relevant gene sets constitutes the scorecard estimate for the lineage that the gene sets represent (Figure 5B, D)

D



E