



Published in final edited form as:

Nat Biotechnol. 2015 November ; 33(11): 1182–1192. doi:10.1038/nbt.3387.

An improved ScoreCard to assess the differentiation potential of human pluripotent stem cells

Alexander M. Tsankov^{1,2,3}, **Veronika Akopian**^{1,2,3}, **Ramona Pop**^{1,2,3}, **Sundari Chetty**^{2,3}, **Casey A. Gifford**^{1,2,3}, **Laurence Daheron**², **Douglas A. Melton**^{2,3,4}, **Nadejda M. Tsankova**^{5,6}, and **Alexander Meissner**^{1,2,3,*}

¹ Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA.

² Harvard Stem Cell Institute, Cambridge, Massachusetts 02138, USA.

³ Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA.

⁴ Howard Hughes Medical Institute, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA.

⁵ Department of Pathology and Cell Biology, Columbia University Medical Center, New York, New York 10032, USA.

⁶ Department of Pathology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

Abstract

Research on human pluripotent stem cells has been hampered by the lack of a standardized, quantitative, scalable assay of pluripotency. We have previously described an assay called ScoreCard that used gene expression signatures to quantify differentiation efficiency. Here we report an improved version of the assay based on qPCR that enables faster, more quantitative assessment of functional pluripotency. We provide an in-depth characterization of the revised signature panel through embryoid body and directed differentiation experiments as well as a detailed comparison to the teratoma assay. We also show that the improved ScoreCard enables applications such as screening of small molecules, genetic perturbations and assessment of culture conditions. Beyond stem cell applications, this approach can in principle be extended to other cell types and lineages.

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

*Correspondence and requests for materials should be addressed to Alexander Meissner (alexander_meissner@harvard.edu).

Author contributions

A.T. and A.M. designed and conceived the study. A.T. optimized all experiments and performed all analysis, V.A. performed cell culture and ran qPCR plates, R.P. performed FACS analysis and ran qPCR plates, C.G. trained A.T. in cell culture and provided experimental advice, S.C. performed PiPS cell culture and EB experiments with supervision from L.D. and D.M., N.T. performed all pathological and histological annotations, A.T. and A.M. interpreted the data and wrote the manuscript.

Financial Disclosures

Harvard University has filed a patent on the original ScoreCard technology and a licensing agreement with Life Technologies (now part of Thermo Fisher). A.M. is an advisor to Life Technologies.

Human pluripotent stem cells (hPSCs) can give rise to all cell types in the body and therefore hold enormous potential for tissue engineering, regenerative medicine and disease modeling. Several major initiatives are under way around the world to produce human induced pluripotent stem cells (hiPSCs) at large scales^{1, 2}. The growing numbers of hiPSC lines and of NIH-registered human embryonic stem cell (hESC) lines are improving access to hPSCs for researchers and should facilitate progress toward therapeutic applications³. These developments underscore the need for hPSC quality standards that are sufficiently stringent to ensure comparable and reproducible results across laboratories⁴. The need for a ‘gold standard’ scalable, quantitative assay of pluripotency is becoming ever more acute as the numbers of lines, culture conditions, and hPSC laboratories continue to increase and as therapies based on hPSCs are advanced to clinical translation.

Formation of teratomas in mice is the most frequently used assay for characterizing the differentiation potential of hPSCs. However, the generation of teratomas requires large numbers of mice and is not scalable to the number of hPSC lines that will be created in the years to come. Moreover, it is a time-consuming assay whose results are highly variable and difficult to quantify^{4, 5}. Recent studies have begun to use genomic approaches^{6, 7} as a more quantitative, efficient way to assess the quality and potential of hPSCs. Although these studies share the principle of gene expression signatures, they measure distinct aspects of pluripotency. PluriTest⁶ measures the molecular signature of pluripotency and uses this to classify pluripotent samples with great sensitivity and specificity. In contrast, the ScoreCard⁷ approach evaluates the molecular signature of pluripotency and expression signatures that indicate functional pluripotency, defined as differentiation into each of the three germ layers. However, the initial ScoreCard was not optimized for early germ layer differentiation, used the NanoString platform that is not available to most laboratories and required customized downstream analysis, restricting its adoption by the community.

To overcome these limitations, we developed a more accessible ScoreCard assay that uses qPCR measurements of a revised set of genes and provides improved statistical analysis, accuracy, and utility for a wider array of applications. We demonstrate applications, including directed differentiation and quantitative screening experiments, that would not be possible using the previous genomic approaches^{6, 7}. Our results further support the advantages of gene expression measurements for the rapid and quantitative characterization of cell types, lineage regulators, and culture conditions.

RESULTS

Characterization of hPSC lines using standard assays

In order to establish a reference point, we selected five commonly used hESC lines from the NIH registry that have shown some variability in their differentiation potential in the past^{7, 8} and performed standard assays to characterize them. All lines displayed the typical morphology (**Fig. 1a**, top row) and stained positive for the pluripotency-associated markers OCT4 and TRA1-60 (**Fig. 1a**, bottom rows). We next performed global expression analysis using RNA-seq of polyadenylated transcripts (**Fig. 1b**; **Supplementary Table 1**) and found expression levels of selected pluripotency associated markers to be 10-1000 times higher

than those of known markers of early differentiation, supporting the molecular pluripotency of these lines. We also performed karyotyping (**Supplementary Fig. 1a**) and injected the five hESC lines as well as an additional hiPSC line (1-51C) into the kidney capsule of immunocompromised mice for teratoma formation (**Fig. 1c; Supplementary Fig. 1b**), which confirmed the functional pluripotency of the selected lines.

To obtain a more accurate and quantitative assessment, we performed detailed pathology analysis using high-throughput imaging (**Fig. 1c; Supplementary Fig. 2; Supplementary Table 2 and Online Methods**). This analysis further showed that each teratoma contained a much higher fraction of ectoderm (EC)- and mesoderm (ME)-derived tissues than endoderm (EN) (**Fig. 1d**), which appears to be consistent across all lines and replicates that we studied. This information is not provided by standard teratoma analysis. Lastly, the teratoma assay shows an inherent variability between biological replicates (same cells, but different mice injected) and even between different sections of the same teratoma. The fraction of EC- and ME-derived nuclei varied between biological replicates of line HUES64 and H9, whereas technical replicates H9_1a and H9_1b were highly similar (**Fig. 1d**). The within-group variability between replicates for the percentage of EC and ME nuclei ($\sigma_{EC} = 16\%$; $\sigma_{ME} = 20\%$) was similar to the variability between cell lines ($\sigma_{EC} = 21\%$; $\sigma_{ME} = 20\%$). Samples also showed high variation between different teratoma sections ($\sigma_{EC} = 15\%$; $\sigma_{ME} = 14\%$) in the pathology quantification (**Fig. 1e**) and in the RNA expression level of germ layer-specific marker genes (**Fig. 1f**). Taken together, the results clearly established the pluripotency of all lines, but despite detailed analysis, the teratomas provided only limited quantitative information to make meaningful statements about the quality of the chosen lines.

Gene expression signatures using a qPCR panel

We have previously shown that gene expression signatures can effectively quantify the differentiation potential of hPSC lines⁷. To improve and expand this approach, we designed a qPCR-based assay for measuring total RNA level of 96 carefully selected genes. Our design included five housekeeping probes for data normalization and gene markers specific to pluripotent stem cells (PL), mesendoderm (MS), and the three embryonic germ layers, EN, ME, and EC (9, 6, 22, 25, and 21 genes, respectively; **Fig. 2a, Supplementary Fig. 3a**). Markers of the three germ layers were chosen based on their uniqueness of expression according to RNA-seq data⁹ on purified HUES64 derived EN, ME and EC populations (**Online Methods**). Notably, this approach allowed us to select germ layer markers with 8-16 fold higher mean uniqueness of expression than those in the previous ScoreCard panel⁷ (**Supplementary Fig. 3b**). Although the markers were selected based on HUES64 data, the uniqueness of expression for the chosen genes correlated very well with the mean uniqueness of expression across all profiled hPSC lines⁹ (**Supplementary Fig. 3c**). We next obtained TaqMan plates containing the selected genes and confirmed that expression levels were highly reproducible between technical replicates on different plate lots, and for lower RNA input (**Supplementary Fig. 4a**).

Next we expanded our sample set to 23 hPSC lines, which we differentiated into embryoid bodies (EB) by withdrawing growth factors in suspension (**Online Methods**). RNA was

collected from cells in the undifferentiated state (day=0) as well as at 2, 5, and 12 days post differentiation. Biological replicates of our EB time course for cell lines with similar passage numbers and more than ten passages apart were well correlated in gene expression values ($R > .92$, **Supplementary Fig. 4b,c**). We then normalized the data against the expression of five housekeeping genes that were most invariant across >80 experiments (**Supplementary Table 3**). The normalization technique reduced variance between replicates of the same cell line relative to the variance between different cell lines (**Supplementary Fig. 4d**).

Characterizing EB differentiation potential

To quantify differentiation potential using EB formation across cell lines, we improved our previous computational analysis⁷. Specifically, for each gene in an EB experiment, we compared its expression relative to the expression of that gene at day 0 of EB differentiation for a reference set of the 23 hPSC lines and calculated a P value P_k (t -test; **Fig. 2a**, middle). The resulting P values were then merged within each gene set (PL, EN, ME, and EC) using the weighted Z -method¹⁰ (**Fig. 2a**, right), which combines the information across multiple statistical tests of the same null hypothesis, weighing tests based on their importance and taking into account dependencies between tests (**Online Methods**). Weights were set to the average cell line expression difference between day 5 or 12 of EB formation and day 0 (**Supplementary Table 4**), and genes with negative weights were excluded from the analysis. Scatter plot of day 5 and day 12 weights showed strong correlation ($R = .93$, **Fig. 2b**). We refer to the weighted Z -test standard normal deviate $Z_{w,GS}$ for each gene set GS as the ‘differentiation potential_{EB},’ which measures the distance traveled away from a pluripotent state.

The differentiation potential_{EB} for the three germ layers gradually increased during the EB differentiation time course (**Fig. 2c**), with endoderm induction lagging behind mesoderm and ectoderm. At day five, 14 of the 23 lines showed significant increases in expression of EC, ME, and EN gene markers ($P < .05$, **Fig. 2d**). At day 12, all 23 lines showed significant ($P < .005$) differential expression from the pluripotent state for all three germ layers (**Fig. 2d**, right), consistent with previous functional characterization of these lines^{7, 11}. Based on this result, we chose day 12 for the quantification of differentiation potential_{EB} in the subsequent experiments. The weighted Z -method increased the overall differentiation potential_{EB} and combined P value for most cell lines when compared with the Stouffer's (unweighted) Z -transform test¹² (**Supplementary Fig. 4e,f**). To contrast the established pluripotent lines, we obtained several partially reprogrammed iPS (PiPS) cell lines and quantified their differentiation potential (**Fig. 2c,d** bottom). Prior to differentiation, these lines' germ layer differentiation potential_{EB} was in the range of other hPSC lines. Not surprisingly, after 5 or 12 days of EB differentiation, the PiPS were unable to activate differentiation in the same manner as hPSCs (**Fig. 2c**, bottom; **Supplementary Fig. 4g**). Notably, we detected a weak signal for ME differentiation, likely because these cell lines were reprogrammed from human fibroblasts, a mesodermal tissue.

To calculate the minimum number of genes needed to accurately measure differentiation potential_{EB}, we performed feature selection using the Lasso regression algorithm

(**Supplementary Fig. 5a-c**). We found that 12 EC, 8 ME, 9 EN, and 5 PL markers were sufficient to calculate a near perfect fit to differentiation potential_{EB} with only a slight decrease in statistical power (**Supplementary Fig. 5d**).

Hierarchical clustering of day 0 and day 12 EB differentiation expression showed a clear separation between the two time points (**Fig. 2e**). It also showed that a subset of lines had specific gene expression programs. Cell lines iPS11C, HUES53, and HUES65 had significantly lower expression of several key EC genes (**Fig. 2e**, top right box) and also the lowest EC differentiation potential_{EB} at day 12 (**Fig. 2c,d**). To assess the stability of the scores and study the effect of continued cell culture on differentiation, we selected three cell lines and passaged them continuously for >10 passages. We found that expression signatures remained highly consistent for each cell line, with stronger sample similarity within than between groups (**Fig. 2f**). We also found that the differentiation potential_{EB} remains largely similar, with a slight increase of the mesoderm signal for later passages of H1 and H9 differentiation (**Fig. 2g**). To compare the predictive power of cell line differentiation potential as quantified by both the teratoma and qPCR assays, we calculated the variance in assay scores between cell lines (between-groups) and between replicates of the same cell line (within-group). The between and within-group variances for teratoma germ layer scores quantified using our pathology analysis or using teratoma RNA expression measured via our signature panel were very similar (**Fig. 2h** left; $P > .2$, F -test). In contrast, EB differentiation potential variance between replicates was significantly less than variance between cell lines, even when considering replicates >10 passages apart (**Fig. 2h** right; $P < .0005$, F -test). We also found that EB formation using the same protocol but with hPSC line stocks from different laboratories (Meissner and Melton) can vary in their EB differentiation potential, such that expression signatures are more similar between different lines within the same laboratory than between the same cell lines from different laboratories (**Fig. 2i**). Taken together, our results show that our qPCR-based measure of cell lines' germ layer propensity is more quantitative and reproducible than the teratoma assay, even when the cell line has remained in culture for several passages; however, as expected different cell line stocks in different labs can have very different starting propensities.

Temporal signature dynamics in EB differentiation

Along with variation in differentiation potential_{EB} between cell lines, we also observed differences in induction dynamics during EB formation. For example, H9 activated EC genes and repressed PL markers by day 5, whereas H1 achieved similar levels of EC induction and PL repression more gradually (**Fig. 3a**).

To further characterize these induction dynamics, we calculated the ratio of day 5 to day 12 mean gene expression (**Fig. 3b**), averaged across all markers within each germ layer. Ratios near or above 1 indicate rapid induction, and ratios near or below 0 indicate delayed germ layer response. This analysis confirmed that H9 activated EC markers faster than H1 (**Fig. 3b**), and that H9 repressed PL markers more rapidly than H1 and HUES53 (**Supplementary Fig. 6a**). The EB gene expression dynamics correlated very well ($R > 0.95$) with differentiation potential_{EB} dynamics for the three germ layers (**Supplementary Fig. 6b,c**), calculated using the ratio of $Z_{w,GS}$ day 5 to day 12. We also note that fast induction did not

necessarily coincide with high differentiation potential_{EB}, as iPSC11c had a rapid EN induction at day 5 but the second lowest EN differentiation potential_{EB}.

To visualize these EB differentiation dynamics in orthogonal vector space, we performed principal component analysis (PCA) of our day 0, 5, and 12 data (**Fig. 3c**). We found that the first principal component (*x*-axis) followed the path of differentiation, as cell lines increased their component 1 scores from day 0 to day 12 (**Fig. 3c,d**). Furthermore, we mapped the direction of EN, ME, and EC differentiation by weighted averaging of the loadings for the marker genes of the respective germ layer (**Fig. 3d**). Amongst the six selected cell lines, HUES53 advanced the smallest distance along the component 1 axis in the direction of the ME vector, indicative of it being ineffective in shutting PL genes down and activating EC and EN markers. Moreover, H9 and HUES64 had a strong EC propensity at day 0, which H9 persisted through day 12, whereas HUES64 moved in the ME/EN direction, activating those markers more effectively.

To quantify EB dynamics for each gene, we calculated the RNA expression ratios of day 2 to day 12 and day 5 to day 12, averaged across the 23 reference cell lines. Positive ratios represent gene induction, whereas negative scores represent overall repression in EB differentiation (**Fig. 3e**). We found that several genes were induced transiently at day 2 and at both day 2 and day 5, while others were transiently repressed at day 2 and day 5. Comparison of day 12 and day 5 weights to day 2 can also be used to visualize gene dynamics on a scatter plot (**Supplementary Fig. 6d,e**). In summary, our EB dynamics analysis revealed rapid activation of EC by day 5 of EB formation, more gradual activation of ME that reached the highest differential potential_{EB} at day 12, and a delayed EN response.

Assessing directed differentiation of hPSCs

Next, we studied directed differentiation of 14 hPSCs to the three embryonic germ layers in two-dimensional (2D) culture^{9, 13}. Hierarchical clustering of the data (**Fig. 4a**, columns) and PCA along the first two components (**Fig. 4b**, top) both showed clear separation between datasets for the germ layers. Furthermore, clustering of the genes (**Fig. 4a**, rows) aggregated markers from the same germ layers across all experiments. As expected, some pluripotent markers intermixed with markers from EN but not the other two germ layers. Similarly, PCA showed that undifferentiated populations mapped most closely to directed dEN and dEC data whereas dME experiments separated the most along the direction of the first component (**Fig. 4b**), which is consistent with prior RNA-seq analysis⁹. Mapping the directions of the marker genes (**Fig. 4b**, bottom) showed that H1 and H9 differentiated most effectively into dEC whereas iPSC1-51C moved along the direction of dME the least.

To quantify the 2D differentiation potential, we made several small modifications to our EB differentiation potential calculation (**Supplementary Fig. 7a**). First, we changed the reference set to the undifferentiated, control populations of the 2D experiments for 11 established hPSC lines (top 11 lines in **Fig. 4c**). Also, we assigned the 2D weights based on cell type uniqueness of expression for each marker (**Online Methods**). The 2D weighted Z-method standard normal deviate for each gene set represents the differentiation potential_{2D}.

Each of the lines we profiled had a highly positive differentiation potential_{2D} at day 5 (**Fig. 4c**, left) and a significant combined P value ($P < .002$, **Fig. 4c**, right), which was not the case when using the unweighted Z -transform test (**Supplementary Fig. 7b**). Moreover, the differentiation potential_{2D} for gene sets outside the pertinent 2D germ layer could quantify the amount of background expression of the other cell types (**Supplementary Fig. 7c,d**). On average, dEN differentiation had a higher background of ME and PL gene expression, which decreased following fluorescent activated cell sorting (FACS; **Supplementary Fig. 7e**)¹⁴.

The differentiation potential_{2D} correlated well with established measures of differentiation efficiency, including abundance of the cell surface markers CD56 (dME and dEC) and CD184 (dEN)⁹. We quantified the percent CD56 positive cells for 11 cell lines and found significant correlation ($R = 0.83$, $P < 10^{-3}$, Pearson correlation) with the pertinent ME and EC differentiation potential_{2D} of those lines, but no significant ($R = 0.35$, $P > 0.1$, Pearson correlation) positive correlation with markers for the other germ layers (**Fig. 4d**). This was also the case for CD184 positive cells and EN differentiation potential_{2D}. The lack of correlation in efficiency with the differentiation potential_{2D} of alternate lineages again argues that there is limited crosstalk in our germ layer marker expression. We also tested whether the three PiPS cell lines that did not make EBs could differentiate into ectoderm in a directed manner. The negative EC differentiation potential_{2D} at day 5 (**Fig. 4e**) indicates that they could not form EC. Again, we detected a high ME background signal characteristic of the fibroblast origin of these cells.

Using the differentiated state as a reference, computing differentiation potential_{2D} either by using the unweighted Z -transform test or by averaging the t scores⁷ lowers the overall correlation with 2D efficiency (**Supplementary Fig. 8a-c**). It is also worth noting that in general we did not find expression at day 0 to be very predictive of the 2D or EB differentiation potentials (**Supplementary Fig. 8d**). Taken together, these results argue that the improved gene selection, using undifferentiated cells as a reference, and weighting of gene observations adds accuracy and flexibility to our computational analysis of the qPCR results.

Utility of assay to quantify the effects of perturbations

The high proportionality between differentiation potential_{2D} and differentiation efficiency provides a quantitative tool for applications beyond line-to-line comparison.

First, we evaluated the effects of different small molecules on endoderm differentiation. We found that dEN protocols that substitute LiCl for WNT3A have a similar differentiation potential (**Fig. 4f**). However, replacing Activin A with the previously reported¹⁵ compound IDE1 decreases LEFTY1 activation and the overall EN differentiation potential.

Next, we used shRNAs to knockdown transcriptional regulators of early differentiation, including EOMES¹⁶, GATA4¹³ and OTX2^{13, 17}, in HUES64 cells. We observed reduced expression of several key lineage markers and a significant overall reduction in the differentiation potential_{2D} of cell lines after knockdown of EOMES ($P = 1 \times 10^{-5}$, Fisher's method) and GATA4 ($P = .003$, Fisher's method; **Fig. 4g**), which is independent of the cell line used (**Fig. 4h**). The knockdowns showed both significant reduction of EN marker

expression and increase of ME markers ($P = .01$, Fisher's method), in line with EOMES's known role in suppressing the mesoderm transcriptional program during endoderm differentiation¹⁶. Similarly, OTX2 knockdown showed lower expression of several key ectoderm markers and reduced EC differentiation potential_{2D} ($P = 6 \times 10^{-5}$, Fisher's method; **Fig. 4i**) as well as higher expression of ME markers ($P = .002$, Fisher's method) during dEC differentiation, suggesting that OTX2 might also play a role in suppressing the mesoderm transcriptional program.

Finally, we wanted to quantify the effects of JQ1 (BRD4 inhibitor) given its previously reported role at “super enhancers”¹⁸. Treatment of the HUES64 cell line with different concentrations of JQ1 during 12 days of EB formation led to increased EN expression and differentiation potential_{EB} ($P < .01$, weighted Z-method) (**Fig. 4j**) along with decreased EC and ME differentiation potential_{EB} (**Supplementary Fig. 8e**). This fits well with recent data showing that super enhancers are enriched for key regulators in EC and ME but not in EN¹³, which may explain the endoderm bias of JQ1-treated EBs.

Quantitative assessment of distinct culture conditions

Our EB and 2D experiments for multiple matching cell lines allowed us to compare these two established differentiation protocols. Scatter plot of the 2D weights versus EB day 5 (**Fig. 5a**) or day 12 weights (**Supplementary Fig. 9a**) shows that the average differential expression of EC and ME gene markers are mostly correlated ($R = 0.48, 0.73$ for EC, ME), with some exception for EC (EN1, NR2F2) and ME (BMP10, NKX2-5, GATA4). In contrast, weights for EN markers are anti-correlated ($R = -0.84$), where several established markers of directed differentiation (FOXA2, EOMES, NODAL, LEFTY1,2) decrease expression in EBs, which may relate to the different signaling components used. For example, directed differentiation into dME uses BMP signaling and GATA4 is known to act downstream of BMP signaling in heart development¹⁹, which could explain differences in BMP10 and GATA4 expression. Moreover, Activin A signaling is crucial for dEN differentiation and proteins NODAL, LEFTY1,2, and EOMES are known downstream effectors¹⁶.

In comparing EB versus 2D undifferentiated samples, we noticed a substantial difference between hPSC lines grown on mouse embryonic fibroblast (MEF) feeder cells and those grown in feeder-free conditions. For 11 lines grown using both feeder and feeder-free culture, we computed the mean difference in expression per gene (**Fig. 5b**) and found that most significantly downregulated genes ($P < .05$, Wilcoxon signed rank test) are EN and ME markers and most upregulated genes are pluripotency and mesendoderm markers. Moreover, GATA4, EOMES, NODAL, LEFTY1,2 are found to decrease in expression in feeder-free culture, which may also contribute to their increased levels in 2D versus EB differentiation. Displaying the distribution of mean expression differences for all markers within the four gene classes showed that pluripotent genes are more highly expressed ($P = 7 \times 10^{-4}$, weighted Z-method) in feeder-free culture (**Fig. 5c**). In contrast, EN genes ($P = 2 \times 10^{-6}$) and to a lesser degree ME and EC genes ($P < .05$, weighted Z-method) are more highly expressed in cells on feeders, suggesting that culture on MEFs introduces higher background

differentiation (**Fig. 5c**). We observed a similar trend when considering the distribution of differentiation potential_{2D} difference (**Supplementary Fig. 9b**).

Next we used our signature panels to quantify the effect of cell line adaptation due to repeated passaging in feeder-free culture. We compared the gene expression profiles of seven hESC lines cultured on Geltrex and after adaptation for one or more passages on Geltrex. The majority of EN and ME genes decreased in mean expression following adaptation, and the effect became more pronounced after six or more passages of adaptation (**Fig. 5d**). In contrast, pluripotent and EC marker expression did not change substantially after Geltrex adaptation.

To quantify the effect of hESC adaptation on subsequent differentiation, we calculated the mean expression difference during directed differentiation for each gene class and found that adapted lines were less efficient but more specific. For example, during dEN differentiation (**Fig. 5e**, right panel) expression of EN markers was slightly lower in adapted cells, whereas background expression of ME and EC markers was significantly lower ($P < .01$, weighted Z-method). Lower background expression of alternate lineages was also observed for dME (**Fig. 5e**, middle panel) and dEC differentiation (left panel). Furthermore, we calculated the difference in differentiation potential_{2D} between adapted and unadapted cells for the seven cell lines (**Fig. 5f**). Again, during dEN differentiation EN potential_{2D} was lower on average, but ME and EC potential_{2D} were much lower. Similarly, dME and dEC differentiation were less efficient on average but more specific. Genes that were significantly down regulated ($P < .05$, *t*-test) due to adaptation (**Fig. 5g**) include EN and ME markers that were also down regulated in adapted cells at day 0 as well as EC markers whose expression was not affected by adaptation prior to differentiation.

DISCUSSION

We describe an improved ScoreCard assay based on qPCR measurements of 96 genes, which is highly scalable, broadly accessible, and 5-10-fold faster than the teratoma assay (**Fig. 6a**). We further show that it is more reproducible than the teratoma assay, in part because of the inherent noise in quantifying sectioned stainings (which arises from limited sampling in two dimensions, unfocused scanning near folded tissue, variability of fixation, variability of antibody concentrations, the presence of host tissue, and the difficulty of classifying undifferentiated and primitive tissue types). We also introduce a computational approach for combining gene signatures in a meaningful and assay-dependent manner using the weighted Z-method, which presents several advantages over PluriTest⁶ and the previous ScoreCard⁷ (**Supplementary Fig. 9c**). The combined statistic allows us to detect the significance of a cell line's ability to differentiate along each of the three germ layers. PluriTest separates pluripotent and differentiated samples along two dimensions—pluripotency and novelty score. As currently designed, it does not give an individual score for the three germ layers.

We provide an in-depth characterization of EB formation, directed differentiation and common hPSC tissue culture conditions spanning 272 experiments (**Supplementary Table 5**). By improving our gene selection (**Supplementary Fig. 3b**) and weighting gene

observations, we increased the statistical power (**Supplementary Fig. 4e, 7c**), accuracy (**Supplementary Fig. 8a,b**), and flexibility of the panel towards use in numerous applications. These include small-molecule screens for dEN and dME protocols, characterizing feeder versus feeder-free culture and their effects on directed differentiation, and quantifying the effect of genetic perturbation of key lineage regulators during directed differentiation.

Our computational measure of differentiation potential_{2D} correlated well with established metrics for differentiation efficiency. Moreover, we performed feature selection using the Lasso regression algorithm to calculate the minimum number of markers needed to accurately fit differentiation potential_{2D} and 2D efficiency (**Supplementary Fig. 10**). We found that 6-8 markers per germ layer are sufficient to calculate a near perfect fit to differentiation potential_{2D} (**Supplementary Fig. 10c**) and 2-3 markers for predicting directed differentiation efficiency (**Supplementary Fig. 10f**). We also observed high correlation with differential potential_{2D} and differentiation efficiency when using an even smaller subset of 2D expression signatures (**Fig. 6b**) down to individual genes (**Fig. 6c**). Overall, we find that our gene panel could be further reduced (**Supplementary Table 4**), while sacrificing little in accuracy and statistical power. This reduction analysis will help guide users to also develop alternative cell line tests using selected markers combined with flow cytometry or possibly reduced RT-qPCR panels.

A major focus of hPSC research is the generation of differentiated cell types relevant to disease therapy. The ScoreCard provides some information in this direction, but was specifically designed for the rapid assessment of molecular and functional pluripotency. One could envision new versions that would include unique gene expression signatures at other developmental timepoints in order to predict capacity to differentiate into specific cell fates of interest.

ONLINE METHODS

Recommended protocol for ScoreCard assay

We have provided a detailed description of the current ScoreCard design and various experimental details including many useful applications. Interested users have several options to take advantage of our results and analysis:

- 1) Life Technologies (now a Thermo Fisher company) offers a commercial version (TaqMan hPSC Scorecard) that 100% matches the described ScoreCard panel used for all 272 experiments in this study. The company website provides a detailed 5-Step protocol including recommended reagents for the day 5 EB protocol, details on assay design and analysis tools. - > To establish a common and comparable standard across laboratories, this appears to be the preferred approach for individual users.
- 2) We provide many details on the various EB time points, the 2D differentiation and the downstream effects on the results and analysis. Users can obtain the TaqMan plates based on our Supplemental Information and use a more customized assay design. As described, informative results can be obtained

within 5 days, but extended time for EB formation (day 12) will increase confidence (we provide culture details in the Tissue culture and Embryoid body (EB) differentiation sections). We recommend using 1ug of total RNA per sample (see RNA extraction, reverse transcription and Taqman qPCR section for more detail). For analysis, we advise to use normalization based on the average between 5 control genes (Data normalization) and using a pluripotent reference of 23 hPSC lines at day 0 (Supplementary Table 5). Calculation of differential potential should be carried out as described here (Computing differentiation potential) using weights calculated for day 12 EBs (Supplementary Table 4).

This option will provide valuable results but requires additional skills compared with option 1, and the results are not directly comparable to those of other laboratories using a different strategy.

- 3) We provide an extensive description of the gene sets and the regression analysis to define the minimal number of informative genes. Users could choose to design their own TaqMan plates using the reduced number of genes and then follow most of the step described in option 2 and the provided **Supplementary Code**.

This option will also provide valuable results but requires more advanced computational skills, and the results are also not as directly comparable to those of other laboratories using a different strategy.

Immunofluorescence and Phase Images

hPSCs were passaged on Geltrex (Life Technologies) coated plates and cells were cultured in mTeSR1 medium (Stem Cell Technologies). Cells were then fixed in 4% paraformaldehyde (PFA) for 15 minutes at room temperature. Permeabilizing and blocking was performed in 5% bovine serum albumin (BSA) and 1% Triton X-100 (Sigma) in 1X phosphate-buffered saline (PBS; Life Technologies) for 30 minutes. Cells were stained with mouse anti-oct3/4 primary antibody (1:1000 dilution; BD Transduction Laboratories, 611202) at 4°C overnight. The secondary antibody (goat anti-mouse DyLight 488; Jackson ImmunoResearch, 115-485-003; 1:1000 dilution) was applied for 2 hours at room temperature. TRA1-60 double staining was performed using StainAlive mouse anti-Human antibody (1:200 dilution; Stemgent, 090068) at room temperature for 2 hours. The nuclei were stained with Hoechst 33342 (1µg/ml; Life Technologies, 33342). Cells were imaged using Olympus IX71 microscope and MetaMorph Advanced software.

RNA extraction and RNA-seq

For measuring expression level, RNA was isolated from hESCs using TRIzol (Invitrogen, 15596-026), further purified with RNeasy columns (QIAGEN, 74104). RNA-seq library construction and data analysis was carried out as described previously⁹.

Tissue culture

Cell culture was done as reported previously⁹. Briefly, hPSCs were maintained on irradiated Murine Embryonic Fibroblasts (MEFs, plated at density of 15,000cells/cm²; Global Stem).

Cells were cultured in KO DMEM (Life Technologies) medium supplemented with 20% Knockout Serum Replacement (KOSR; Life Technologies), 1% GlutaMAX (Life Technologies), 1X MEM Non-Essential Amino Acid (Life Technologies), 10ng/ml bFGF (Millipore), and 55 μ M 2-mercaptoethanol. Cells were passaged every 4-5 days using 1mg/ml Collagenase IV (Life Technologies).

Teratoma formation

Teratoma formation was performed by the Genome Modification Facilities. hPSC lines HUES9, HUES53, HUES64, H1, H9, and iPSC1-51C were expanded in culture and each line was injected into the kidney capsules of 3 immuno-suppressed mice, using 1 million cells per animal. The teratomas were isolated 7-8 weeks following injection. Next, the teratomas were fixed in 4% paraformaldehyde (PFA) pH 7.4 for 24 hours, embedded in paraffin, sectioned at a thickness of 10-12 μ m, and stained with hematoxylin and eosin for examination. The Histology Core Facility at Harvard Stem Cell Institute performed the steps following teratoma formation.

Teratoma analysis

Histological analysis and quantification of the different germinal layers in the teratomas was performed using the Leica automated scanner (SCN400 F) and the Leica Biosystems Tissue Image Analysis software (Version 4.0.4). A board-certified pathologist examined the tissue, and annotated all areas of mesoderm, ectoderm, and endoderm on all of the tissue present on the scanned slide. Annotation was done blindly, without prior knowledge of the identity of the cell lines and samples. Specialized tissues such as muscle, nerve, cartilage, neuroectoderm, and bone were given a separate annotation (**Supplementary Table 2**) before being summed into the three germ layers for final analysis. The amount of tissue present in each germinal layer was quantified by measuring the number of cell nuclei within each annotation, using the “Measure Stained Cell” algorithm. This algorithm was custom-designed to measure hematoxylin-stained nuclei (tissue threshold = 220, nuclei heterogeneity = 2, strength of nuclear counterstaining = 2, nuclear window radius size = 50, maximum cell radius = 100, % of stained area in a nucleus cutoff = 10%, strong/moderate/weak nuclear staining intensity cutoff = 173-203, nuclear staining intensity cutoff = 220). The nuclei count for all annotation was automated, using the same algorithm. For the quantification of total cell nuclei within each annotation, moderately- and strongly-stained nuclei were added together; weakly-stained nuclei were not used in the analysis as they often did not detect a true nuclei (**Fig. 1c**).

Immunohistochemistry and antibody staining

We picked 3 antibodies for each germ layer (total of 9 antibodies) that were uniquely expressed in our directed differentiation data, reasoning that since they are unique markers of early development they should be a good marker in teratomas which often contain tissues that are not fully differentiated and matured in structure. For each antibody, we performed IHC at different dilutions and picked one antibody per germ layer with the best specificity in staining, independent of our pathological analysis. FOXA2, HAND2, and PAX6 were hence chosen in this manner.

Immunohistochemical stainings for FOXA2 (R&D Systems AF2400, 1:500), HAND2 (Abcam ab60037, 1:100), and PAX6 (Abcam ab5790, 1:800) were carried out by the Histology Core Facility at Harvard Stem Cell Institute. In summary, PFA-fixed and unfixed teratoma cryosections were mounted on slides. Unfixed sections were thawed for 10 minutes at room temperature and then fixed in cold acetone for 2 hours. Antigen retrieval for both acetone-fixed and formalin-fixed sections was achieved by boiling in citrate buffer (10mM citric acid, 0.05%, pH 6) for 15 minutes. Endogenous peroxidase activity was quenched with 3% hydrogen peroxide for 20 minutes. Sections were blocked in 3% normal horse serum for 20 minutes and then incubated overnight at 4°C with either FOXA2, HAND2, or PAX6 primary antibodies at dilutions indicated above. Secondary antibody incubation was for 30 minutes at room temperature with either anti-goat (Vector Labs PK-4005, 1:200, for FOXA2) or anti-rabbit (Vector Labs BA-1000, 1:200, for HAND2 and PAX6). The ABC amplification system (Vector Labs, PK-4005) and DAB substrate detection (Vector Labs, sk-4100) were used following the provided kit protocols. Standard washes were used for all procedures. The slides were counterstained in hematoxylin for 2 minutes, dehydrated in series of alcohol and water washes, and coverslipped.

Taqman qPCR signature panel design

We selected gene markers for each of the four gene categories (EC, ME, EN, PL) based on uniqueness of expression for that marker in published RNA-seq for directed differentiation of hESC line HUES64 into purified germ layer populations⁹. Uniqueness of expression for each gene was defined as the difference in expression for each gene in a given category and the closest expression level of that gene in the other three cell types. For example, minimal uniqueness of expression of EC marker k with $\log_2(FPKM)$ expression value dEC_k in a purified ectoderm population dEC (see **Supplementary Fig. 3b**) can be calculated as follows:

$$\min \Delta \log_2 (FPKM) = dEC_k - \max (dME_k, dEN_k, dPSC_k).$$

The expression values dME_k , dEN_k , $dPSC_k$ represent $\log_2(FPKM)$ expression values in purified mesoderm, endoderm, and pluripotent populations, respectively. Similarly the mean uniqueness of expression of the same ectoderm marker k can be calculated as follows:

$$\text{mean } \Delta \log_2 (FPKM) = dEC_k - \text{mean} (dME_k, dEN_k, dPSC_k).$$

For the final panel design, genes were ranked based on having the highest minimal uniqueness of expression and selected with priority given to highly expressed genes, to genes also highly uniquely expressed in Nanostring data⁹ from the same cell types, and to known markers from the literature. Mesendoderm (MS) markers were selected using recent RNA-seq data from an earlier HUES64 directed differentiation mesendoderm population¹³ in the same manner, but now comparing uniqueness of expression to hPSC and the three germ layers (4 cell types). Genes known to be invariant in expression level were selected as control genes and further filtered to use for data normalization (see below). Identified panel was used to design the set of 96 genes in replicates of 4 to generate a 384 well panel which

is now commercially available as TaqMan hPSC Scorecard panel (Life Technologies). Subsequent experiments were carried out using the commercially available product.

Correlation analysis

The Pearson correlation coefficient (R) between biological and technical replicates was computed in a pairwise manner using all genes with non-empty CT expression values for both samples. The distributions of CT values are Gaussian. Analysis was performed using MATLAB version 8.1 release name R2013a, built-in function `corrcoef.m` with default settings and 'rows' parameter set to 'pairwise'.

Data normalization

We normalized the data using housekeeping genes that we found most invariant in expression between experiments and most unbiased in differentiation into the 3 germ layers. To quantify each gene's expression variability, we measured all possible differences in expression between four cell types (dEC, dME, dEN, hPSC), averaged across 21 independent 2D experiments (**Supplementary Table 3**). We found the following 5 gene probes to be the most invariant in expression on average during germ layer differentiation-- SMAD1, EP300, CTCF, ACTB, and ACTB. Moreover, we calculated the average bias between different germ layers for these 5 genes (**Supplementary Table 3**). In all cell type comparisons, the average bias for a given comparison was half of a CT value or less. As a result, we decided to normalize the data using the average expression level of the 5 control genes. This normalization factor was subtracted from the CT value of each gene per experiment.

Embryoid body (EB) differentiation

Embryoid body formation was done as previously described⁷ with slight modifications. Briefly, hPSCs were expanded in 10cm² dishes on irradiated MEFs. Once the cells were confluent, we harvested them using 1mg/ml Collagenase IV and washed three times with PBS. We collected a portion of the cells as a day 0 undifferentiated control, and the remainder were plated for EB formation on 6-well low attachment dishes (Corning) in KO DMEM medium with 20% KOSR, 1% GlutaMAX, and 1X MEM Non-Essential Amino Acid. The media was changed every 2-3 days. EBs were collected at days 2, 5, and 12, washed three times with PBS, flash frozen as pellets, and stored at -80°C.

Derivation of partially reprogrammed cell lines

Adult fibroblasts (samples from patients with Spinal Muscular Atrophy (SMA) procured under IRB-approved protocols) were plated at a density of 100,000 cells into one well of a 6-well plate. The cells were transduced for two consecutive days with 4 retroviruses (1ml of MIG-Oct4, 1ml of MIG-Klf4, 1ml of MIG-Sox2 and 200µl of MIG c-Myc). Five days later, the transduced fibroblasts were collected and replated on feeder (irradiated mouse embryonic fibroblasts). Beginning the following day, the cells were fed daily with hESC medium (400ml DMEM/F12; 100ml KO-SR; 2x L-Glutamine, 1X MEM-NEAA, 10ng/ml bFGF) until colonies were ready to pick. Fully reprogrammed iPSC lines were identified by viral GFP-silencing and characteristic pluripotent stem cell colony morphology. Partially

reprogrammed iPS cell lines (or PiPS) were picked based on uncharacteristic shape and GFP expression.

RNA extraction, reverse transcription and Taqman qPCR

Total RNA was extracted from frozen cell pellets using Ambion Pure Link RNA Mini Kit (Life Technologies). Eight cDNA reactions were set-up from 1µg of total RNA per sample using High-Capacity cDNA RT Kit (Life Technologies). qPCR was performed on 384-well TaqMan hPSC Scorecard plates using Viia7 RUO software and Applied Biosystems ViiA7 instrument. The qPCR assay used in this study is commercially available as the TaqMan hPSC Scorecard panel (Life Technologies).

Computing differentiation potential

To quantify the differentiation potential of cell lines using EB differentiation data, we built on our previous computational analysis³. Specifically, for each gene in an EB experiment, we compared its expression relative to the expression of that gene prior to EB differentiation (day 0) for 23 hPSC reference lines and calculated a P value P_k using the one-tailed, one sample t -test. The distribution of CT expression values for the reference set closely resembled a Gaussian distribution, as has been observed for gene expression data in general, which justifies the use of the t -test. The resulting P values are then combined within each gene set GS (PL, EN, ME, or EC) using the weighted Z -method

$$P_{w,GS} = 1 - \Phi \left(\frac{\sum_{k=1}^N w_k Z_k}{\sqrt{\sum_{k=1}^N w_k^2 + 2 \sum_{k < j} w_k w_j r_{k,j}}} \right),$$

where $Z_k = \Phi^{-1}(1 - P_k)$, P_k is the P value for the k -th gene in each gene set GS that contains N genes, Φ and Φ^{-1} denote the standard normal cumulative distribution function and its inverse, w_k are the corresponding weights for the k -th gene, and $r_{k,j}$ is the correlation of Z_k and Z_j . Weights were set to the CT expression difference between day 5, 12 of EB formation and day 0, averaging across experiments for 23 cell lines (**Supplementary Table 4**). This can be interpreted as the average differentiation power of each gene during EB differentiation. Genes with negative weights were excluded from the EB analysis ($w_k = 0$). The correlations $r_{k,j}$ were estimated using Z_k from all 23 day 12 EB reference cell lines. The combined P value for each gene set GS , $P_{w,GS}$, gives a statistical basis for rejecting the null hypothesis that a cell line cannot differentially express gene markers characteristic for a given germ layer. We refer to the weighted Z -test standard normal deviate $Z_{w,GS} = \Phi^{-1}(1 - P_{w,GS})$ for each gene set GS as the differentiation potential_{EB}, since it measures the distance traveled away from a pluripotent state. The weighted Z -test assumes one-tailed P values. To satisfy this assumption, we use the one-tailed version of the one sample t -test. Partially reprogrammed cell lines had a number of missing values for differentiation markers due to low expression of these genes. Hence, for these samples we assigned a CT value of 45, which estimates the highest measureable CT value, for all markers with missing values.

To calculate the differentiation potential_{2D}, we made several small modifications to our computational approach. First, we changed the reference set to the pluripotent populations

cultured in feeder-free conditions for 11 established hPSC lines (top 11 lines in **Fig. 4c**). Also, we assigned the 2D weights based on minimal uniqueness of expression for each marker amongst the four cell types (see Taqman qPCR signature panel design), averaged across all cell lines. The correlation terms $r_{k,j}$ for all EC markers were estimated using Z_k from all 14 dEC experiments (cell lines in **Fig. 4c**). Similarly, the ME, EN, and PL marker correlation terms $r_{k,j}$ were estimated using all 14 dME, dEN, and hPSC experiments, respectively. Analysis was performed using MATLAB version 8.1 release name R2013a, using functions provided in the **Supplementary Code**.

Measuring within and between group variance

To calculate between-group variance for teratoma pathology analysis, we computed the variance of the normalized pathology scores for the 6 cell lines displayed in **Figure 1d**, where scores from biological replicates for HUES64 and H9 were averaged prior to the calculation. We computed the scores for EC and ME individually but excluded the EN score due to the low sampling of these tissues in the teratoma sections and the resulting high level of variability and noise. The between group variance for the EB analysis was calculated for EC, ME, and EN day 12 differentiation potential_{EB} for the 23 reference cell lines. The EC, ME, and EN differentiation potential_{EB} scores were divided by the sum of the differentiation potential_{EB} scores of all three germ layers for each cell line and multiplied by 100, to normalize the differentiation potential_{EB} to 100% and make the scoring equivalent with that of the teratoma quantification (% nuclei).

To calculate the within-group variance for teratoma and EBs, we calculated the variance between different replicates within each cell line and then performed weighted averaging of the within-group variances of all cell lines with replicates. The weights were proportional to the degrees of freedom, or the sample size minus one. For the teratoma sections, the replicates used were the teratoma sections of each cell line. For teratoma section RNA, we calculated the differentiation potential for each section and normalized the scores to 100% as described above for the EB analysis. For EBs, we calculated the within group variances separately for biological replicates with similar passages (less than 4 passages apart) and replicates that are more than 10 passages apart. The significance of the ratio of between versus within group variance was calculated using the F-test, using the total number of degrees of freedom for each between and within group variance calculation. Analysis was performed using MATLAB version 8.1 release name R2013a, using built-in function `vartest2.m` modified to calculate P value between two samples based on variance and degrees of freedom of each sample.

Hierarchical clustering and PCA

Hierarchical clustering was performed across all day 0 and day 12 EB data and all genes in the 4 gene categories. We used Pearson correlation coefficient as a measure of similarity and average linkage to join clusters. Results using Euclidian distance metric were highly similar, again separating all day 0 and day 12 samples. For clustering all dEC, dME, and dEN directed differentiation experiments and all genes in the 4 gene categories, we used the same clustering procedure; again clustering by Euclidean distance was highly similar and also separated 2D differentiation experiments by germ layer identity. Analysis was performed

using MATLAB version 8.1 release name R2013a, using built-in function clustergram.m with the following parameter settings--'ImputeFun', @knnimpute, 'RowPdist', 'Correlation', 'ColumnPdist', 'Correlation', 'linkage', 'average', 'cluster', 3, 'OptimalLeafOrder', false.

Principal component analysis (PCA) was carried out for all EB experiments across all time points and across all genes (observations) from the 4 gene categories. To visualize the EB dynamics in 2 dimensions, we mapped the scores of the first and second principal component for each cell line on a scatter plot. We also mapped the direction of PL, EN, ME, and EC gene sets in two-dimensional space by creating an expression vector that only expresses genes in the relevant gene set and multiplying it by the PCA loadings to obtain the gene set specific PCA scores and direction. We weighted this expression vector using our day 12 weights_{EB} for each gene set to be consistent with our measure of differentiation potential_{EB}. To aid visualization of the vectors, the weighted averages of the component 1 and 2 loadings in **Fig. 3b** and **Fig. 4f** were scaled by a factor of 30. The same approach was used for performing PCA for all the 2D data and across all genes (observations) from the 4 gene categories, where the 2D weights were now used for mapping the directions of gene sets PL, EN, ME, and EC. Analysis was performed using MATLAB version 8.1 release name R2013a, using built-in function knnimpute.m and pca.m with default settings.

EB dynamics analysis

To quantify each cell line's EB induction dynamics, we calculated the ratio of day 5 to day 12 mean gene expression within each germ layer, averaging across all markers. Ratios near or above 1 indicate rapid induction, while near or below 0 indicate delayed germ layer response. For the pluripotent gene set, we similarly calculated each cell line's EB repression dynamics using the ratio of day 5 to day 12 mean gene expression, where a positive numbers now indicate rapid repression. To quantify each gene's EB induction/repression dynamics, we calculated the RNA expression ratios of day 2 to day 12 and day 5 to day 12, averaged across the 23 reference cell lines before calculating the ratio. Positive ratios represent gene induction, while negative scores represent overall repression in EB differentiation. We note that since EBs are inherently heterogeneous and we are measuring population average gene expression, induction/repression refers to the average change in expression of a gene across the population, and not actual up/down regulation in any individual cell.

Directed differentiation into the three germ layers

When hPSCs reached 60-70% confluency on MEFs, the cells were plated as clumps at a low density on 6-well plates coated with Geltrex in mTeSR1 medium. We maintained the cells for three days in feeder-free culture and then induced directed differentiation towards endoderm, mesoderm and ectoderm. To induce endoderm differentiation, cells were cultured for 5 days in RPMI medium supplemented with 100 ng/ml Activin A (Life Technologies), 2 μ M/ml Lithium Chloride (Sigma), 0.5% FBS, 1% GlutaMax, 1X MEM Non-Essential Amino Acid, and 55 μ M 2-mercaptoethanol. For the endoderm protocol test in **Figure 4f**, 50 nM WNT3A (R&D Systems) was used in place of LiCl and 0.5 μ M IDE1 (Stemgent) instead of Activin A. To induce mesoderm differentiation, for the first 24 hours cells were cultured in DMEM/F12 medium supplemented with 100 ng/ml Activin A (Life

Technologies), 10 ng/ml bFGF (Millipore), 100 ng/ml BMP4 (Life Technologies), 100 ng/ml VEGF 100 ng/ml (Life Technologies), 0.5% FBS, 1X MEM Non-Essential Amino Acid, 55 μ M 2-mercaptoethanol, and GlutaMAX (Life Technologies). For days 1 through 5 of mesoderm differentiation, Activin A was removed from culture. For the mesoderm protocol test in **Figure 4f**, recombinant proteins Activin A, BMP4, and VEGF were purchased from R&D Systems. To induce ectoderm differentiation, cells were cultured for 5 days in DMEM/F12 differentiation media supplemented with 2 μ M TGFb inhibitor (Tocris, A83-01), 2 μ M WNT3A inhibitor (Tocris, PNU-74654), 2 μ M Dorsomorphin BMP inhibitor (Tocris), 55 μ M 2-mercaptoethanol, 1X MEM Non-Essential Amino (Life Technologies), and 15% KOSR (Life Technologies). Media was changed daily. Directed differentiation and pluripotent control samples were cultured simultaneously and collected for RNA extraction at the end of the differentiation. For each 2D protocol, plating densities were kept consistent across all cell lines.

FACS analysis of hESCs

hESCs and cells differentiated into mesoderm and ectoderm were stained with 1 μ l UV LIVE/DEAD (Molecular Probes, Cat Nr. L23105) to assess viability. Cells were then fixed with 4% PFA for 15 minutes at RT and labeled with antibodies directed at CD56 (BD 340724) and CD326 (BD 347199). Endoderm differentiation was assessed using CD184 (BD 555974) and CD326 (BD 347200) cell surface markers, on live cells. Forward scatter (FSC) and side scatter (SSC) gates were used to select single cells, excluding debris and larger aggregates. Immunofluorescence was measured on a BD LSRII instrument (BD Biosciences, CA) and data was analyzed using FloJo software (v9.4.11; Tree Star, CA).

shRNA infection and knockdown experiments

ES cells were maintain MEFs in KSR culture media as described above and passaged onto Geltrex coated dishes in mTeSR1 culture medium prior to infection. When cells were ~75% confluent, cells were collected with Accutase as single cells or small clumps. 100,000 ES cells were plated per well of 12 well plate coated with Geltrex and in mTeSR1 culture medium. After 24 hours, ES cells were infected twice on separate days for 3 hours with approximately 30 viral particles per cell. 48 hours after the last infection, cells were selected with 1 μ g/ml puromycin until the non-infected ES cells die off (usually within 3 days). Knockdown (KD) and control shRNA-infected ES cell lines were then maintained as described above. We then performed directed differentiation of three control and KD cell lines into 5-day dEN, dME, and dEC. We collected cells and carried out RNA extraction as described above. cDNA reaction was set-up from 1 μ g of total RNA per sample using High-Capacity cDNA RT Kit (Life Technologies). qPCR was performed on 384-well TaqMan hPSC Scorecard plates using ViiA7 RUO software and Applied Biosystems ViiA7 instrument. CT values were normalized using two probes of the ACTN housekeeping gene. Other normalization genes were not used here, since they are reduced in expression in the KDs. We used pLKO.1 cloning vector with the following target sequences for EOMES (CCGTTTCAGAAGGAGACATTT, CCCAGATGATAGTCTTACAAT, CCATAAAGTGAGGACATTA), GATA4 (CCAGAGATTCTGCAACACGAA, CGAGGAGATGCGTCCCATCAA, CCCGGCTTACATGGCCGACGT), and OTX2 (GCACTGAAACTTTACGACAAA, GCTGGCTCAACTTCTACTTT,

CCATGACCTATACTCAGGCTT). The shRNA control cell lines targeted gene products not present in the human genome using the same cloning vector with the following target sequences: TGACCCTGAAGTTCATCTGCA (GFP) and CACTCGGATATTTGATATGTG (LUCIFERASE).

Computing shRNA knockdown *P* values

To compute significance in change of differentiation potential in all knockdown experiments we set the relevant control lines as the reference set and computed the differentiation potential_{2D} and corresponding *P* value for each shRNA knockdown experiment. The resulting *P* value calculation accounted for dependencies between genes as described above. For expected decrease in gene expression compared to the reference we used the left-sided, one sample *t*-test and for an expected increase we used the right-sided, one sample *t*-test. Analysis was performed using MATLAB version 8.1 release name R2013a, using the scorecardv5d0.m function provided in the **Supplemental Code**. *P*-values for the differentiation potential_{2D} of the three independent shRNA knockdown experiments were then combined using Fisher's method to compute the total *P* value of the knockdown lines compared to control lines.

Feeder versus feeder-free culture analysis

To compare feeder and feeder-free culture conditions, for each gene we computed the mean difference in CT expression, averaged over 11 matching cell lines grown in both culture conditions. Significance of the difference in expression across all cell lines was calculated both using the paired *t*-test and the Wilcoxon signed rank test (non-parametric), with both methods yielding very similar results. The distributions of mean difference in CT expression were then displayed for each gene set (EC, ME, EN, PL) using Box plots.

To quantify the effect of cell line adaptation due to repeated passaging in feeder-free culture, for each gene we again computed the mean difference in CT expression, averaged over 7 matching cell lines. To quantify the effect of hESC adaptation on subsequent differentiation, for each gene we calculated the mean expression difference during dEC, dME, and dEN differentiation. The distributions of mean difference in CT expression were then displayed for each gene set (EC, ME, EN, PL) and each experiment class (hESC, dEC, dME, dEN) using Box plots. Analysis was performed using MATLAB version 8.1 release name R2013a, using built-in functions `ttest.m` and `signrank.m` with 'tail' parameter set to 'both'.

To compute significance of the distribution of mean expression difference for each gene class, we combined the paired *t*-test *P* values for all genes within a gene class using the weighted *Z*-method, as described above. The resulting *P* values calculation accounted for dependencies between genes. For expected decrease in gene expression compared to the reference we used the left-sided, paired *t*-test and for an expected increase we used the right-sided, paired *t*-test. Analysis was performed using MATLAB version 8.1 release name R2013a, using the scorecardv5d0.m function provided in the **Supplemental Code**.

Lasso regression fitting

We performed Lasso regression algorithm to find the most sparse set of coefficients within the EC, ME, EN, and PL gene sets that best fit the differentiation potential_{EB} of the 23 cell lines in the day 12 EB reference, the differentiation potential_{2D} of the 14 cell lines in **Figure 4c**, and the 2D efficiency measures (**Fig. 4d**). Lasso was fit to the standard normal deviates $Z_k = \Phi^{-1}(1 - P_k)$, where P_k is the P value for the k -th gene in each gene set GS calculated using the one-tailed, one sample t -test relative to the reference set (see **Computing differentiation potential**) and Φ^{-1} denote the inverse standard normal cumulative distribution function (**Fig. 2a**). For each lambda we compute the Lasso regression coefficients based on all the data and the corresponding Mean Squared Error (MSE) for that fit. We then used five-fold cross validation (where 80% of the data was used for training and 20% for testing) to estimate the standard error of the MSE for each lambda fit. The Lasso coefficients chosen for all analyses were the ones with the largest lambda such that the MSE is within one standard error of the minimum MSE (labeled “coef_1SE” in **Supplementary Table 4**). This lambda makes the sparsest model within one standard error of the minimum MSE. Grey error bars in the MSE plots (**Supplementary Figs. 5, 10**) indicate the standard error computed using the cross-validation for each value of lambda. The green circles indicate the Lambda with a minimum MSE. The blue circles indicate the largest lambda such that the MSE is within one standard error of the minimum MSE. We found that the coefficients chosen within one standard error of the minimum MSE explained at least 99% of the differentiation potential variance ($R^2 = .99$). We find that 5-12 markers per germ layer are sufficient to calculate a near perfect fit to differentiation potential_{EB} (**Supplementary Fig. 5c**), 6-8 markers for 2D differentiation potential and 2-3 markers for predicting 2D efficiency (new **Supplementary Fig. 10f**). The Lasso fits to 2D efficiency using 2-3 genes correlated at $R = .94$ for all three germ layers, which was higher than the correlation with the corresponding 2D differentiation potential scores. Analysis was performed using MATLAB version 8.1 release name R2013a, using built-in functions lasso.m with parameter ‘CV’ set to 5 and lassoPlot.m with parameter ‘PlotType’ set to ‘CV’.

Marker reduction analysis

To assess the predictive power of our assay using a subset of markers (**Fig. 6b,c**), we calculated the differentiation potential_{2Dsubset} using a subset of EN, ME, and EC marker genes and then found the Pearson correlation coefficient (R) of the differentiation potential_{2Dsubset} with the overall differentiation potential_{2D} using all markers and with the 2D efficiency, described in the FACS analysis section. Subsets of marker genes and corresponding weights were chosen using Lasso regression algorithm fits as described above at different values of lambda. Correlations to the overall differentiation potential_{2D} and 2D efficiency for all lambdas with n non-zero coefficients were averaged to obtain one correlation value for all fits with n markers.

We also measured the Pearson correlation coefficient (R) of 2D efficiency versus the standard normal deviate $Z_k = \Phi^{-1}(1 - P_k)$ for each individual gene k , where P_k is the P value for the k -th gene calculated using the one-tailed, one sample t -test relative to the reference set (see **Computing differentiation potential**) and Φ^{-1} denote the inverse standard normal cumulative distribution function. These single gene correlations are displayed in **Figure 6c**.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank all members of the Meissner lab as well as Deborah Tieberg, and Uma Lakshmiathy for their valuable support and insightful feedback, Lauren Solomon and Leslie Gaffney for graphical support, Ik Soo Kim and Bradley Bernstein for the JQ1 inhibitor molecule and insight, Lee Rubin for the partially reprogrammed cell lines, and Catherine MacGillivray (HSCI Histology core) for teratoma sectioning, fixation, and histochemical staining for markers of the 3 germ layers. A.M.T is supported by NIH under Ruth L. Kirschstein NRSA fellowship 5F32DK095537. This work was supported by NIGMS (P01GM099117), NHGRI (1P50HG006193), the New York Stem Cell Foundation and a research grant from Life Technologies and the New York Stem Cell Foundation. A.M. is a New York Stem Cell Foundation, Robertson Investigator.

References

1. Takahashi K, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*. 2007; 131:861–872. [PubMed: 18035408]
2. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 2006; 126:663–676. [PubMed: 16904174]
3. Daley GQ. The promise and perils of stem cell therapeutics. *Cell stem cell*. 2012; 10:740–749. [PubMed: 22704514]
4. Dolgin E. Putting stem cells to the test. *Nat Med*. 2010; 16:1354–1357. [PubMed: 21135828]
5. Muller FJ, Goldmann J, Loser P, Loring JF. A call to standardize teratoma assays used to define human pluripotent cell lines. *Cell stem cell*. 2010; 6:412–414. [PubMed: 20452314]
6. Muller FJ, et al. A bioinformatic assay for pluripotency in human cells. *Nature methods*. 2011; 8:315–317. [PubMed: 21378979]
7. Bock C, et al. Reference Maps of Human ES and iPS Cell Variation Enable High-Throughput Characterization of Pluripotent Cell Lines. *Cell*. 2011; 144:439–452. [PubMed: 21295703]
8. Wang Z, Oron E, Nelson B, Razis S, Ivanova N. Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. *Cell stem cell*. 2012; 10:440–454. [PubMed: 22482508]
9. Gifford CA, et al. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell*. 2013; 153:1149–1163. [PubMed: 23664763]
10. Lipták T. On the combination of independent tests. *Magyar Tud Akad Mat Kutato Int Közl*. 1958; 3:171–196.
11. Boulting GL, et al. A functionally characterized test set of human induced pluripotent stem cells. *Nature biotechnology*. 2011; 29:279–286.
12. Stouffer, S.; DeVinney, L.; Suchmen, E. *The American soldier: Adjustment during army life*. Vol. 1. Princeton University Press; Princeton, US: 1949.
13. Tsankov AM, et al. Transcription factor binding dynamics during human ESC differentiation. *Nature*. 2015 in press.
14. Liao J, et al. Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. *Nature Genetics*. 2015 in press.
15. Borowiak M, et al. Small molecules efficiently direct endodermal differentiation of mouse and human embryonic stem cells. *Cell stem cell*. 2009; 4:348–358. [PubMed: 19341624]
16. Teo AK, et al. Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes Dev*. 2011; 25:238–250. [PubMed: 21245162]
17. Ziller MJ, et al. Dissecting neural differentiation regulatory networks through epigenetic footprinting. *Nature*. 2014
18. Loven J, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*. 2013; 153:320–334. [PubMed: 23582323]

19. Pikkarainen S, Tokola H, Kerkela R, Ruskoaho H. GATA transcription factors in the developing and adult heart. *Cardiovasc Res.* 2004; 63:196–207. [PubMed: 15249177]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

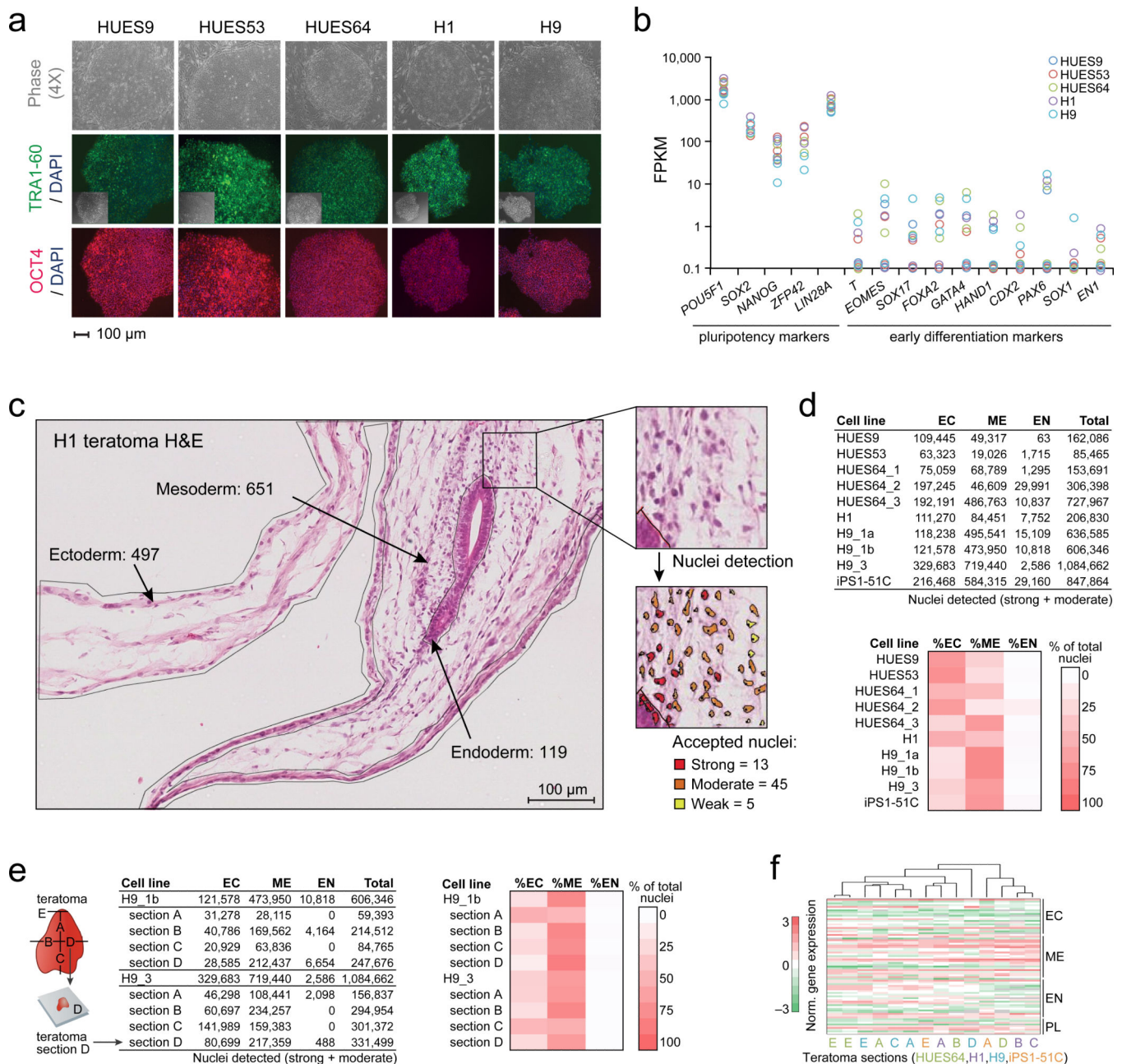


Figure 1. Common assays to characterize pluripotent cells

a. Phase images (4X magnification), OCT4 and TRA1-60 (10X magnification) immunofluorescent stainings (rows) for five selected human stem cell lines (columns). Scale bar corresponds to the 10X images.

b. RNA-seq expression level (FPKM) in replicate of several pluripotency (left) and early differentiation (right) markers for five selected cell lines, color-coded as shown in top right.

c. H1 teratoma H&E stain shows presence of tissue from the three embryonic germ layers, where numbers represent nuclei within each tissue annotation (black lines). Nuclei were digitally detected, counting only strong and moderate nuclei.

d. Top: Number of moderate plus strong nuclei detected within annotations for tissues derived from the three germ layers for six cell lines (rows). Biological replicates (teratomas grown in different mice) are indicated as number following the underscore and technical replicates (parallel sections from the same teratoma) are indicated as letters following the number. Bottom: Heatmap normalizes the nuclei counts to percentages per teratoma and shows a higher presence of ectoderm (EC) and mesoderm (ME) tissue than endoderm (EN) in all cell lines.

e. Left: Schematic of a three-dimensional teratoma and the location of the 3-5 two-dimensional sections used in the pathologist analysis. Middle: Number of moderate plus strong nuclei detected within all germ layer annotations for 4 independent sections A-D from teratoma replicates H9_1b and H9_3. Right: Heatmap shows that the normalized counts of ectoderm and mesoderm are highly variable between sections.

f. Hierarchical clustering of normalized RNA levels from different teratoma sections (columns) across gene markers (rows) for the 3 germ layers (EC, ME, EN) and pluripotency (PL). Clustering was carried out using correlation coefficient as a similarity measure and average linkage. Color codes for sections A-E indicate the cell line of origin, summarized in the legend below the x-axis.

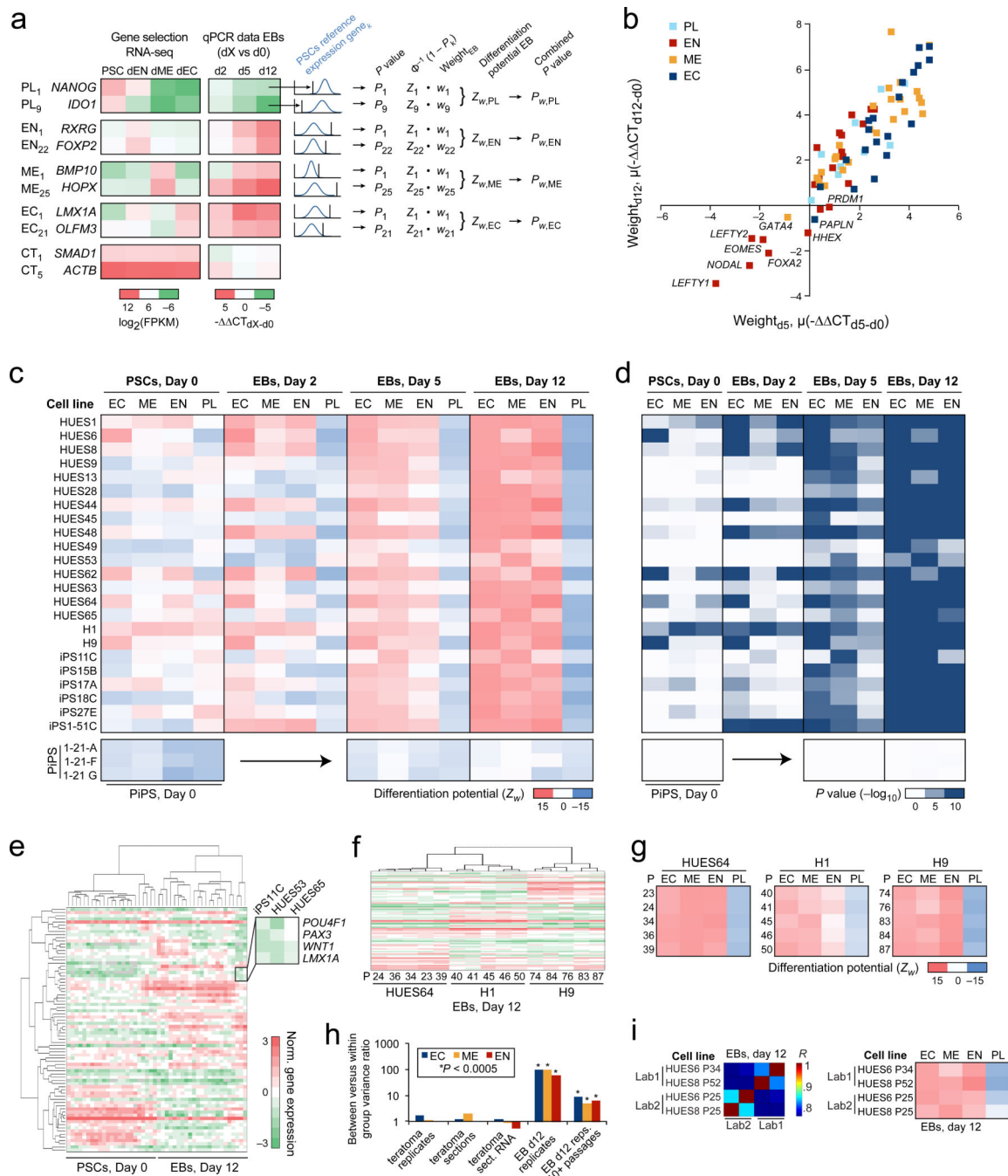


Figure 2. Design of qPCR assay and characterization of EB formation

a. Schematic of gene selection (left), EB time-course expression data (middle) and computational technique (right) used for quantifying EB differentiation potential for PL, EN, ME, and EC gene classes. For gene selection (Methods), we used RNA-seq data of HUES64 derived endoderm (dEN), mesoderm (dME), and ectoderm (dEC) populations.

b. Scatter plot of the day 12 EB weights versus day 5 EB weights. Genes above/below the diagonal line $y=x$ are more highly upregulated in day 12/day 5. Negative weights are set to 0 for all EB analyses.

- c. Heatmap of the differentiation potential_{EB} for all gene classes (EC, ME, EN, PL) at day 0, 2, 5, and 12 of EB formation. Partially reprogrammed lines (bottom) exclude the day 2 timepoint.
- d. Combined *P* values corresponding to the differentiation potential_{EB} for all three germ layers at day 0, 2, 5, and 12 of EB formation.
- e. Hierarchical clustering of normalized RNA levels from all hPSCs' day 0 and day 12 EB experiments (columns) and clustering of gene markers (rows) from the 4 gene classes using correlation coefficient as a similarity measure and average linkage. Rectangular box (top right) zooms in on three cell lines with low expression for several ectoderm markers. Normalized RNA levels are displayed in legend (bottom right).
- f. Hierarchical clustering of normalized RNA levels for day 12 EB experiments from three cell lines maintained over five different passages (columns) and clustering of gene markers (rows) from the 4 gene classes. Legend for normalized RNA levels is shown in panel e.
- g. Heatmap of the differentiation potential_{EB} per gene class (EC, ME, EN, PL; columns) at day 12 of EB formation for HUES64, H1, and H9 cultured over several passages (rows).
- h. Ratio of between to within group variance for germ layer differentiation potential as quantified by teratoma formation (left) and by our qPCR based assay (right). Germ layer variance ratios are shown using different colored bars and the asterick above bars correspond to a significantly lower variance between replicates than between cell lines ($P < 0.0005$, *F*-test).
- i. Pearson correlation coefficient (*R*) matrix (left) and heatmap of the differentiation potential_{EB} (right) of day 12 EB gene expression signatures for HUES6 and HUES8 cell line stocks from the Meissner ("Lab1") and Melton ("Lab2") labs. Legend for differentiation potential_{EB} is shown in the bottom right of panel g.

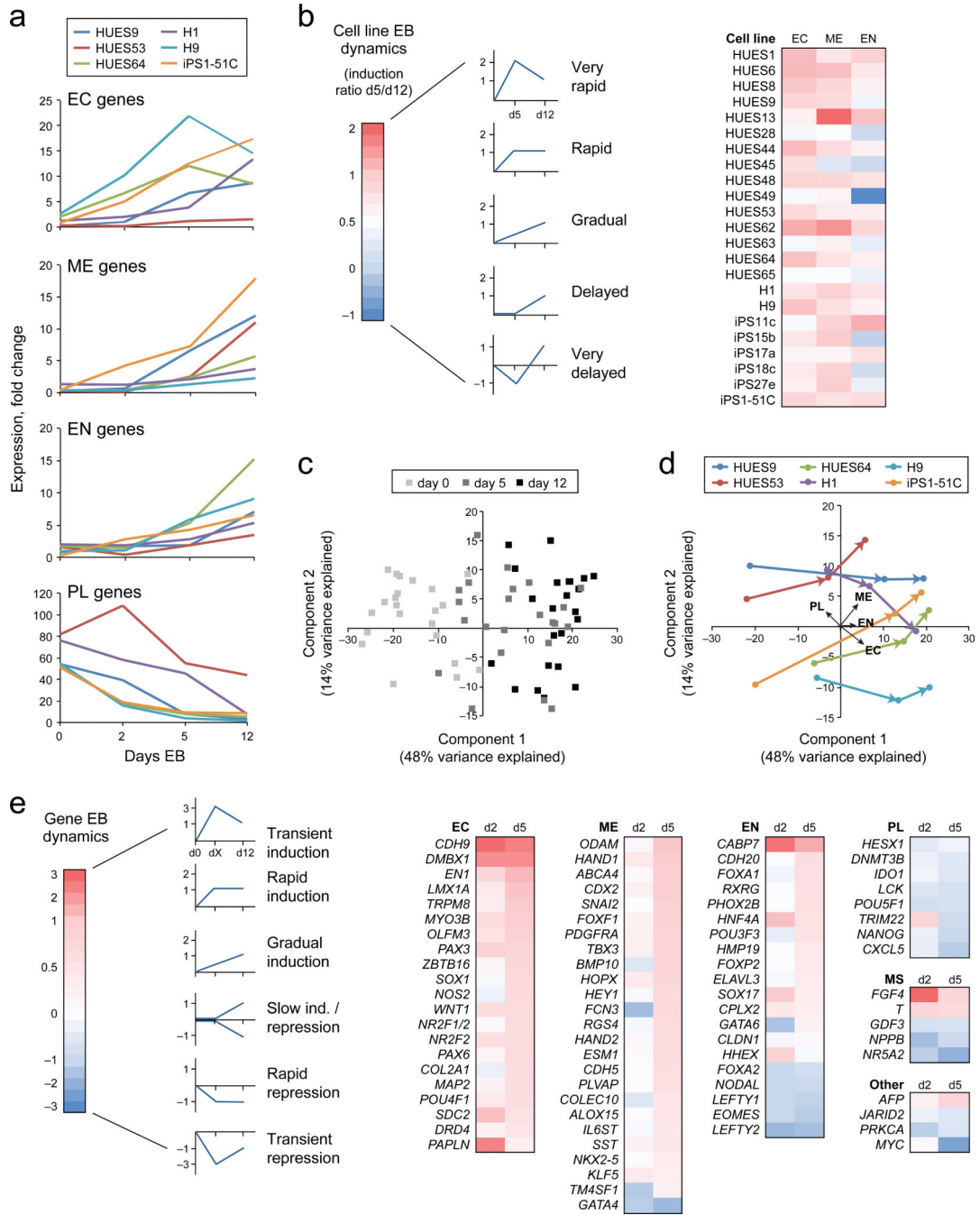


Figure 3. Quantifying EB differentiation dynamics

a. Average ectoderm (top), mesoderm, endoderm, and pluripotent (bottom) gene expression for six selected lines. Genes with $\text{weight}_{\text{EB}} < 0$ were excluded from the EB data analysis.

b. Quantification of cell line dynamics during EB differentiation. Heatmap shows the ratio of day 5 to day 12 gene expression, averaged across all markers for the three germ layers (EC, ME, EN). Ratios near or above 1 indicate rapid induction, while near or below 0 indicate delayed germ layer response.

- c. Scatter plot of the first and second principal components for all day 0, 5 and 12 EB data show that differentiation follows along the direction of component 1. The starting point in differentiation is day 0 (on the left) while the end point is day12 (on the right).
- d. Scatter plot of the first and second principal components of day 0, 5, and 12 EB data for the six selected cell lines. Small arrows show the direction of differentiation trajectory and the successive points correspond to day 0, 5 and 12 respectively. Vectors at the intersection of the axes map the direction of PL, EN, ME, and EC gene sets.
- e. Quantification of gene expression dynamics during EB differentiation for different gene classes. Heatmapt shows the ratio of day 2, 5 to day 12 expression levels, relative to day 0 baseline expression. Positive/negative values correspond to different induction/repression dynamics (see panels on left).

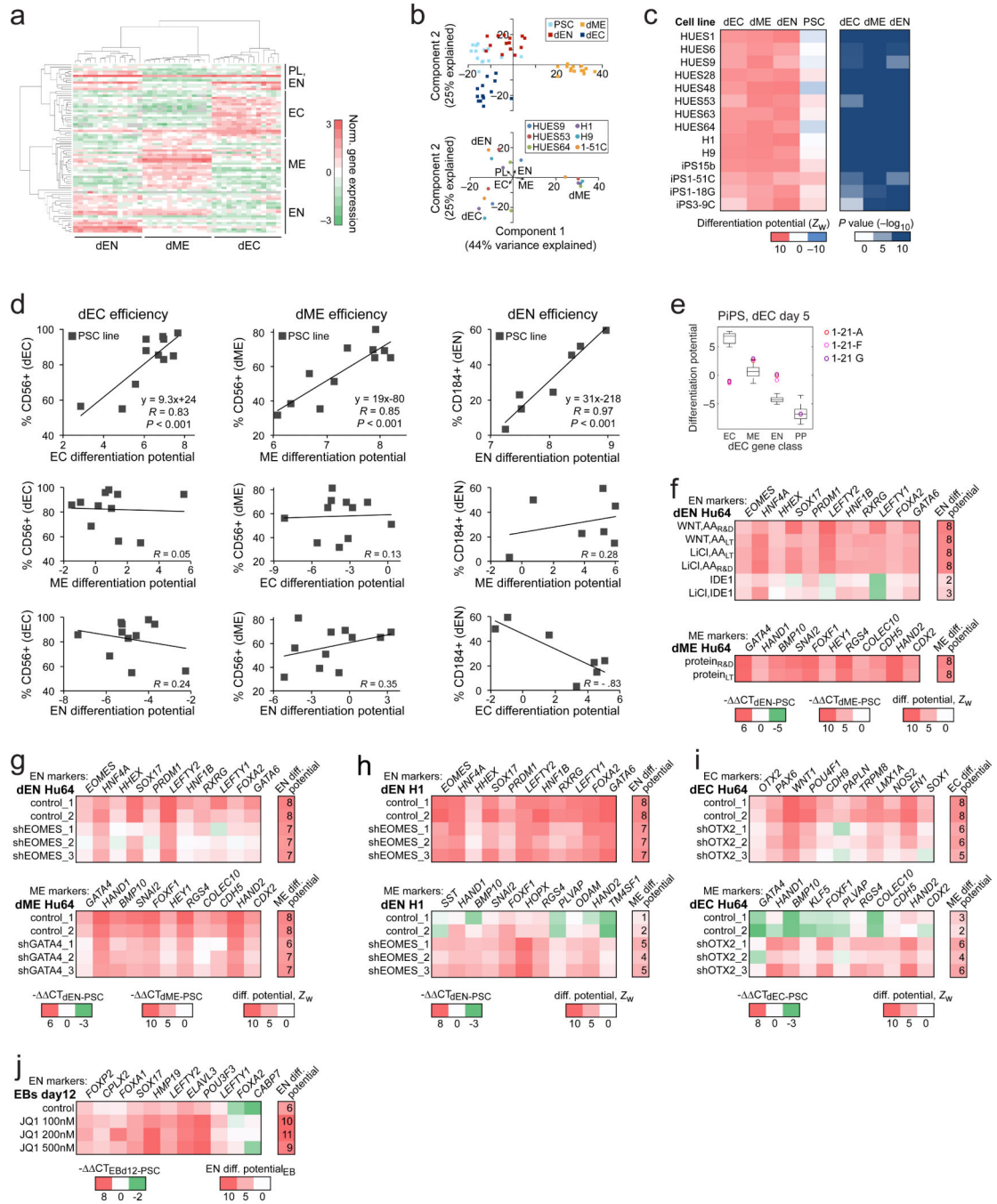


Figure 4. Directed differentiation of hPSCs and new applications

a. Hierarchical clustering for all dEN, dME, and dEC 2D experiments (columns) and gene markers (rows) from the 4 gene classes using correlation coefficient as a similarity measure and average linkage.

b. Scatter plot of the first and second principal components for all directed differentiation experiments (top) and for six selected cell lines (bottom) show clear separation between samples for the 3 germ layers. Vectors at the intersection of the axes map the direction of PL, EN, ME, and EC gene sets.

- c. Heatmap of the differentiation potential_{2D} (left) and combined *P* value (right) of the four different gene classes for 14 cell lines.
- d. Top: Linear regression shows a high correlation between differentiation potential_{2D} of the pertinent gene class and FACS quantification for CD56 in dEC (left) and dME (middle), and CD184 in dEN (right). Middle/Bottom: Correlation between alternate gene classes and dEC, dME, and dEN efficiency do not correlate positively.
- e. Box plots illustrate the distribution of differentiation potential_{2D} for the 4 gene sets (EC, ME, EN, and PL) of the 14 reference dEC experiments. Circles represent all three partially reprogrammed iPS (PiPS) cell lines tested following 5 day directed differentiation into ectoderm.
- f. Heatmaps showing gene expression level (left) and differentiation potential_{2D} (right) of several different protocols for dEN (top) and dME (bottom) differentiation. LT=Life Technologies, R&D=R&D Systems, WNT=WNT3A_{R&D}, AA=Activin A, protein=recombinant proteins AA, BMP4, and VEGF.
- g. Heatmaps showing gene expression level (left) and differentiation potential_{2D} (right) for three shRNA knockdowns of EOMES and GATA4 during HUES64 dEN and dME differentiation, respectively. We observe decreased EN expression and differentiation potential_{2D} (top) for EOMES and decreased ME expression and differentiation potential_{2D} (bottom) for GATA4 in the knockdowns compared to control experiments.
- h. Heatmaps showing gene expression level (left) and differentiation potential_{2D} (right) for three shRNA knockdowns of EOMES during dEN differentiation in H1. We observe decreased EN expression and differentiation potential_{2D} (top) and increased ME expression (bottom) in the knockdowns compared to control experiments.
- i. Heatmaps showing gene expression level (left) and differentiation potential_{2D} (right) for three shRNA knockdowns of OTX2 during dEC differentiation in HUES64. We observe decreased EC expression and differentiation potential_{2D} (top) and increased ME expression (bottom) in the knockdowns compared to control experiments.
- j. Heatmaps showing gene expression level (left) and differentiation potential_{EB} (right) after 12 days of EB formation using HUES64 cells treated with different concentrations of JQ1 bromodomain inhibitor (100nM, 200nM, 500nM, and 0nM untreated control). We observe increased expression of several EN markers and increased EN differentiation potential_{EB} in EBs treated with JQ1.

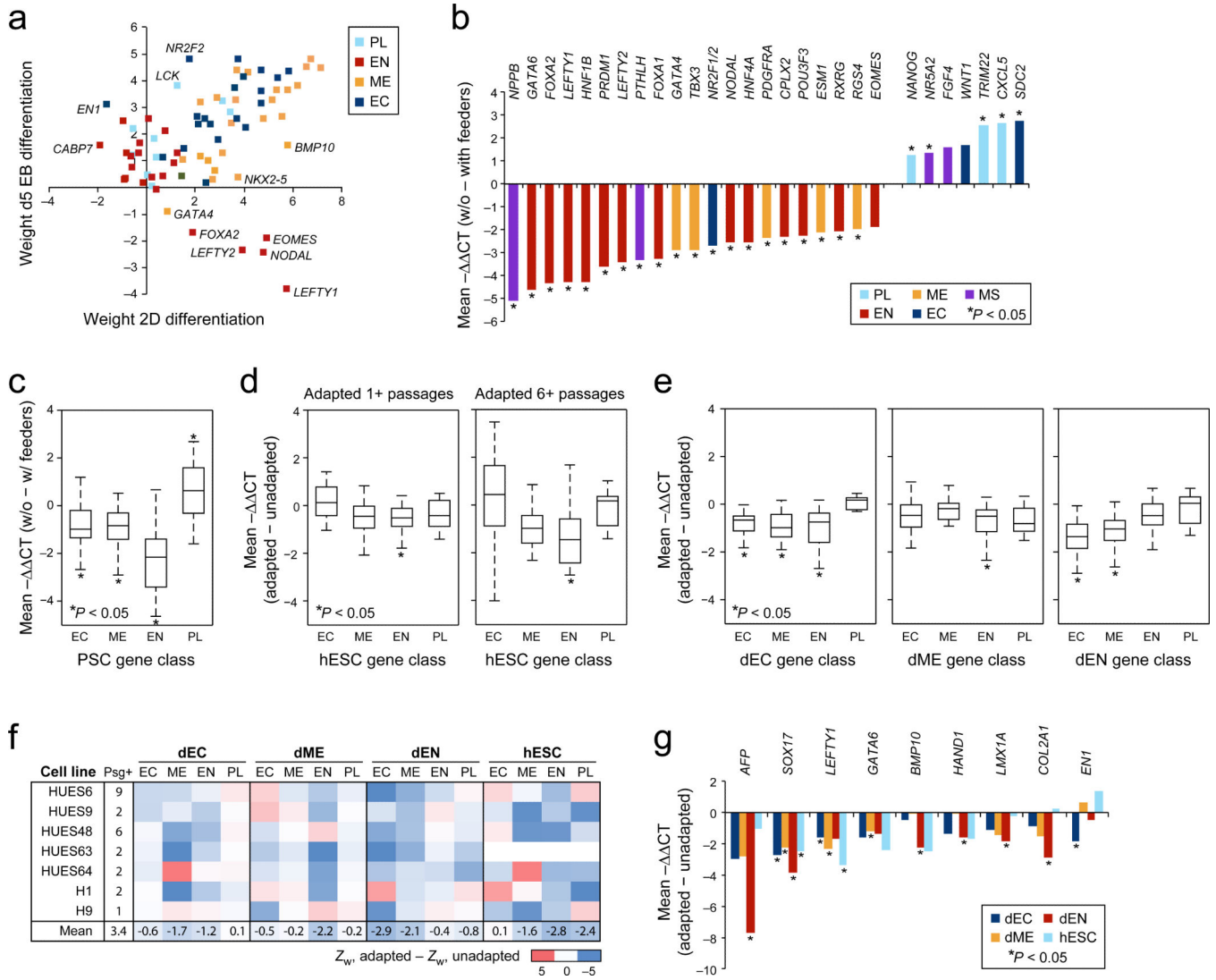


Figure 5. Tissue culture quantification

- Scatter plot of day 5 EB weights versus 2D weights. Labeled genes above/below the diagonal line $y=x$ are more highly upregulated in day 5 EB/2D experiments relative to day 0.
- Bar graph of significantly ($P < .05$, Wilcoxon signed rank test; marked with asterisk) downregulated (left) and upregulated (right) genes in feeder-free versus feeder tissue culture. Y-axis measures the mean expression difference, averaged over 11 matching cell lines. Bars are colored according to gene classes (MS = mesendoderm).
- Box plot of the distribution of mean expression difference between feeder-free and feeder cultured hPSC lines for all genes belonging to the four gene classes (EC, ME, EN, PL).
- Box plots of the distribution of mean expression difference between adapted and unadapted hESC lines for all genes belonging to the four gene classes. EN and ME mean expression decreases after adaptation of lines for 1+ passages (left) and 6+ passages (right) in feeder-free culture.

- e. Box plot of the distribution of mean expression difference between adapted and unadapted hESC lines following directed differentiation into dEC (left), dME (middle) and dEN (right). Gene expression differences were categorized into four gene classes (EC, ME, EN, PL).
- f. Differentiation potential between adapted and unadapted hESC lines (first column) after directed differentiation into dEC (left), dME, dEN and for the undifferentiated control (right). Mean differences for all cell lines is shown on the bottom row.
- g. Bar graph of significantly ($P < .05$, t -test; marked with asterisk) downregulated genes in cell lines adapted versus unadapted to feeder-free tissue culture. Y-axis measures the mean expression difference, averaged over 7 matching cell line experiments. Bars are colored according to cell type.

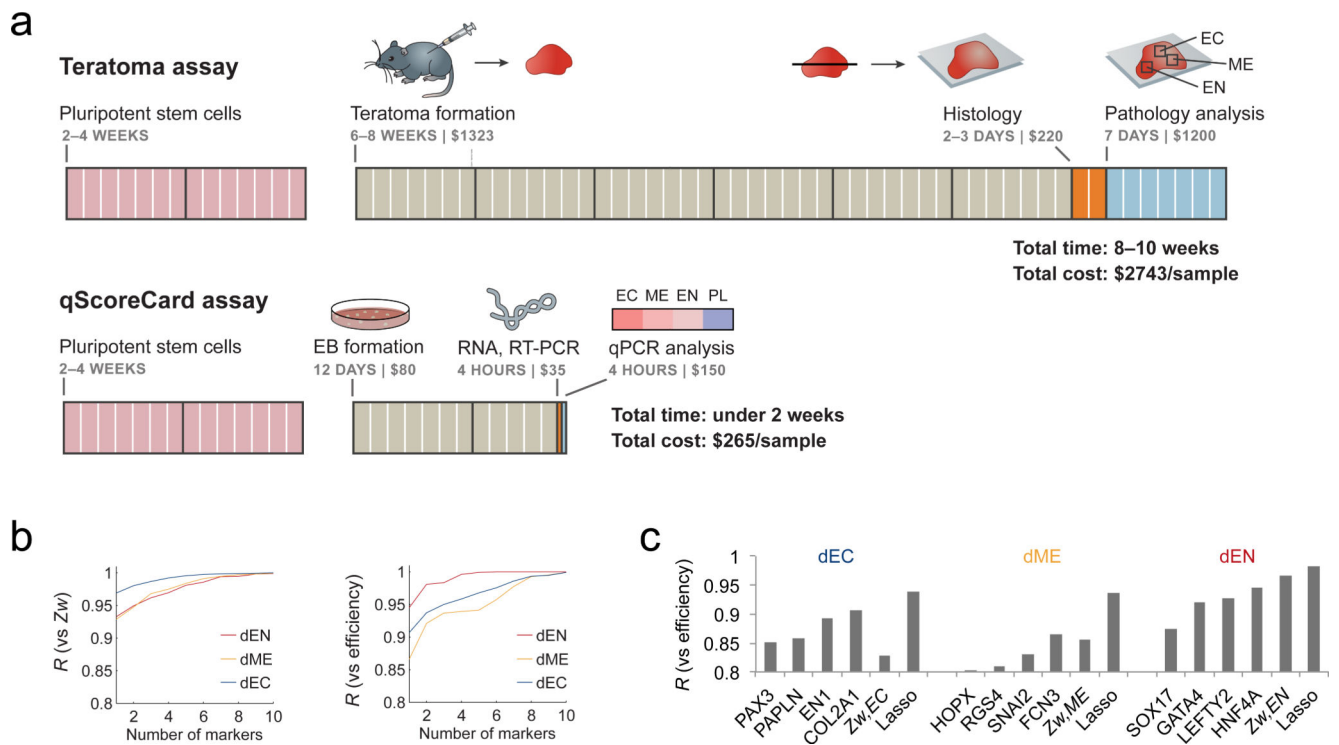


Figure 6. Future outlook of qPCR expression assay

a. Schematic of the timelines and costs for performing teratoma formation (top) versus qPCR expression assay for assessing hPSC utility (bottom).

b. Pearson correlation coefficient (R) of the differentiation potential_{2D} (calculated using a subset of EN, ME, and EC markers) is plotted when correlated with the differentiation potential_{2D} using all markers (left) or with 2D efficiency (right). Subsets of marker genes and weights were chosen using the Lasso regression algorithm. R increases with number of marker genes used (x-axis).

c. Pearson correlation coefficient (R) of 2D efficiency versus the differentiation potential_{2D} calculated for each germ layer (left: dEC; middle: dME; right: dEN) using the expression of individual markers, using all markers (Z_w , G_S), and using 2-3 marker genes and weights learned by Lasso regression (rightmost bars).