Elsevier Editorial System(tm) for Cell Stem Cell
Manuscript Draft

Manuscript Number: CELL-STEM-CELL-D-11-00358R3

Title: Molecular signatures of human induced pluripotent stem cells (hiPSC) highlight sex differences and cancer genes

Article Type: Research Article

Keywords: hiPSC; X inactivation; XIST; epigenetic; microarray; sex differences

Corresponding Author: Professor Jeannie Lee,

Corresponding Author's Institution: MGH

First Author: Montserrat Anguera

Order of Authors: Montserrat Anguera; Jeannie Lee

Abstract: Although human induced pluripotent stem cells (hiPSCs) have enormous potential in regenerative medicine, their epigenetic variability suggests that some lines may not be suitable for human therapy. There are currently few benchmarks for assessing quality. Here we show that X-inactivation markers can be used to separate hiPSC lines into distinct epigenetic classes and that the classes are phenotypically distinct. Loss of XIST expression is strongly correlated with upregulation of X-linked oncogenes, accelerated growth rate in vitro, and poorer differentiation in vivo. Whereas differences in X-inactivation potential result in epigenetic variability of female hiPSCs lines, male hiPSC lines generally resemble each other and do not overexpress the oncogenes. Neither physiological oxygen levels nor HDAC inhibitors offer advantages to culturing female hiPSC lines. We conclude that female hiPSCs may be epigenetically less stable in culture and caution that loss of XIST may result in qualitatively less desirable stem cell lines.

*Cover Letter

HHMI
HOWARD HUGHES MEDICAL INSTITUTE

February 7, 2012

Re: CELL-STEM-CELL-D-11-00358-R2

Deborah Sweet, PhD
Editor
Cell Stem Cell
600 Technology Square, 5th floor
Cambridge, Massachusetts 02139

Dear Debbie,

We thank you and the Reviewers for additional recommendations concerning our manuscript, "*Molecular signatures of human induced pluripotent stem cells (hiPSC) highlight sex differences and cancer genes*" by Anguera et al. (CELL-STEM-CELL-D-11-00358). We have now revised the manuscript accordingly, and a point-by-point response to the Reviewer comments follows this letter. Changes to the manuscript include:

(1) More teratomas added to Figure 4: We now have a few more teratomas from Class II and III hiPSC lines to support our original conclusion. A total of four distinct Class II and three distinct Class III lines are tested, and each made multiple teratomas (4-11). All Class II teratomas are solid and well diferentiated, whereas all Class III teratomas are cystic and poorly diferentiated. Fig. 4D now shows representative teratomas made from two matched hiPSC lines, comparing solid teratomas generated from the Class II state versus the cystic teratomas of the Class III state. The contrast is dramatic and very reproducible, further adding to our conclusion that XIST+ and XIST- sublines from the same original cell line are phenotypically distinct.

(2) Toned down conclusions: In response to the remaining concerns of Rev2 and the new comments of Rev4, we have toned down statements in the Abstract and Discussion. We also mention specific caveats to our data and discussion where Rev2 and Rev4 have raised alternative interpretations. Please see the point-by-point response for examples.

(3) Elimination of ~9K characters from the text: from ~68K to ~59K characters. We did so by eliminating redundancies in wording and unnecessary detail in the Methods. We have also reduced the Abstract count to <148 words.

We would also like to comment on three points raised in your decision letter:

(i) Use of VPA during reprogramming: VPA was used in standard practice at the time we initiated the study (October 2009). However, we agree that the addition of VPA during

reprogramming appears out of date compared to current reprogramming protocols. We used VPA to create the first generation of hiPSC lines (hiPS-1,2,3,9,10,11,12), using the most efficient reprogramming protocols available at the time (October 2009). Importantly, all passage 0 hiPSC lines (using 3 different fibroblasts) and male fibroblasts were reprogrammed in June 2011 without VPA. We have updated the methods section. Note that there were no major differences with respect to occurrence of Class I, II, and III cells whether VPA is used or not (Table I). Thus, we do not believe that VPA has a noticeable impact on X-reactivation and XCI states in general. We have also generated more hiPSC lines from IMR-90 cells without VPA (data not shown in this manuscript) and they also have the same percentage of XIST-positive/negative cells in the undifferentiated state as lines shown in Table I. ***Importantly, all lines – regardless of whether VPA or ROCK inhibitor was used – had similar proportions of Class I, II , and III (Table I).***

(ii) Passage number: Rev2's request to study all cell lines at the same passage number is a nearly impossible one. **No papers in the stem cell field adhere to this ideal !!** In our response to Rev2, you will find a summary of what other major labs have done with regard to passage numbers – and it will be clear that none of the hiPSC/hESC comparisons made by the papers of Envers et al., the Meissner/Eggan/Hochedlinger Labs, and the Plath Lab have actually used passage-number-matched cell lines. None. Our attempts to use low-passage and closely passage-matched cell lines are at least as rigorous as theirs.

Apart of the impracticality of doing this, it should also be mentioned that exact passage numbers are not an exact science. A passage is simply defined as a split in culture. It is not standardized to the number of cell divisions. Therefore, two vials of "passage 5" cells may have undergone different numbers of cell divisions; conversely, two vials, one of "passage 3" and the other of "passage 8", may have had the same number of divisions. We think it is more meaningful and practical to separate sublines by "low" versus "high" passage.

There is a major difference between our study and the three referenced by the Rev2 to highlight his/her point about passage numbers. The hESCs from the three references have abnormal karyotypes and were studied at much higher passage numbers than our study. Importantly, by contrast, all of our hiSPC lines have normal karyotypes and relatively low passage numbers. In the study by Yang et al., 'normal passage' of their HES-3 line was observed up to passage 44, aberrant change at passage 44-99, and complex changes at passage 142-182 (determined by CGH and karyotype analyses). The study of Enver et al. observes 'adaptation' at "more than 100 passages". For almost all of our experiments, the hiPSC lines were taken at passages 0-40. The highest passage used in any experiment was ~p.50 (just one line). And all were karyotypically normal, several verified to be normal by CGH.

(iii) We also think that Rev2's concerns regarding use of the XXY cell line for HDACi studies are misplaced. None of the other Reviewers shared this concern, including Rev4, who seems familiar with X-inactivation. As stated previously, one must distinguish between sex *per se* (maleness, femaleness) and X-chromosome number (and consequently whether a cell requires

Jeannie T. Lee, M.D., Ph.D.
Investigator

Massachusetts General Hospital, Department of Molecular Biology
Harvard Medical School, Department of Genetics
Simches Research Center, CPZN7250
185 Cambridge Street, Boston, MA 02114
lee@molbio.mgh.harvard.edu

dosage compensation, XCI). They are not the same issue. For example, XO individuals are female but they do not undergo XCI; whereas XXY individuals are male and they do undergo XCI. Therefore, sex is not a determinant of the need to undergo XCI. The point of our study is the aberration of XCI patterns in hiPSC – not the female sex itself – is associated with cancer-like gene profiles. Therefore, our decision to use the XXY cells for reprogramming not only provided much more efficient reprogramming and a more sensitive assay (as stated in the Results) but also helped support this idea. Note that XXY embryos count correctly and silence one X, expressing XIST from one X in spite of the presence of the Y-chromosome. Thus, the fact that the XXY hiPSC line is also a heterogeneous mixture of Class I, II, and III cell is a significant finding, indicating that the aberrant XCI status is not related to sex *per se* but is an aberration of the dosage compensation mechanism in hiPSC grown *in vitro*. ***Note that the XXY line was used only in the context of the HDACi studies (for their enhanced reprogramming efficiencies). These tests yielded negative results. We therefore consider this section of the manuscript to be relatively minor. The major and most interesting results of the manuscript used XX cell lines.***

With these additional changes, I hope you find the work acceptable for publication in *Cell Stem Cell.* We appreciate your reconsideration and look forward to hearing from you.

Sincerely,

Professor of Genetics
Harvard Medical School

Jeannie T. Lee, M.D., Ph.D.
Investigator

Massachusetts General Hospital, Department of Molecular Biology
Harvard Medical School, Department of Genetics
Simches Research Center, CPZN7250
185 Cambridge Street, Boston, MA 02114
lee@molbio.mgh.harvard.edu

RESPONSE TO REVIEWERS

Reviewer 1:

The revised version of the manuscript entitled "Molecular signatures of human induced pluripotent stem cells (hiPSC) highlight sex differences and cancer genes" by Anguera et al. presents additional data and contains changes to the text that have largely improved on the weaknesses of the original submission. In particular, the authors now show an expression analysis of the X-linked PGK gene along with XIST and importantly a polymorphic gene expression analysis. These additional data are important in that they now clearly demonstrate that cells with two active X chromosomes exist to variable degree in the cultures and that upon differentiation a clonal X inactivation pattern is observed which is consistent with earlier reports from other groups. Furthermore, the revision shows that class III iPSCs have a much reduced developmental potential in an in vitro differentiation assay and thereby providing a link between the classification system and a biologically relevant aspect. Together these further experiments address my major concerns. However, there are some necessary corrections that need attendance before publication should be considered.

1. Figure 3: The labelling of the figure panels and reference in the legend do not match. Especially panel D seems to be missing from the figure which makes the assessment of the data presented difficult.

We have corrected the labeling for Figure 3 in the figure legends. Thank you for bringing this mistake to our attention.

2. Page 18: The description of the expression analysis and the multiple correlations are hard to follow. Especially the sentence "Dark red values for male hiPSCs (labelled blue)..." is confusing. The analysis of the expression data sets should be presented in a precise and statistically significant format trying as much as possible to avoid vague descriptions such as "in general" and without citing "exceptions" to strengthen the text.

We have re-written this section between pp.16-18. We think it is now easier to follow.

3. Page 23: The references from loss of Polycomb proteins from the Xi after conditional Xist deletion appear wrong or incomplete.

There was an Endnote glitch. We have removed the Poplinsky reference from this sentence – Thank you.

4. In general the manuscript appears to propose XIST expression as a molecular marker for the quality of female iPSCs. However, XIST expression is only suitable to pick out class II cells and both class I (highly desirable) and class III (less desirable) express low or no XIST. Therefore, the question needs to be discussed if this is a viable strategy if it potentially eliminates class I and what advantage XIST classification has over direct testing of the differentiation potential in an assay as demonstrated in figure 4.

Our study emphasizes less the idea that positive XIST expression is a marker of good hiPSC. It remains to be seen if all XIST+ female hiPSC are in fact "good." Our study instead makes the point that loss of XIST expression (Class III) is quite undesirable, because of (a) the upregulation of X-linked cancer genes, (b) faster-than-normal growth rates *ex vivo*, and (c) poor differentiation potential *in vivo*. We suggest going forward that, the XIST status in combination with differentiation potential *in vivo* be used together to assess the quality of hIPSC. We have added this point to the Discussion (p.22).

Reviewer 2:

In the revised manuscript the authors added a significant body of new information; for example, they have included three additional iPS cell lines (reprogrammed in other labs), and have performed allelic specific expression on X chromosome. The additional data and analysis significantly improve the manuscript. However, the authors failed to properly address several of the comments: for instance, they have analyzed the effects of oxygen and HDAC inhibitors in a male cell line with Klinefelter syndrome (XXY). This cell line is clearly aberrant, and although it can obviously inactivate one of its X chromosomes, it is not a female cell line, and has three copies of pseudo-autosomal regions.

As we stated previously, one must distinguish between sex *per se* (maleness, femaleness) and X-chromosome number (and consequently whether a cell requires dosage compensation, XCI). They are not the same issue. For example, XO individuals are female but they do not undergo XCI; whereas XXY individuals are male and they do undergo XCI. Therefore, sex is not a determinant of the need to undergo XCI.

The point of our study is the aberration of XCI patterns in hiPSC – not the female sex itself – is associated with cancer-like gene profiles. Therefore, our decision to use the XXY cells for reprogramming not only provided much more efficient reprogramming and a more sensitive assay (as stated in the Results) but also helped support this idea. Note that XXY embryos count correctly and silence one X, expressing XIST from one X in spite of the presence of the Y-chromosome. There is no evidence that the pseudoautosomal region has any effect on reprogramming or XIST expression.

In truth, the fact that the XXY hiPSC line is also a heterogeneous mixture of Class I, II, and III cell is a significant finding, indicating that the aberrant XCI status is not related to sex *per se* but is an aberration of the dosage compensation mechanism in hiPSC grown *in vitro*.

Also, please note that the XXY line was used only in the context of the HDACi studies and only because this cell line – for whatever reason – gave us a much higher efficiency of reprogramming, which therefore facilitated the HDACi tests. *Note that these tests yielded negative results. We therefore consider this section of the manuscript to be relatively minor. The major and most interesting results of the manuscript used XX cell lines.*

The major unresolved issue is the expression of cancer related genes in class III cells. The authors compared class II and class III cell lines, when actually comparing cells from the same cell line at low and high passage number. Culture-adapted pluripotent stem cells (at high passage number) have been shown to change genetically and epigenetically in culture, see:
Enver et al.: "Cellular differentiation hierarchies in normal and culture-adapted human embryonic stem cells." Hum Mol Genet. 14:3129-40 (2005).
Yang et al.: "Tumor progression of culture-adapted human embryonic stem cells during long-term culture." Genes Chromosomes Cancer. 47:665-79 (2008).
Olariu et al.: "Modeling the evolution of culture-adapted human embryonic stem cells."

Stem Cell Res. 4:50-6 (2009).
Thus, their analysis is problematic, and suffers from a conceptual flaw. Instead of properly comparing multiple cell lines at the same passage number in either class II or class III status, they argue that they "believe" that the effects are not actually related to passage number.

We hope the Reviewer agrees that it is almost impossible to study all cell lines at the same passage number. To our knowledge, no paper in the stem cell field adheres to this ideal ! Passage numbers are often not reported at all. For example:

(a) Tchieu et al., 2010, Cell Stem Cell (Plath Lab, as cited): Figure S2 makes comparisons across several female hiPSC at different passage numbers (only a few at the same passage number) ….

STEMCCA 1 p.6
STEMCCA 4 p.9
STEMCCA 4 p.23
STEMCCA 6 p.9
MIP2 p.3
MIP3 p.3
STEMCCA E p.16
STEMCCA G p.16

(b) Enver et al. (Peter Andrews' Lab, as cited): H7 hESCs. Passage numbers not declared, hESC passages were classified as 'early' and 'late'.

(c) Yang et al (Guang-Xiu Lu's Lab): HES-3 hESCs. For microarrays, they used 'normal' at p.30, 'aberrant' at p.72, and 'complex' at p.182.

(d) Takahashi et al., 2006, Cell (Yamanaka's Lab, as cited): for male human iPSCs there's no passage info for microarray experiments.

(e) Park et al. 2008, Cell (George Daley's Lab, as cited): for their disease model hiPSCs, there is no passage information for microarray experiments.

(f) Bock et al., 2011, Cell (Kevin Eggan & Alex Messiner's Labs, as cited): hESCs passage numbers vary from p.17-65 (microarrays). The ES & EB comparisons were also not taken at the same passage numbers (off by 1-5 passages). hiPSC passage numbers were also unmatched and varied from p.13-43.

Apart of the impracticality of matching passage numbers precisely, it should also be mentioned that exact passage numbers are not an exact science. A passage is simply defined as a split in culture. It is not standardized to the number of cell divisions. Therefore, two vials of "passage 5" cells may have undergone different numbers of cell divisions; conversely, two vials, one of "passage 3" and the other

of "passage 8", may have had the same number of divisions. We think it is more meaningful and practical to separate sublines by "low" versus "high" passage.

We should also note that there is one major difference between our study and the three referenced by the Reviewer. The hESCs from the three references have abnormal karyotypes and were studied at much higher passage numbers than our study. Importantly, by contrast, all of our hiSPC lines have normal karyotypes and relatively low passage numbers. In the study by Yang et al., 'normal passage' of their HES-3 line was observed up to passage 44, aberrant change at passage 44-99, and complex changes at passage 142-182 (determined by CGH and karyotype analyses). The study of Enver et al. observes 'adaptation' at "more than 100 passages". For almost all of our experiments, the hiPSC lines were taken at passages 0-40. The highest passage used in any experiment was ~p.50 (just one line). And all were karyotypically normal, several verified to be normal by CGH.

Reviewer 3:

In this revision, the authors addressed most of the concerns raised by previous reviewers. Concerning the concurrent presence of class I, II, and III cells in female iPSCs, it is still unclear whether the minor population of class I is from the reprogramming of previously inactivated X in somatic cells. However, given the substantial data and detailed analyses of class II and III of female iPSCs in gene expression and cell growth rate, I think that this paper is now of general interests to the readers of Cell Stem Cell. I would therefore recommend the acceptance of this paper with a minor revision.

Specifically, in the section of "Class III association with upregulation of cancer-related genes" (page 13), a previous study by Shen et al. (2008) as cited in this manuscript showed that 44 X-linked genes are reactivated in the absence of XIST marker in female hESCs. This study could be cited together with the evidence of mouse studies by Csankovszki et al., (2001) and Zhang et al.(2007). In fact, it is of interest to examine whether the X-linked genes that are reactivated in XIST- class III iPSCs overlap with the genes identified in class III hESCs. It is also possible that reactivation in each line of female class III iPSCs could be cell-line specific. This can be done by the pair-wise comparison between class II and class III iPSCs.

We thank Reviewer #3 for the suggestion to intersect our lists of genes characteristic of Class III with the list of 44 reactivated X-linked genes from Shen et al (2008). We have performed this analysis and found that none of these 44 X-linked genes appears on our lists (Tables II, III, S-III). Shen et al. examined differentially expressed genes for one set of human ESCs (HSF6; XIST-positive vs negative). We speculate that X-reactivation may occur differently in hESC and hiPSC, though a more exhaustive study would be required to make conclusive statements. Also possible is that X-reactivation occurs stochastically and in a piecemeal fashion along the X. In fact, this is what we observed among the 5 Class III hiPSC lines -- note different expression profiles for the list of 30 Class III-reference genes (Table II) for the 3 Class III lines (hiPS 6C-1, hiPS 12D-1, hiPS-2 LO) shown in Figure 3B. If we examine the 6 X-linked genes in this list and compare the expression levels for all 5 Class III lines (shown in the heatmap of Figure 3B), we observe that some of these genes (HDHD1A, TCEAL3) follow the same X-reactivation pattern as the Class III average (set to zero and shown in black), and some genes exhibiting greater X-reactivation (FTX).

The bottom line is that, even amongst hiPSC lines, there appears to be stochastic differences in X-reactivation. We might expect hESC lines to behave similarly. Importantly, in the Class III lines, some X-linked cancer genes are consistently upregulated (these may be genes that are easiest to reactivate stochastically), which is why we believe that Class III lines may in general grow faster and have poorer differentiation potential.

Reviewer 4:

Lee and colleagues describe data directed at resolving the controversy about the X chromosome status in female human iPSCs. Previously, several studies proposed that the inactive X chromosome of female somatic cells is not reactivated during reprogramming to iPSCs, but that XIST expression is lost upon extended culturing. Other reports argued that the inactive X is reactivated during reprogramming allowing random silencing upon induction of differentiation. In addition to addressing this controversy, the authors attempt to show that the analysis of the X state in female human iPSCs can be used to group iPSCs into distinct classes that differ generally in gene expression and cell growth. Obviously, these topics are of great interest and the current controversy about the X chromosome in reprogramming needs to be resolved. The points I am making below summarize some issues I have with the data.

The authors make the following points:

1. They conclude that the X is reactivated during reprogramming, at least in some cells such that iPSCs when first obtained are XaXa. The majority of iPSCs at early passage are XiXa.

One important point to make is that the authors reprograme in VPA - i.e. the cells are exposed to VPA for 10 days during the reprogramming process. I also found in the methods that the authors add Rock Inhibitor to the cultures for the initial expansion of iPSCs, which is not a common practice. VPA, a histone deacetylase inhibitor has been shown to enhance human cell reprogramming dramatically. Histone deacetylation also affects X status (Ware et al). Even under those improved conditions the authors only get 1-6 iPSC clones per 50K starting cells - i.e. the reprogramming is rather inefficient and relatively slow. I am summarizing all these things here because they may demonstrate that the reprogramming condition are suboptimal, but more importantly, that the epigenetic state of the X could be altered during reprogramming by inhibition of deacetylases. Therefore, it is not clear to me that the data in this paper can be directly compared to studies that argued there is no reactivation during reprogramming. I would hope the authors put this important information into the abstract as it may be important for the interpretation of the data. To help clarify the controversy about Xi reactivation versus no Xi reactivation, I would suggest that the authors repeat their experiments under VPA free reprogramming conditions.

Reprogramming efficiencies do vary between labs and investigators, and we believe that this has to do with genetic backgrounds of starting cell lines, OSKM virus, and inter-lab variability in culture conditions. Unfortunately, this is an industry-wide occurrence and, until conditions are standardized in this nascent field, laboratories will continue to experience such inter-lab differences. In truth, our reprogramming efficiencies are not out of range with those of other laboratories. More to the point, the relative proportions of Class I, II, and III lines are similar in all of our cell lines, irrespective of whether VPA is used and genetic background (IMR90 versus XXY versus MM versus twins).

We agree that the addition of VPA during reprogramming appears out of date compared to current reprogramming protocols. We used VPA to create the first generation of hiPSC lines (hiPS-1,2,3,9,10,11,12), using the most efficient

reprogramming protocols available at the time (October 2009). Importantly, all passage 0 hiPSC lines (using 3 different fibroblasts) and male fibroblasts were reprogrammed in June 2011 without VPA. We have updated the methods section. Note that there were no major differences with respect to occurrence of Class I, II, and III cells whether VPA is used or not (Table I). Thus, we do not believe that VPA has a noticeable impact on X-reactivation and XCI states in general. We have also generated more hiPSC lines from IMR-90 cells without VPA (data not shown in this manuscript) and they also have the same percentage of XIST-positive/negative cells in the undifferentiated state as lines shown in Table I.

Beyond IMR-iPSC lines, the authors mention iPSCs from other starting cells, but it is not clear how much work was done on these lines, particularly regarding class III state. i.e. are conclusions regarding class II versus III only derived from iPSCs from one reprogramming experiment?

The conclusions were derived from more than one reprogramming experiment. Cell lines from 5 different genetic backgrounds have been studied for the Class I-II-III states (Table I). As noted above, usage of VPA made no difference to the outcome.

One interesting note the authors make is that SNP analysis did not detect transcription from both X chromosomes even in lines that are up to 14% XaXa. They claim that this is below detection level of their SNP expression analysis. To ensure this is the correct conclusion, why don't the authors mix two different male lines (with different SNPs) at various ratios and show that 14% of line A cannot be detected. Based on the data provided I am not convinced that there is Xi-reactivation, even in the VPA-induced reprogramming.

We believe that because Class I cells comprise only 8-14% of the cell lines, microarray analysis would not have been sensitive enough to pick up any biallelic expression. We therefore performed RNA FISH analysis using two approaches. In one, we queried X-linked Cot1 transcription in the XIST- subpopulation and saw clear Cot1 expression from both X-chromosomes in 8-14% of cells (Table I). In the second strategy, we examined expression of a specific X-linked gene (*PGK1*), and observed a mixture of cell populations. About half of XIST- cells showed biallelic nascent transcripts and the other half showed monoallelic expression. This is consistent with the mixture of Class I and Class III cells in the XIST- population: The Class I cells would be expected to show biallelic PGK1, whereas the Class III cells would be expected to show monoallelic PGK1 consistent with the occurrence of XCI in spite of the loss of XIST. Alternatively, the 50:50 result could mean that there is only partial X-reactivation in the Class I XIST- cells, in spite of the repetitive element fraction (Cot1) being generally active. Combined, these data tell us that the vast majority of hiPS-11 and -12 are Class II cells with a pre-fixed pattern of XCI, but that a small minority of cells have at least a partially reactivated X-chromosome.

2. Loss of XIST expression correlates with expression of cancer genes from the X chromosomes.

The authors describe that XiXa iPSCs show loss of Xist. This conclusion has been made prior to this work.

We agree. The study by Tchieu et al., as cited, was the first to demonstrate XIST loss upon prolonged passage of hiPSCs (Class III). The novelty of our manuscript lies in the biological consequences of the Class III state – (a) the upregulation of X-linked cancer genes, (b) faster-than-normal growth rates *ex vivo*, and (c) poor differentiation potential *in vivo*. Our study indicates that XIST loss leads to a qualitative difference between hiPSC lines. This important point has not been demonstrated before.

How do the conclusions made by the authors fit with the Meissner/Eggan data (Cell) that iPSCs and ESCs have a wide range of gene expression difference for each gene. Generally, the Meissner/Eggan data (Cell) argue that ESCs and iPSCs are not different from each other, even though they exclude the sex chromosomes for most of their analysis. More importantly, I am bringing up this paper because it describes nicely that a large number of cell lines need to be analyzed to reveal the normal spread of gene expression among all ESCs or iPSCs. Did the authors analyze enough class III and II lines (with two lines) to make the argument that a specific set of genes is normally deregulated in class III cells? I am not convinced. Is it always the same genes that are deregulated? There are only few genes that are misexpressed more than 2x - are the authors suggesting those make a signature? The claim that there is a unique genome-wide signature upon loss of Xist expression is probably an overstatement.

Yes, in fact, we do see consistent upregulation of specific genes. We found that, among genes showing significant differential expression (FDR < 0.05), only 10 coding genes were consistently upregulated more than 2-fold in Class III compared to all Class II lines (Table II-A; Fig. S6). Interestingly, among the genes upregulated in Class III hiPSC lines, X-linked genes were over-represented (4 out of 10 genes, $P = 7 \times 10^{-5}$). This is highly significant. Note that our microarray analysis was applied to multiple cell lines (Fig. 3,4).

We did not directly compare our work to those of Meissner/Eggan for the exact reason that the Reviewer noted: The *Cell* paper largely excludes analysis of X-linked genes. It is also difficult to compare our gene list to the Meissner/Eggan data because they grouped male and female lines together, and may have therefore diluted out possible heterogeneities. Thus, note: Another important point of our study is the sex differences and the implication that male and female lines should be studied separately in the future. Indeed, we found that published male hiPSCs have very different global gene expression patterns compared to female hiPSCs (Figure 3A), and hierarchical clustering also distinguishes both groups (data not shown).

Our findings are significant because we found expression differences of Class III cells despite high similarity among all female hiPSC lines. The heatmap and principal component analysis plots of Figure 2 demonstrate that: (a) the two Class III lines cluster together and apart from XIST+ hiPSCs; (b) all ten female hiPSC lines are quite similar to each other (R values are shown in Color Key for the heatmap in Fig. 3A). We generated lists of differentially expressed genes two

ways: (a) using a strict requirement that the gene must have at least 2-fold change for Class III samples and ALL 8 XIST+ samples (Table II); (b) using a relaxed requirement that the gene expression exhibit at least 2-fold change for 6 out of 8 XIST+ samples (Supplemental Tables S-II, III). It is very likely that analysis of more Class III hiPSC (correcting for genetic background) will yield different lists of genes compared to our lists. However, our data predicts that the gene expression profile of Class III state will always differ from the XIST+ matched sample. *In sum, we do NOT expect any one gene to be specific to any one class. Rather, we would find statistically significant the association of a group of genes with one class. In this case, we see a group of upregulated X-genes and a subset of X-oncogenes in Class III.*

I am also not clear from the presented data if class III iPSCs actually express more X-linked genes from the Xa or Xi relative to class II cells. This would be an important point to address to understand a potential mechanism? It is important to point out that class III ESCs have been associated with partial Xi reactivation. Thus, the presented data are not be very surprising?

The Reviewer is asking whether increased expression from X is due to upregulation from Xa or Xi. We assume that increased expression comes from Xi. There is no reason to think that Xa expression would be increased (no precedent for this whatsoever). Xi origin would be in agreement with partial and piecemeal X-reactivation on Xi observed by Shen et al. (2008), Zhang et al. (2007), and Csankovskii et al. (2001), as cited. Of course, we agree that partial X-reactivation with XIST loss is not new (this is acknowledged in the paper with mutliple citations which we indicated were consistent with our observation of partial X-reactivation). And while we agree that understanding potential mechanisms of X-reactivation is important, this question is – in our opinion – outside the scope of the current study. As stated above, our study makes the important and novel point of there being biological consequences of the Class III state – (a) the upregulation of X-linked cancer genes, (b) faster-than-normal growth rates *ex vivo*, and (c) poor differentiation potential *in vivo*. Based on a large amount of data in the paper, we believe that XIST loss leads to a qualitative difference between hiPSC lines.

Is there Tsix expression (Figure 2D)?

Yes. There appears to be TSIX expression in the hiPSC lines with a mixture of Class I, II, and III cells (Class II-predominant); and much less TSIX in lines with 100% Class III cells (see red line in Fig. 2D). We believe this difference in TSIX levels to be due to the presence of a fraction of Class I cells in the Class II-predominant lines. Undifferentiated Class I lines best resemble the XaXa ES cells in the mouse system – these undifferentiated XaXa cells are the ones that express Tsix RNA in the mouse.

One concern I have regarding figure 2 is that I am not sure it is clear, with the low number of lines, particularly class III lines, what portion of the gene expression differences are culture induced versus X specific. For example, where the two class III

lines simply always passaged together and not frozen as often? The authors should show data from many independent reprogramming experiments, with different fibroblast lines, to show that there is a difference between class II and III. It seems this experiment needs to be done at same passage, exactly same culturing when only comparing few lines.

We hope the Reviewer agrees that it is almost impossible to study all cell lines at the same passage number. We are not aware of any papers in the stem cell field that strictly adhere to this ideal. Apart of the impracticality of doing this, it should also be mentioned that exact passage numbers are not an exact science. A passage is simply defined as a split in culture. It is not standardized to the number of cell divisions. Therefore, two vials of "passage 5" cells may have undergone different numbers of cell divisions; conversely, two vials, one of "passage 3" and the other of "passage 8", may have had the same number of divisions. We think it is more meaningful and practical to separate sublines by "low" versus "high" passage.

Nearly all the hiPSC lines (exceptions: hiPS-2 p.9, hiPS-9 p.7) used for the microarray experiments (Figure 2) were frozen just one time and cultured together at the same time in the same incubator (hiPS-11 LO was cultured at low oxygen at the same time as the other lines). We agree that it would be best to have RNA samples from same passages (as done for hiPS-11 for high vs low oxygen). The basis for selecting samples for microarray submission was RNA quality and yield; some of the RNA samples underwent degradation during storage or isolation, forcing us to select samples from higher passages. Note that, in the end, we used 8 biological replicates of Class II lines and several independent Class III comparators (hiPS-9, -12, L3(cIII)LO, and L3(cIII)Dis; Fig. 3B).

Figures 2A,B,C demonstrate that:

(a) Class III lines (despite different passage numbers) cluster together and far away from both their XIST+ matched line and other (different) XIST+ hiPSCs.

(b) All 10 samples are not very different (color key shows range of R values), yet we see these groupings irrespective of passage number.

Figure 3 evaluates expression patterns of published male, female iPSCs and ESCs. First, there appears to be the same Xist expression differences between male iPSCs and class III iPSCs in 3B? I am rather confused. Then, how do the authors know what the class status of published cell lines is? Xist levels appear to be highly variable, but it is not clear whether this is because the cells are potentially XaXa/XaXi mixes or class II/classIII mixes? I am not sure I get anything out of figure 3. Notably, L3 appears very different from everything else in 3B - so class III cells differ dramatically from each other? The differences between these class III lines appear larger than those to the other lines?

We apologize for the confusion. The explanation is actually rather straightforward and is now stated on p. 17. XIST levels in male hiPSC were lower than in Class II lines but higher than in Class III lines. Presence of low-level XIST in males is consistent with low-level pinpoint XIST expression seen in undifferentiated ES

cells of mice, both male and female, and is one indication that the male lines have been reprogrammed to an ES-like state.

We infer the Class III state from published array studies by the absence of XIST expression in female lines. This is a fairly objective and accurate way of determining the Class III state. Of course, when XIST levels are variable, we cannot infer the relative proportions of Class I, II, versus III cells within the line. In this scenario, we classify the line as "Class II" – which is how we classified all of our other lines with a mixture of Class I, II, and III cells when Class II cells predominate (see Table I).

Figure 3B is actually an informative figure, though we agree that it takes some patience and analysis to take full advantage of this figure panel. As stated in the text and figure legend, L3 is the AVERAGE of the two Class III lines, hiPS-9 and -12. We set the levels for this averaged sample to zero across 30 genes (hence, the black color on the heatmap). All other samples are compared to this average (shades of green and red). The combination of Figures 3A, 3B, and 3C enable us to make the critical statements that Class III cells group differently from other lines in PCA in multiple dimensions (Fig. 3A), that certain oncogenes are upregulated in Class III cells (Fig. 3B), and that there is an excellent anti-correlation between XIST levels and degree of genic upregulation (Fig. 3C). Note excellent Pearson correlations (R).

3. The expression of cancer genes mirrors a cancer-like gene expression (abstract). This conclusion may be an overstatement.

We agree and have revised the statement to say: "We conclude that female hiPSC may be epigenetically less stable in culture and caution that loss of XIST may result in qualitatively less desirable stem cell lines."

4. iPSCs with different X chromosome state differ in cell growth and differentiation, particularly, class III cells.

The last figure suggests that class III lines grow faster in vitro but differentiate less efficiently. I am surprised to see that the doubling time is around 100hours. Does that mean the authors have a lot of cell death in each passage? In 4B - lines with 83 or 29% Xist have many cells without Xist - are those XaXa or XaXi? No data are presented. Thus, it would be better to show exactly what the X status of these lines is and test more lines as I am not sure the differences are significant? The teratoma assay is difficult to perform in a quantitative manner and needs to be done with many class II and III lines. Directed differentiation may be better?

We agree that the doubling times are long, but they are not so different from those published. In Cowan et al. (NEJM 2004), low passage hESC double every ~150 hours!! Later passage lines might have faster doubling times of 24-48 hrs (HUES1,4,6-9) or 60-72 hrs (HUES3,5, 10). For male hiPSC, Takahashi et al. (2007) reported similar doubling times to hESC of 47+/-12 hrs, 43 +/- 12 hrs, 48 +/- 6 hrs. These were rates observed for high O2 levels, which generally give rise to faster doubling times. Thus, our doubling times in Figure 4 are really not so different

from published accounts.

Regardless, using the same conditions on cells grown in parallel, we observed that Class III lines have faster doubling rates. This is true in both norm-oxic and hypoxic conditions. Because the experiments are internally controlled, we feel comfortable with the conclusion that Class III cells grow faster, regardless of what our doubling times are relative to those of other labs.

On the basis of RNA FISH, we know that the XIST^{-} cells are XaXi and represent Class III. We tested this not only in Table I but for the cell growth experiments.

We agree that different lines should be injected in the teratoma assay to make a significant conclusion. This is why we tested 4 different Class II hiPSCs (hiPS-2, -10, -11, -12) and 3 different Class III hiPSC lines (hiPS-2, -12, 6C-1 c.III). All Class II injections produced solid tumors (5/5, plus countless published teratomas), whereas all Class III injections yielded cystic ones (11/11). The numbers are clarified in the new Results section. We now also show two distinct examples in Figure 4D.

We have subsequently also tested 2 different Class II-III transition intermediates (hiPS-9, -12). The Class II-III transition intermediate hiPSC lines yield almost equal numbers of solid and cystic tumors, consistent with original observations for Class II and III teratomas (data not shown).

Cell Stem Cell Conflict of Interest Form

Cell Press, 600 Technology Square, 5th floor, Cambridge, MA 02139

Please complete this form electronically and upload the file with your final submission.

Cell Stem Cell requires all authors to disclose any financial interest that might be construed to influence the results or interpretation of their manuscript.

As a guideline, any affiliation associated with a payment or financial benefit exceeding $10,000 p.a. or 5% ownership of a company or research funding by a company with related interests would constitute a financial interest that must be declared. This policy applies to all submitted research manuscripts and review material.

Examples of statement language include: AUTHOR is an employee and shareholder of COMPANY; AUTHOR is a founder of COMPANY and a member of its scientific advisory board. This work was supported in part by a grant from COMPANY.

Please disclose any such interest below on behalf of all authors of this manuscript.

Please check one of the following:

☒ None of the authors of this manuscript have a financial interest related to this work.

☐ Please print the following Disclosure Statement in the Acknowledgments section:

Please provide the following information:

☒ Please check this box to indicate that you have asked every author of this work to declare any conflicts of interest. Your answers on this form are on behalf of every author of this work.

Manuscript #: CELL-STEM-CELL-D-11-00358

Article Title: Molecular signatures of human induced pluripotent stem cells (hiPSC) highlight sex differences and cancer genes

Author List: Montserrat C. Anguera, Ruslan Sadreyev , Zhaoqing Zhang, Attila Szanto, Bernhard Payer, Steven D. Sheridan, Showming Kwok, Stephen J. Haggarty, Mriganka Sur, Jason Alvarez, Alexander Gimelbrant, Maisam Mitalipova, James E. Kirby, and Jeannie T. Lee

Your Name: Jeannie T Lee

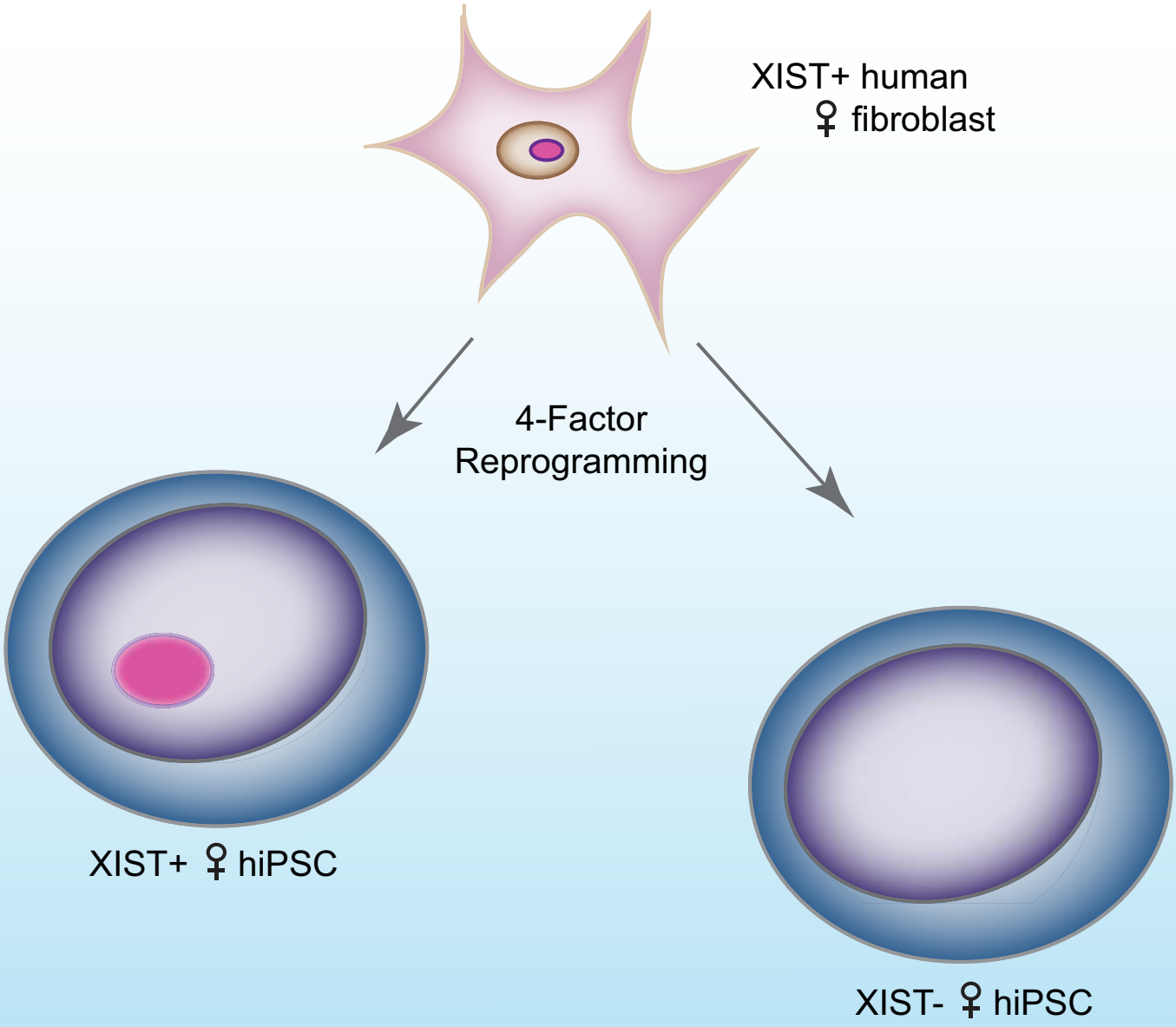
Date: 2/29/12

E-TOC

hiPSCs have enormous potential in regenerative medicine, but their epigenetic variability suggests that some lines may not be suitable. We show that XIST RNA can be used to assess quality of hiPSCs. XIST loss correlates with oncogene upregulation, faster growth rate, and poor cell differentiation.

Graphical Abstract

Click here to download Graphical Abstract: GraphicalAbstract.eps



Cancer gene expression
Hyperproliferation
Poor differentiation

HIGHLIGHTS

- hiPSCs have huge potential in medicine, but they are epigenetically unstable.
- XIST RNA can be used to assess quality of female hiPSCs.
- XIST loss correlates with faster growth rates and poorer cell differentiation.
- Female hiPSCs are more difficult to maintain in culture, resulting in qualitatively less desirable lines.

Molecular signatures of human induced pluripotent stem cells (hiPSCs) highlight sex differences and cancer genes

Montserrat C. Anguera[1,2,3], Ruslan Sadreyev [1,2,3], Zhaoqing Zhang[4], Attila Szanto[1,2,3], Bernhard Payer[1,2,3], Steven D. Sheridan[5,6], Showming Kwok[5], Stephen J. Haggarty[6], Mriganka Sur[5], Jason Alvarez[3,7], Alexander Gimelbrant[3,7], Maisam Mitalipova[8], James E. Kirby[9], and Jeannie T. Lee[1,2,3]

[1]Howard Hughes Medical Institute

[2]Department of Molecular Biology, Massachusetts General Hospital, and

[3]Department of Genetics, Harvard Medical School, Boston, MA 02114, USA.

[4]SAB Biosciences, Qiagen, 6951 Executive Way, Suite 100, Frederick, MD 21703, USA.

[5]Department of Brain and Cognitive Sciences, Picower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[6]Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston MA 02114, USA

[7]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA

[8]Whitehead Institute for Biomedical Sciences, 9 Cambridge Center, Cambridge, MA 02142, USA

[9]Department of Pathology, Beth Israel Deaconess Medical Center, 330 Brookline Ave, Boston, MA 02215, USA.

*Corresponding Author: lee@molbio.mgh.harvard.edu

Running Title: Sex differences in hiPSCs

SUMMARY

Although human induced pluripotent stem cells (hiPSCs) have enormous potential in regenerative medicine, their epigenetic variability suggests that some lines may not be suitable for human therapy. There are currently few benchmarks for assessing quality. Here we show that X-inactivation markers can be used to separate hiPSCs lines into distinct epigenetic classes and that the classes are phenotypically distinct. Loss of XIST expression is strongly correlated with upregulation of X-linked oncogenes, accelerated growth rate *in vitro*, and poorer differentiation *in vivo*. Whereas differences in X-inactivation potential result in epigenetic variability of female hiPSC lines, male hiPSC lines generally resemble each other and do not overexpress the oncogenes. Neither physiological oxygen levels nor HDAC inhibitors offer advantages to culturing female hiPSC lines. We conclude that female hiPSC may be epigenetically less stable in culture and caution that loss of XIST may result in qualitatively less desirable stem cell lines.

INTRODUCTION

With the potential to differentiate into cells of three germ lineages *ex vivo*, human embryonic stem cells (hESCs) hold immense promise for the field of regenerative medicine. Their derivation from the early human embryos has, however, limited the extent to which hESCs can be generated to meet the needs of an immunologically diverse population. A major breakthrough, therefore, has been creation of patient-specific hESC-like cells from somatic cells by reprogramming through defined pluripotency factors, OCT4, KLF4, c-MYC, and SOX2 (OSKM) (Takahashi and Yamanaka, 2006; Yu et al., 2007). These "human induced pluripotent stem cells" (hiPSCs) share similar gene expression profiles, morphologies, and differentiation potential with hESCs (Wernig et al., 2007; Maherali et al., 2008), and mouse-derived iPSCs can be passaged through the germline (Okita et al., 2007). While hiPSCs solve major ethical issues, recent studies have revealed that they may be as genetically and epigenetically fluid as hESCs (Kim et al., 2010; Bock et al., 2011; Gore et al., 2011). There may also be greater expression anomalies in hiPSCs (Rugg-Gunn et al., 2005; Adewumi et al., 2007; Rugg-Gunn et al., 2007; Pick et al., 2009). Because mutation and epigenetic change can lead to cancer and other diseases, these observations imply that some hiPSCs lines may not be suitable in a clinical setting. However, apart from karyotype and a limited panel of differentiation markers, there are currently few established benchmarks for assessing hiPSCs quality and suitability.

Interestingly, unlike mouse embryonic stem cells (mESCs), hESCs vary tremendously in their potential to undergo X-chromosome inactivation (XCI) (Hoffman et al., 2005; Adewumi et al., 2007; Hall et al., 2008; Shen et al., 2008; Silva et al., 2008; Dvash et al., 2010; Lengner et al., 2010), an epigenetic event that is tightly coupled to cell differentiation both *in vivo* during epiblast differentiation and *ex vivo* in cultured

embryonic stem cells (Payer and Lee, 2008). During XCI, one of two female X-chromosomes is transcriptionally repressed to achieve similar X-linked gene dosage as males (Payer and Lee, 2008). This process depends on expression of the long non-coding Xist RNA, which is upregulated just prior to the initiation of chromosome-wide silencing (Marahrens et al., 1997; Wutz and Jaenisch, 2000) and results in recruitment of repressive chromatin to the X. Whereas all XX female mESCs faithfully recapitulate XCI in culture, female hESCs have been grouped into three classes based on differences in their ability to do so: Class I lines initially carry two active Xs (XaXa) but can upregulate XIST and undergo XCI during cell differentiation, suggesting that they most closely approximate the mESCs ideal. Class II lines already possess one inactive X (XaXi) and may therefore be partially differentiated. Class III lines largely already underwent XCI but subsequently lost XIST expression, raising questions about their epigenetic stability (Silva et al., 2008; Dvash et al., 2010). Whether these epigenetic classes themselves have practical implications remains unclear. However, because the XCI phenotype may correlate with differentiation potential, XIST has been proposed as a benchmark for assessing hESCs quality (Silva et al., 2008).

Indeed, use of the XIST marker led to identification of more XaXa hESCs at early passage (Dvash et al., 2010), discovery that physiological oxygen concentrations are preferable for deriving Class I cell lines,, and demonstration that stressful *ex vivo* conditions are associated with conversion to the Class III epigenotype (Lengner et al., 2010). These observations have lately generated much interest in the X-chromosome status of hiPSCs and raised the question of whether XIST could also be used as a benchmark for hiPSCs quality. In the mouse system, reprogramming of somatic cells to iPSC is accompanied by X-reactivation (Maherali et al., 2007). By contrast, recent studies in the human system have reported varying results, with some demonstrating

that reprogramming does not reactivate the Xi of parental fibroblasts (Tchieu et al., 2010; Cheung et al., 2011), and others observing that some hiPSCs lines have reactivated Xi (Lagarkova et al., 2010; Marchetto et al., 2010). Unresolved, therefore, is whether hiPSCs ever attain the XaXa state associated with pluripotency in the mouse system and in Class I hESCs. Also unclear is whether X-chromosome states can be used as a readout for female hiPSCs quality. Below, we investigate the XCI status and implications of XCI differences in female hiPSCs. We report genomewide signatures associated with loss of XIST expression and demonstrate sex-specific differences, the combination of which caution that female hiPSCs may be inherently more difficult to maintain using existing protocols.

RESULTS

X reactivation and inactivation in female hiPSCs

Given contrasting reports on X-reactivation during establishment of hiPSCs lines, we revisited this question by following the X-transcriptional status of new female hiPSCs lines derived from IMR-90, a diploid human fetal fibroblast line that has been used extensively to generate hiPSCs (Yu et al., 2007). Cells were reprogrammed using virally expressed OSKM (Park et al., 2008a) and colonies picked between days 28-32 (Fig. S1A). Immunostaining (Fig. S1B) and qRT-PCR (Fig. S1C) showed expression of pluripotency markers; bisulfite sequencing showed appropriate demethylation of endogenous OCT4 and NANOG promoters (Fig. S1D); and qRT-PCR demonstrated silencing of viral factors (Fig. S1E). The hiPSCs could differentiate into three germ lineages (Fig. S2A) and form teratomas in NOD-SCID mice (Fig. S2B). Furthermore, differentiation induced expression of lineage-specific markers (Fig. S2C) and karyotypes confirmed a 46XX constitution (Fig. S3). These data demonstrate successful generation of new female hiPSC lines. For XIST analysis, we used earliest possible passages (p. 0-7) to circumvent potential problems associated with long-term culture.

To examine XCI status, we performed RNA FISH and observed one XIST cloud in 58-84% of nuclei immediately following reprogramming (p.0) (Fig. 1A,B, Table I). These 14 distinct clones were derived from five different individuals, including IMR-90 (46XX), a 47XXY cell line, a 46XX 'MM' line, and two lines from MM's twin daughters ('TA' and 'TB'). An H3K27me3 domain indicative of XIST-mediated Polycomb recruitment was also present (Fig. 1C). These findings demonstrate XCI in a majority of cells in each line. Allele-specific analyses of gene expression using SNP arrays showed monoallelic expression of X-linked genes for two hiPSCs lines (Fig. S4), consistent with

the absence of X-reactivation following reprogramming (Tchieu et al., 2010; Cheung et al., 2011).

However, we also observed absence of XIST in 15-40% of cells, raising several possibilities with respect to XCI. First, $XIST^-$ cells could represent cells that underwent X-reactivation and attained the XaXa state of pluripotent stem cells. In this scenario, reprogramming would be accompanied by X-reactivation, and XaXi cells might represent those that spontaneously re-inactivated one X, as is often observed for hESCs (Higgins et al., 2007; Dvash et al., 2010; Lengner et al., 2010). Alternatively, the large number of $XIST^+$ cells might indicate that X-reactivation never occurred during reprogramming and resulting hiPSCs merely retained the Xi of parental cells, as proposed by two previous studies (Tchieu et al., 2010; Cheung et al., 2011). In this scenario, $XIST^-$ cells would represent spontaneous loss of XIST expression characteristic of Class III hESCs (Silva et al., 2008; Dvash et al., 2010).

To distinguish these possibilities, we performed serial RNA-DNA FISH. We first carried out two-color RNA FISH on undenatured nuclei to visualize XIST and Cot-1 expression. The Cot-1 staining pattern provides an overview of nascent transcription from a nuclear domain (Hall et al., 2002; Clemson et al., 2006; Namekawa et al., 2010). Following RNA FISH, we denatured the samples and performed DNA FISH using X-painting probes to locate the X-chromosomes (Fig. 1B). Two types of $XIST^-$ cells were observed in all hiPSCs, irrespective of reprogramming method (e.g., with or without VPA), passage number, and genetic background (Table I).

One type of $XIST^-$ cells showed two $Cot\text{-}1^+$ X-chromosomes, implying active transcription of both Xs (Class I). This inferred XaXa state suggests occurrence of X-reactivation during reprogramming. To determine whether percentages of $XIST^+$ cells

must increased during cell differentiation, we placed hiPSCs in differentiation conditions for 14-50 days. Indeed, XIST expression increased (Fig. 1D, Table I; e.g., hiPS-2, -11, -12), as would be expected of differentiating XaXa cells. We suggest that hiPS-2, -9, -10, and -11 contain a small fraction of Class I cells mixed with Class II and III cells. Because Class I cells accounted for only 2-14% of cells (Table I), biallelic expression from this subpopulation would not have been discernible by allele-specific SNP analysis (Fig. S4).

The second type of XIST$^-$ cells displayed one Cot-1$^-$ X-chromosome, indicating X-chromosome repression in spite of being XIST$^-$ (Class III state)(Fig. 1B). This Class III phenotype resembled spontaneous conversion to a Class III phenotype in hESCs (Silva et al., 2008; Dvash et al., 2010). Initially, Class III cells comprised less than one-third of each hiPSCs line (Table I). During routine culture, three of the Class II-predominant hiPSCs lines evolved to 100% Class III (Table I: hiPS-2, hiPS-9, and hiPS-12). In the three sub-lines, XIST expression was absent before and after differentiation, and one X-chromosome lay within a Cot-1 hole (Fig. 1B,F). Two previously published female hiPSCs lines, HD 12D-1 and JDM 6C-1, derived from patients with Huntington's Disease and Type I diabetes mellitus, respectively (Park et al., 2008a), were also Class III, with 0% XIST expression before and after differentiation (Fig. 1E). Examination of nascent transcription from X-linked *PGK1* indicated that, of XIST$^-$ cells with detectable PGK1 signal, approximately half showed biallelic PGK1 and half showed monoallelic expression in hiPS-9, -10, and -11 lines on d0 (data not shown), consistent with the idea that the XIST$^-$ subpopulation is a Class I-III mixture.

Thus, hiPSCs and hESCs share the tendency to lose XIST expression in culture. Once lost, XIST expression was never regained (data not shown), though a Cot-1 hole indicative of repetitive element silencing persisted. These data show that our female hiPSC lines consist of a mixture of Class I, II, and III cells. The presence of XaXa cells

(Class I) argues that X-reactivation takes place in a fraction of cells during reprogramming. The XaXi cells (Class II) indicates either that XaXa cells spontaneously undergo re-inactivation of one X or that a fraction of hiPSCs never underwent X-reactivation. Although the hiPSC lines have a mixed population, Class II cells dominate at early passage. The tendency to become Class III during culture demonstrates a level of epigenetic fluidity characteristic of female hESCs.

Effects of oxygen and HDAC inhibitors

Previous work showed that physiological (4%) oxygen instead of ambient (20%) levels preserves the Class I state of hESCs (Lengner et al., 2010) and enhances reprogramming to iPSCs (Utikal et al., 2009; Yoshida et al., 2009). We investigated whether physiological oxygen might be beneficial for hiPSCs. Here we used a fibroblast line, derived from an 18-week-old 47,XXY fetus. It yielded 39 colonies at 20% oxygen, in contrast to IMR-90, which typically yielded 1-6 colonies from 50,000 starting cells. Furthermore, when reprogrammed at 4% oxygen, the XXY line produced twice as many colonies (~80) (Utikal et al., 2009; Yoshida et al., 2009). The XXY fibroblasts behaved similarly to 46XX cells with respect to XCI, as XIST RNA was expressed from a single X (Poplinski et al., 2010). We expanded four clones reprogrammed in ambient oxygen (hiPS-XXY-H1, -H2, -H3, -H5) and three in physiological oxygen (hiPS-XXY-L1, -L3, -L4). Immunostaining and qRT-PCR confirmed expression of pluripotency markers in all 7; each also demonstrated EB formation and outgrowth during differentiation (Fig. S5A and data not shown). In general, hiPSCs maintained in 4% oxygen better preserved their morphology and showed less oxidative stress (Fig. S5B) (Prigione et al., 2010). XIST RNA FISH showed that reprogramming at 4% oxygen had no effect on XIST expression

(Table I), as the XXY lines remained predominantly Class II, with 66-89% expressing XIST on d0 and 84-94% after differentiation (Table I, Fig. 1F). The Class I subpopulation was invariably low (2.8-7%). Class III cells were also present in each isolate. Thus, oxygen levels have no major effect on XIST in hiPSCs.

Previous work also showed that HDAC inhibitors promote a more favorable epigenetic state for hESCs. Specifically, treating H9 containing a mixture of XaXa/XaXi cells resulted in a homogeneous XaXa population capable of upregulating XIST upon differentiation (Ware et al., 2009). To determine if the effects extended to hiPSCs, we treated hiPSCs with HDAC inhibitors, sodium butyrate and vorinostat, for 5-8 passages and examined XIST during differentiation. HDAC inhibition did not change XIST profiles from d0-d18 in any of six lines (Table S-II). All Class II-predominant lines continued to show XIST clouds in 40-70% of cells, and three Class III lines from three distinct individuals showed no rescue of XIST expression. Therefore, HDAC inhibition has no obvious beneficial effect for XIST in hiPSCs.

To determine if the effect might be specific to hESCs, we treated two Class I-predominant hESCs lines, HUES-9 and H9 (Silva et al., 2008; Ware et al., 2009), for 5 passages and examined XIST (Table S-I). Consistent with previous analysis (Ware et al., 2009), HDAC inhibition increased the number of XaXa cells on d0, and yielded cells with XIST clouds after differentiation. Treatment of HUES-9 resulted in a modest increase of $XIST^{+}$ cells during differentiation, but did not increase the number of XaXa cells on d0, as observed with H9. In our hands, recovery after cryopreservation, general growth, and morphology of both hESCs and hiPSCs were enhanced, consistent with the previous report (Ware et al., 2009). We conclude that HDAC inhibitors do not improve XIST profiles for hiPSCs, but may better rescue XIST in female hESCs (Diaz Perez et al., 2012).

Genomewide transcription profiling reveals class-specific differences

Whether the Class III state has significant biological consequences is currently unknown. To address this question, we compared genomewide expression profiles by microarray analysis of 10 hiPSCs lines and sublines, all of which were derived from IMR-90. Hierarchical clustering revealed that all Class II-predominant cell lines showed highly correlated expression patterns among each other (Fig. 2A). Furthermore, Class III lines were strongly correlated with each other. Intriguingly, Class III sublines of hiPS-9 and hiPS-12 resembled each other more than they resembled Class II parents and other Class II lines. Departures from their parental lines were more dramatic than differences for hiPSCs grown in 20% versus 4% oxygen (hiPS-11 vs hiPS-11LO2).

Principal component analysis (PCA) supported these observations (Fig. 2B,C). In multiple dimensions, Class III hiPS-9 and hiPS-12 sublines were significantly closer to each other than to parental Class II hiPS-9 and hiPS-12 counterparts and to all other Class II lines. Interestingly, the hiPS-1 profile was closer to Class III than to other Class II. This correlated with the larger subpopulation of Class III cells within hiPS-1 (33%, Table I). hiPS-1 may be in transition to Class III. Thus, loss of XIST expression is associated with significant shifts in global expression profiles, suggesting that the Class III state is a distinct epigenotype that develops during culture.

Class III association with upregulation of cancer-related genes

To determine what genes were affected, we looked for class-specific differences in gene expression. We used ANOVA-based estimates of statistical significance with

conservative modeling of gene-specific inter-sample variance implemented in NIA Array Analysis webtool (Sharov et al., 2005). Among genes showing significant differential expression (FDR < 0.05), only 10 coding genes were consistently upregulated more than 2-fold in Class III compared to all Class II lines (Table II-A; Fig. S6). Interestingly, among the genes upregulated in Class III hiPSCs lines, X-linked genes were significantly over-represented (4 out of 10 genes, $P = 7x10^{-5}$). This caught our attention, given that loss of *Xist* has been shown to result in partial X-reactivation in murine cells (Csankovszki et al., 2001; Zhang et al., 2007) and overexpression of X-genes has been correlated with cancer (Richardson et al., 2006; Pageau et al., 2007).

Two of the upregulated X-linked genes, *MAGEA2* and *MAGEA6*, are highly expressed in cancers (Rogner et al., 1995). Overexpression of five others has also been implicated in cancer and metastasis, including *RAB6B*, a member of the *RAS* oncogene family; *CHP2* in ovarian tumors; *ACP5* in various cancers; and *AIF1* in breast tumor growth. *TCEAL3*, *LOC100131199*, and *LOC285965* have no known function. Thus, at least 6 of 10 upregulated coding genes specific to Class III lines are previously identified cancer genes.

We then asked which genes were consistently downregulated by at least 2-fold in Class III cells compared to all Class II samples (Table II-B; Fig. S6). X-linked genes were not over-represented in this list, as might be expected since XIST is an X-silencer. Apart from *XIST*, the only other X-linked locus in the top hits list was *FTX*, a noncoding gene near *XIST* with undefined function (Chureau et al., 2002). Other downregulated genes of interest were known tumor suppressors, including *FN1*, a fibronectin involved in cell adhesion. Noncoding RNAs MALAT1 and NEAT1 (of nuclear speckles and paraspeckles associated with cancers) were also downregulated (Ji et al., 2003; Sunwoo et al., 2009).

Taken together, the genome-wide profiles argue for class-specific associations with cancer genes and raises the question of how many of the Class III changes could be attributed to or strongly correlated with loss of XIST. To address this, we identified genes whose expression levels had highest Pearson correlation coefficients with XIST levels across all samples. We included genes that were upregulated in all Class III lines compared to at least 6 of 8 Class II lines. Several hundred met these criteria (Table S-II for complete list), of which 30 with greatest correlation are shown in Table III-A. Intersecting the list of XIST-correlated genes with known cancer genes from MSKCC CancerGenes resource (Higgins et al., 2007) revealed 9 tumor suppressors (*CDC14B, CDK6, CNOT7, IDH1, IGFBP5, PCDH10, PLXNC1, RBBP4, STK4*) and 7 oncogenes (*BCL11A, CHD1L, FGFR1, FUS, FYN, RAB12, SOS1*) that were differentially expressed between Class II and Class III. Genes with highest correlation with XIST include *SEMA6A*, *MALAT1*, and *FTX*, and genes for oxidative stress response, *COX1* (R = 0.94) and *PRDX2* (R = 0.838)(Fig. 2D, Table III-A). Also highly correlated were members of the Mediator complex (*MED6, MED17*), a transcriptional co-activator complex found at promoters of active genes in pluripotent cells; *MLL2*, a histone 3 lysine-4 (H3K4) methyltransferase responsible for bulk methylation of H3K4me3 associated with transcriptional activation. Using DAVID Bioinformatics Resource (Huang da et al., 2009), we observed significant enrichment for genes involved in transcription (FDR = 0.001), transcriptional repression (FDR = 0.0017), and transcriptional regulation (FDR = 6.02x10-4). DAVID analysis of Table S-II also confirmed enrichment for many genes involved in RNA processing (FDR = 0.006), splicing (FDR = 0.054), binding (FDR = 0.034) stability and export (e.g., *FUS, HNRNPA1, SFPQ, HNRNPD, SFRS15, SFRS4*).

The same analysis was applied to genes whose expression levels had the greatest anti-correlation (negative Pearson correlation coefficients) with XIST levels

across all samples (Table III-B; complete gene list in Table S-III), of which 12 with highest negative Pearson correlation coefficients are plotted in Figure 2E. There was considerable overlap between Tables II-A and III-B, and X-linked genes were significantly over-represented ($P = 2x10^{-8}$, for 9 of 21 genes being X-linked in Table III-B). Again, cancer genes were also highly represented. In addition to those in Table II, *CSAG2, NUCKS1, REPS2, MTA2*, *RAB6B, RAP2C, and VAV1* showed anti-correlation with *XIST*. Oncogenes *MAGE2A* (R = -0.980) and *MAGE6A* (R = -0.944) showed especially high correlation. DAVID analysis of Table S-III yielded no significant enrichment for any group of genes. Notably, oncogenes as a general class were not significantly enriched. Taken together, these argue for enriched expression only of oncogenes residing on the X, due to loss of XIST-mediated suppression *in cis*. We conclude that loss of XIST expression is strongly correlated with X-gene over-expression, hyper-expression of select X-linked oncogenes, and with repression of select tumor suppressors.

Male and female differences in hiPSC quality

Because male cells do not undergo XCI, male hiPSCs cannot be subclassified by XIST expression. However, the strong genomewide positive and negative correlations identified above for female hiPSCs might be used in lieu of XIST to address male hiPSC quality. Could male hiPSCs be subcategorized on the basis of genomewide expression profiles? Do some male hiPSCs exhibit aberrant expression of cancer genes? To address these questions, we analyzed gene expression profiles of published male and female hiPSCs derived from normal fibroblasts by reprogramming with either virally introduced factors, modified RNA, or direct protein delivery (Maherali et al., 2008; Park

et al., 2008b; Kim et al., 2009; Jia et al., 2010; Mayshar et al., 2010; Tchieu et al., 2010; Warren et al., 2010). We also queried whether variability occurred in hESCs lines (Westfall et al., 2008; Chen et al., 2011), and compared male hiPSCs and female hESCs profiles to our female hiPSCs, including a low-oxygen line (hiPS-2 cIII LO) and two disease-model hiPSC lines (12D-1, 6C-1) created elsewhere but passaged in our laboratory (Park et al., 2008a). Our diverse sampling therefore tested cell lines of distinct provenance, with fibroblasts derived from multiple individuals and hiPSCs created in 12 different labs.

We first performed hierarchical clustering and PCA loading analyses. Because hiPSC lines are known to have a tendency to cluster by laboratory of origin (Guenther et al., 2010; Laurent et al., 2011) and because variations could arise from biases between microarray batches ('batch effects'), we analyzed the data without *in silico* correction of batch effects or with correction using the ComBat method (Johnson et al., 2007)(Fig. 3A,B). Several patterns emerged using either method. First, there is a tendency for each type of cell line to cluster together, irrespective of lab origin. For example, female hESCs clustered together (black), as did female hiPSCs (pink), and male hiPSCs (blue). Furthermore, Class III female lines (green) grouped together but away from female hiPSCs and hESCs. In general, female hiPSCs lines showed greater variation among one another than did male hiPSC lines among themselves (blue male lines versus pink female lines; Fig. 3A,B). Interestingly, while male hiPSCs tend to group together apart from female hiPSCs, RNA-reprogrammed male lines (R4, R5) (Warren et al., 2010) appeared to better resemble female hESCs and hiPSCs. The secondarily reprogrammed male hiPSCs line (H4-2) was also set apart from other male hiPSCs (most evident with ComBat correction; Fig. 3A). These differences pertained to cell lines derived not only in

different laboratories but also within any given laboratory, as evidenced by loose groupings observed in multiple PCA dimensions (Fig. 3A,B).

Given a strong positive association between XIST loss and overexpression of select X-linked oncogenes (Tables II, III, S-III), we next asked whether molecular signatures of male hiPSCs could be compared against those of female hiPSCs to infer stem cell quality. For the genes differentiatlly expressed in Class III vs II (Table II), we evaluated expression profiles in female hiPSCs, male hiPSCs, and female hESCs, and compared them to the average of Class III lines, hiPS-9 and -12 [Fig. 3C: L3 (cIII)], as the basis for comparison. As expected, all three Class III lines (green) showed low XIST expression, whereas Class II lines of various provenance (pink) showed significantly more XIST expression. The dark-red values for male hiPSCs (blue) were consistent with low-level XIST expression known to occur in male mouse embryonic stem cells (Payer et al., 2011), consistent with their successful reprogramming.

In general, male hiPSCs expression profiles resembled those of Class II female hiPSCs (except XIST levels were lower than in Class II lines but higher than in Class III lines, consistent with pinpoint XIST expression in undifferentiated ES cells of mice) (Fig. 3C). The male profiles, however, significantly deviated from those of the hiPS-9/-12 Class III average. For example, male hiPSCs lines generally did not show increased expression of the oncogenes upregulated in Class III lines (e.g., *MAGEA2, MAGEA6, RAB6B*, *TCEAL3*, and *ACP5*). Main exceptions were male D6(3) and D6(32), which displayed increased *MAGEA2* and *MAGEA6* expression, and the secondarily reprogrammed male lines (shown as an average, H4-2º)), which showed increased expression of many genes upregulated in Class III lines (e.g., *TCEAL3*, *ACP5*, *CHP2*, *RAB6B*; Table III). By contrast, the additional Class III female lines exhibited a trend towards greater expression of the most correlated marker genes from Table II. For

example, L3(cIII)LO (hiPS-11 grown in low oxygen) and L3(cIII)Dis (Disease lines, hiPS 6C-1 and 12D-1) had similarly increased expression of *TCEAL3*, *RAB6B*, *LOC285965*, and *CHP2*.

Even among Class II female lines, casual examination hinted at a correlation between degree of XIST expression and likeness to the Class III profile. For instance, hiPS-1 (L3-1, Fig. 3C), shown above to be a Class II-III intermediate (Fig. 2B,C), resembled the Class III profile (Fig. 3C). This suspicion was confirmed by direct quantitative analysis of profile similarities calculated as Pearson correlation coefficients of expression values on the set of genes differentially expressed in Class III (Table II), excluding XIST itself. This analysis revealed a trend of monotonic decrease with increasing level of XIST expression (Fig. 3D). Two loose groupings of female hiPSCs were apparent. Cell lines with highest XIST expression occupied the bottom right region of the plot, demonstrating the highest dissimilarity to Class III. Those with intermediate XIST expression were located in the center (e.g., hiPS-1 [a.k.a. L3-1]), demonstrating a drift towards the Class III reference in the top-left corner.

Taken together, these results argue that upregulation of X-linked oncogenes and other loci revealed in Tables II and III is a property of female hiPSCs when they lose XIST expression. Though not generally a feature of male hiPSCs, secondarily reprogrammed male lines may more closely resemble Class III female lines. We believe that expression differences in Class III lines are due to epigenetic change rather than to genomic alterations, as microarray-based comparative genomic hybridization (CGH) on paired sets of Class II and III lines revealed no gross copy number changes (Fig. S7). No deletions within the *XIST* locus were observed in each case. Thus, we demonstrate that hiPSCs of both male and female origin could be evaluated by comparison to the deviant Class III profiles.

Loss of XIST results in accelerated growth *in vitro* and poor differentiation *in vivo*

Here we tested whether resemblance to Class III has functional consequences. In light of increased oncogene expression, we asked whether Class III lines grow faster in culture. We measured growth rates of multiple undifferentiated Class II and III cells in four independent experiments over 20-35 days, and plotted numbers of cells (Fig. 4A,B) and colonies (Fig. 4C). In multiple replicates, Class III hiPSCs lines generally exhibited a shorter doubling time than their Class II parents and other Class II lines (Fig. 4A,B). They also grew more quickly than male hiPSCs. This was the case in high and physiological oxygen conditions. Interestingly, hiPS-1 – the Class II-III transitional cell line – exhibited a growth rate more similar to Class III cells (Fig. 4A), thus correlating with its Class III-like expression profiles (Fig. 3B,C). Other transitional lines (identified by fewer XIST+ clouds) also displayed faster growth rates than their Class II parents (compare hiPS-12 p.28 to hiPS-12 p.32). Faster growth rates did not appear to be a consequence of adaptation to culture, as we tested Class II cell lines (hiPS-2, -11, -12) at early (p.14) and later passages (p.32) and found no consistent significant change in growth rates (Fig. 4B). Interestingly, we also observed that Class III lines recovered faster after routine passaging, as these lines typically yielded greater colony numbers and larger colony sizes 1 day after passaging when compared to Class II lines and male hiPSCs.

We next investigated the ability of Class III cells to form teratomas. In general, hiPSCs are known to form teratomas when injected into nude mice. Although both Class II and III lines could do so, their *in vivo* differentiation capacities were markedly different (Fig. 4D,E). Class II teratomas showed prominent differentiation into structures

recapitulating adult organs and tissues, such as cartilage and small intestine, including a range of cell types found in mature intestine including mucin-producing epithelial cells and Paneth cells, secondary organization into villi, and investing layers of circular and longitudinal smooth muscle (Fig. 4D,E; S2-A). Intriguingly, all teratomas derived from four representative Class II sublines, hiPS-2, -10, -11 and -12, formed solid masses (5 of 5); by contrast, teratomas derived from the matched Class III sublines, hiPS-2 and -12, and disease model line, 6C-1, were all cystic (11 of 11), with the cysts lined by simple epithelia and undifferentiated mesenchymal tissue, with little to no differentiated cell types (Fig. 4D,E). Two of the 11 cystic teratomas had small solid masses with a low degree of differentiation into all three germ layers. Notably, one prior report found that male hESCs formed solid teratomas but one female hESCs line with unknown XIST expression status produced cystic teratomas (Mikkola et al., 2006). Our observed differences between matched Class II-III lines argue that Class III cells may generally form poorly differentiated teratomas of cystic nature. The poor differentiation is consistent with a cancer-like state. On the basis of these observations, we suggest that XCI class designations of female hiPSCs may have practical implications for stem cell therapy.

DISCUSSION

Here, we have studied genomewide expression profiles of multiple new and existing hiPSCs lines and shown that the XCI marker, XIST RNA, can be used as a readout to assess one aspect of female hiPSCs quality. The gene expression profiles have identified molecular signatures that distinguish $XIST^{+}$ (Class II) and $XIST^{-}$ (Class III) female hiPSCs lines. Loss of XIST expression in Class III cells is associated with upregulation of oncogenes, several of which are X-linked, and downregulation of several

tumor suppressors. We do not know whether loss of XIST is directly responsible for these expression differences. An alternative possibility is that conditions which lead to loss of XIST expression cause other changes genome-wide. In either case, we presume that these changes are generally undesirable and can therefore be used as additional benchmarks of hiPSCs quality. Indeed, the Class III changes correlate with faster growth in culture. Notably, these changes are not generally observed for male hiPSCs lines. We also observed differences in differentiation *in vivo*, as shown by formation of predominantly cystic, poorly differentiated teratomas in nude mice.

These data argue for class- and sex-specific differences in epigenetic stability of hiPSCs that depend in large part on the ability to maintain XCI. One major implication is that the epigenetic state of female hiPSCs may be more difficult to maintain in culture, at least using current protocols. Neither physiological oxygen nor HDAC inhibitors offered any advantage nor more efficient X-reactivation. Several recent works suggest that hiPSCs are not equivalent to the more extensively reprogrammed 'naïve' female hiPSCs, which apparently contain two Xa and may therefore represent the best model for X-reactivation in hiPSCs (Hanna et al., 2010; Pomp et al., 2011; Wang et al., 2011). However, it is not clear whether these naïve hiPSCs contain a pure population of Class I cells or rather a mixture of Class I-III cells. The epigenetic stability of XIST after extended culture is also uncertain. Better protocols are needed in order to avoid the potentially unfavorable genomewide changes seen in many female hiPSCs lines.

Another major implication may be that Class III female hiPSCs lines are best avoided for *in vivo* human therapy due to (a) the upregulation of some X-linked cancer genes, (b) faster-than-normal growth rates *ex vivo*, and (c) poor differentiation potential *in vivo*. Some hiPSCs lines may evolve into the Class III state more readily than others, perhaps because of underlying genetic and copy number variation between parental cell

lines, or the number of viral OSKM integrations. Although we do not know whether upregulation of X-linked and other oncogenes is a direct consequence of XIST repression, we surmise that the absence of XIST in Class III lines may promote reactivation of undesirable X-linked genes, given recent work showing that conditionally deleting *Xist* on Xi of mouse somatic cells and loss of XIST in hESCs can result in piecemeal X-reactivation (Csankovszki et al., 2001; Shen et al., 2008; Diaz Perez et al., 2012). If X-reactivation occurs in these lines, they do not occur on all X-genes at once (one X still resides in a Cot-1 hole). Nevertheless, the possibility of general X-reactivation over time should present significant concern and urge caution in using some female hiPSCs lines in cell regeneration programs. We therefore encourage the use of XCI markers as a benchmark to assess quality of all female hiPSCs and, by inference, hESCs lines. Going forward, we suggest that, XIST expression in combination with differentiation potential be used to assess stem cell quality.

METHODS

hiPSC culture and derivation

IMR-90 fibroblast line (ATCC; CCL-186) was cultured in EMEM medium with 10% FBS and XXY line (Corielle GM03102) with 15% FBS. Human iPSCs were maintained on irradiated MEFs with hES medium (DMEM/F12, 10% knockout serum replacement (Invitrogen), L-glutamine, non-essential amino acids, 2-mercaptoethanol, penicillin/streptomycin, and 10 ug bFGF). For hiPSC derivation, 10^5 fibroblasts were infected with retrovirus (pEYK cassette with 4F) (Park et al., 2008a) at ambient oxygen, then 48 hrs later transferred to MEF-coated plates at either ambient or 4% oxygen. MM, TA, and TB fibroblasts were reprogrammed using the tet-inducible lentiviral STEMCCA and rtTA (without MEFs, no VPA during reprogramming or ROCK inihibitor). For hiPS-1,2,3,9,10,11,and 12: cells were treated with 1mM VPA for 7 days (10d for XXY lines), and colonies were picked one month after infection. Use of VPA did not impact occurrence of Class I, II, or III cells, as MM, TB, TA lines were reprogrammed without it. ROCK inhibitor Y-27632 (Calbiochem) was used for the first 2 days during the first 2 passages and after thawing cells. All hiPSCs were passaged manually. For HDACi treatment, sodium butyrate (0.1mM; Sigma) and vorinostat (400nM; Cayman) were freshly diluted and added daily. For cell growth experiments, colonies were split using cell rollers (Invitrogen) and 1 colony (about 10 clumps) or 3 colonies for each line were transferred to 1 well of a 96-well plate (1 well of 12-well plate for 3 colonies). Cells were plated in triplicate. After d7, duplicate wells containing 3-5 colonies were transferred to 2 wells (12-well plate). Cells were harvested weekly. Half the culture was counted and the other half passaged onto MEF-coated wells for 4-6 passages. Colony number determined by counting undifferentiated colonies at each passage.

In vitro and *in vivo* differentiation of hiPSCs

Human iPSC colonies were dislodged with cell scraper and transferred to low attachment 6-well plates containing hES differentiation media (hiPS media without b-FGF with 20% FBS). EBs were transferred to gelatin-coated plates (d7) and cultured for additional 8-14 days. For EB germ lineage testing, hiPSCs were dispersed then grown in ultra-low attachment 6-well plates (Nunc) in hES media without bFGF supplemented with 1% FBS for d19. EBs were fixed in PBS with 4% paraformaldehyde (PFA), pelleted in low-melt agarose, paraffin-embedded, sectioned (5 µm), then stained with H&E. For teratoma injections, 1-2 10cm plates of confluent hiPSCs (no MEFs), pelleted and mixed with equal volume of 2X Matrigel (200uL/injection). Tumors appeared 6-12 weeks after injection, dissected and fixed overnight with 4% PFA, then sectioned and stained with H&E.

Microarray Experiments

Total RNA isolated with Trizol and converted to cDNA using NuGEN Ovation V2 Amplification system. cDNA was hybridized to Affymetrix Human Genome U133 Plus 2.0 Arrays (Microarray Core Facility, Dana Farber Cancer Institute). Samples: hiPS-1 p.31, hiPS-2 p.9, hiPS-3 p.14, hiPS-9 p.7, hiPS-10 p.24, hiPS-11 p.16, hiPS-11 p.16 (4% O2), hiPS-12 p.23 (4% O2), hiPS-9 p.17 (XIST-), and hiPS-12 p.30 (XIST-).

Microarray expression data were normalized by RMA (Irizarry et al., 2003). Hierarchical clustering and PCA were performed on the total sets of RMA expression values. MAS5 (Liu et al., 2002) and MBEI (Li and Hung Wong, 2001) normalization produced similar

results. For the analysis of differential expression, we used the estimates of false discovery rate (FDR; ANOVA-based) with variance adjustment, implemented in NIA Array Analysis (Sharov et al., 2005). The tool was used with default parameters, except for Z-score threshold for outliers set to 10000. The definition of differentially expressed genes was based on the combination of high statistical significance (FDR cutoff 0.05) and the magnitude of expression change. When overall expression was consistent between samples and the two compared groups corresponded to tight clusters, differentially expressed transcripts were defined with FDR < 0.05 and ≥2-fold change between group expression means. In the absence of tight clustering (Class II female samples), differentially expressed transcripts were defined based on FDR cutoffs (FDR < 0.05) and N (individual samples that consistently showed at least 2-fold expression change compared to other group). For the 8 samples of Class II, we used the strict cutoff of N=8 (all samples) and a more relaxed cutoff of N=6. Correlation with XIST expression was measured by Pearson correlation coefficient calculated for the expression values. ComBat method for the compensation of microarray batch effects (Johnson et al., 2007) was run with default parameters on RMA expression values in the set of Affymetrix microarrays for different cell lines. DAVID functional annotation tool (Huang da et al., 2009) was run online on the extended sets of differentially expressed genes.

See Supplemental Information for RNA/DNA FISH, immunostaining, real-time PCR

Bisulfite sequencing, allele-specific expression analyses, and CGH analysis.

GEO accession numbers: Deposition in progress. Information will be forwarded shortly.

ACKNOWLEDGEMENTS

We thank members of the Lee laboratory for critical insight and discussion; G. Daley for pEYK, HD 12D-1, and JDM 6C-1; G. Mostoslovasky for STEMCCA lentiviral cassette; H. Willard and J. Lawrence for PGK1 probe; L. Daheron, T. Ahfeldt, and R. Alagappan for technical assistance; and C. Ware for advice on HDACi. This work was funded by NIH RO1-GM58839 and ARRA supplement to J.T.L; and NIH R21MH087896 and a Stanley Medical Research Institute award to S.J.H. J.T.L. is an Investigator of the Howard Hughes Medical Institute.

REFERENCES

Adewumi, O., Aflatoonian, B., Ahrlund-Richter, L., Amit, M., Andrews, P.W., Beighton, G., Bello, P.A., Benvenisty, N., Berry, L.S., Bevan, S.*, et al.* (2007). Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. Nat Biotechnol *25*, 803-816

Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z.D., Ziller, M., Croft, G.F., Amoroso, M.W., Oakley, D.H.*, et al.* (2011). Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. Cell *144*, 439-452

Chen, B.Z., Yu, S.L., Singh, S., Kao, L.P., Tsai, Z.Y., Yang, P.C., Chen, B.H., and Shoei-Lung Li, S. (2011). Identification of microRNAs expressed highly in pancreatic islet-like cell clusters differentiated from human embryonic stem cells. Cell biology international *35*, 29-37

Cheung, A.Y., Horvath, L.M., Grafodatskaya, D., Pasceri, P., Weksberg, R., Hotta, A., Carrel, L., and Ellis, J. (2011). Isolation of MECP2-null Rett Syndrome patient hiPS cells and isogenic controls through X-chromosome inactivation. Hum Mol Genet

Chureau, C., Prissette, M., Bourdet, A., Barbe, V., Cattolico, L., Jones, L., Eggen, A., Avner, P., and Duret, L. (2002). Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. Genome Res *12*, 894-908

Clemson, C.M., Hall, L.L., Byron, M., McNeil, J., and Lawrence, J.B. (2006). The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences. Proc Natl Acad Sci U S A *103*, 7688-7693

Csankovszki, G., Nagy, A., and Jaenisch, R. (2001). Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. J Cell Biol *153*, 773-784

Diaz Perez, S.V., Kim, R., Li, Z., Marquez, V.E., Patel, S., Plath, K., and Clark, A.T. (2012). Derivation of new human embryonic stem cell lines reveals rapid epigenetic progression in vitro that can be prevented by chemical modification of chromatin. Human molecular genetics *21*, 751-764

Dvash, T., Lavon, N., and Fan, G. (2010). Variations of X chromosome inactivation occur in early passages of female human embryonic stem cells. PLoS One *5*, e11330

Gore, A., Li, Z., Fung, H.L., Young, J.E., Agarwal, S., Antosiewicz-Bourget, J., Canto, I., Giorgetti, A., Israel, M.A., Kiskinis, E.*, et al.* (2011). Somatic coding mutations in human induced pluripotent stem cells. Nature *471*, 63-67

Guenther, M.G., Frampton, G.M., Soldner, F., Hockemeyer, D., Mitalipova, M., Jaenisch, R., and Young, R.A. (2010). Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. Cell Stem Cell *7*, 249-257

Hall, L.L., Byron, M., Butler, J., Becker, K.A., Nelson, A., Amit, M., Itskovitz-Eldor, J., Stein, J., Stein, G., Ware, C.*, et al.* (2008). X-inactivation reveals epigenetic anomalies in most hESC but identifies sublines that initiate as expected. J Cell Physiol *216*, 445-452

Hall, L.L., Byron, M., Sakai, K., Carrel, L., Willard, H.F., and Lawrence, J.B. (2002). An ectopic human XIST gene can induce chromosome inactivation in postdifferentiation human HT-1080 cells. Proc Natl Acad Sci U S A *99*, 8677-8682

Hanna, J., Cheng, A.W., Saha, K., Kim, J., Lengner, C.J., Soldner, F., Cassady, J.P., Muffat, J., Carey, B.W., and Jaenisch, R. (2010). Human embryonic stem cells with biological and

epigenetic characteristics similar to those of mouse ESCs. Proceedings of the National Academy of Sciences of the United States of America *107*, 9222-9227

Higgins, M.E., Claremont, M., Major, J.E., Sander, C., and Lash, A.E. (2007). CancerGenes: a gene selection resource for cancer genome projects. Nucleic Acids Res *35*, D721-726

Hoffman, L.M., Hall, L., Batten, J.L., Young, H., Pardasani, D., Baetge, E.E., Lawrence, J., and Carpenter, M.K. (2005). X-inactivation status varies in human embryonic stem cell lines. Stem Cells *23*, 1468-1478

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57

Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. (2003). Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res *31*, e15

Ji, P., Diederichs, S., Wang, W., Boing, S., Metzger, R., Schneider, P.M., Tidow, N., Brandt, B., Buerger, H., Bulk, E.*, et al.* (2003). MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene *22*, 8031-8041

Jia, F., Wilson, K.D., Sun, N., Gupta, D.M., Huang, M., Li, Z., Panetta, N.J., Chen, Z.Y., Robbins, R.C., Kay, M.A.*, et al.* (2010). A nonviral minicircle vector for deriving human iPS cells. Nat Methods *7*, 197-199

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118-127

Kim, D., Kim, C.H., Moon, J.I., Chung, Y.G., Chang, M.Y., Han, B.S., Ko, S., Yang, E., Cha, K.Y., Lanza, R.*, et al.* (2009). Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. Cell stem cell *4*, 472-476

Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I.*, et al.* (2010). Epigenetic memory in induced pluripotent stem cells. Nature *467*, 285-290

Lagarkova, M.A., Shutova, M.V., Bogomazova, A.N., Vassina, E.M., Glazov, E.A., Zhang, P., Rizvanov, A.A., Chestkov, I.V., and Kiselev, S.L. (2010). Induction of pluripotency in human endothelial cells resets epigenetic profile on genome scale. Cell Cycle *9*, 937-946

Laurent, L.C., Ulitsky, I., Slavin, I., Tran, H., Schork, A., Morey, R., Lynch, C., Harness, J.V., Lee, S., Barrero, M.J.*, et al.* (2011). Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. Cell Stem Cell *8*, 106-118

Lengner, C.J., Gimelbrant, A.A., Erwin, J.A., Cheng, A.W., Guenther, M.G., Welstead, G.G., Alagappan, R., Frampton, G.M., Xu, P., Muffat, J.*, et al.* (2010). Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations. Cell *141*, 872-883

Li, C., and Hung Wong, W. (2001). Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. Genome Biol *2*, RESEARCH0032

Liu, W.M., Mei, R., Di, X., Ryder, T.B., Hubbell, E., Dee, S., Webster, T.A., Harrington, C.A., Ho, M.H., Baid, J.*, et al.* (2002). Analysis of high density expression microarrays with signed-rank call algorithms. Bioinformatics *18*, 1593-1599

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

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R.*, et al.* (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell *1*, 55-70

Marahrens, Y., Panning, B., Dausman, J., Strauss, W., and Jaenisch, R. (1997). Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev *11*, 156-166

Marchetto, M.C., Carromeu, C., Acab, A., Yu, D., Yeo, G.W., Mu, Y., Chen, G., Gage, F.H., and Muotri, A.R. (2010). A model for neural development and treatment of rett syndrome using human induced pluripotent stem cells. Cell *143*, 527-539

Mayshar, Y., Ben-David, U., Lavon, N., Biancotti, J.C., Yakir, B., Clark, A.T., Plath, K., Lowry, W.E., and Benvenisty, N. (2010). Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. Cell stem cell *7*, 521-531

Mikkola, M., Olsson, C., Palgi, J., Ustinov, J., Palomaki, T., Horelli-Kuitunen, N., Knuutila, S., Lundin, K., Otonkoski, T., and Tuuri, T. (2006). Distinct differentiation characteristics of individual human embryonic stem cell lines. BMC Dev Biol *6*, 40

Namekawa, S.H., Payer, B., Huynh, K.D., Jaenisch, R., and Lee, J.T. (2010). Two-step imprinted X inactivation: repeat versus genic silencing in the mouse. Mol Cell Biol *30*, 3187-3205

Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature *448*, 313-317

Pageau, G.J., Hall, L.L., Ganesan, S., Livingston, D.M., and Lawrence, J.B. (2007). The disappearing Barr body in breast and ovarian cancers. Nat Rev Cancer *7*, 628-633

Park, I.H., Arora, N., Huo, H., Maherali, N., Ahfeldt, T., Shimamura, A., Lensch, M.W., Cowan, C., Hochedlinger, K., and Daley, G.Q. (2008a). Disease-specific induced pluripotent stem cells. Cell *134*, 877-886

Park, I.H., Zhao, R., West, J.A., Yabuuchi, A., Huo, H., Ince, T.A., Lerou, P.H., Lensch, M.W., and Daley, G.Q. (2008b). Reprogramming of human somatic cells to pluripotency with defined factors. Nature *451*, 141-146

Payer, B., and Lee, J.T. (2008). X chromosome dosage compensation: how mammals keep the balance. Annu Rev Genet *42*, 733-772

Payer, B., Lee, J.T., and Namekawa, S.H. (2011). X-inactivation and X-reactivation: epigenetic hallmarks of mammalian reproduction and pluripotent stem cells. Hum Genet *130*, 265-280

Pick, M., Stelzer, Y., Bar-Nur, O., Mayshar, Y., Eden, A., and Benvenisty, N. (2009). Clone- and gene-specific aberrations of parental imprinting in human induced pluripotent stem cells. Stem Cells *27*, 2686-2690

Pomp, O., Dreesen, O., Leong, D.F., Meller-Pomp, O., Tan, T.T., Zhou, F., and Colman, A. (2011). Unexpected X chromosome skewing during culture and reprogramming of human somatic cells can be alleviated by exogenous telomerase. Cell stem cell *9*, 156-165

Poplinski, A., Wieacker, P., Kliesch, S., and Gromoll, J. (2010). Severe XIST hypomethylation clearly distinguishes (SRY+) 46,XX-maleness from Klinefelter syndrome. Eur J Endocrinol *162*, 169-175

Prigione, A., Fauler, B., Lurz, R., Lehrach, H., and Adjaye, J. (2010). The senescence-related mitochondrial/oxidative stress pathway is repressed in human induced pluripotent stem cells. Stem Cells *28*, 721-733

Richardson, A.L., Wang, Z.C., De Nicolo, A., Lu, X., Brown, M., Miron, A., Liao, X., Iglehart, J.D., Livingston, D.M., and Ganesan, S. (2006). X chromosomal abnormalities in basal-like human breast cancer. Cancer Cell *9*, 121-132

Rogner, U.C., Wilke, K., Steck, E., Korn, B., and Poustka, A. (1995). The melanoma antigen gene (MAGE) family is clustered in the chromosomal band Xq28. Genomics *29*, 725-731

Rugg-Gunn, P.J., Ferguson-Smith, A.C., and Pedersen, R.A. (2005). Epigenetic status of human embryonic stem cells. Nat Genet *37*, 585-587

Rugg-Gunn, P.J., Ferguson-Smith, A.C., and Pedersen, R.A. (2007). Status of genomic imprinting in human embryonic stem cells as revealed by a large cohort of independently derived and maintained lines. Hum Mol Genet *16 Spec No. 2*, R243-251

Sharov, A.A., Dudekula, D.B., and Ko, M.S. (2005). A web-based tool for principal component and significance analysis of microarray data. Bioinformatics *21*, 2548-2549

Shen, Y., Matsuno, Y., Fouse, S.D., Rao, N., Root, S., Xu, R., Pellegrini, M., Riggs, A.D., and Fan, G. (2008). X-inactivation in female human embryonic stem cells is in a nonrandom pattern and prone to epigenetic alterations. Proc Natl Acad Sci U S A *105*, 4709-4714

Silva, S.S., Rowntree, R.K., Mekhoubad, S., and Lee, J.T. (2008). X-chromosome inactivation and epigenetic fluidity in human embryonic stem cells. Proc Natl Acad Sci U S A *105*, 4820-4825

Sunwoo, H., Dinger, M.E., Wilusz, J.E., Amaral, P.P., Mattick, J.S., and Spector, D.L. (2009). MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. Genome Res *19*, 347-359

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676

Tchieu, J., Kuoy, E., Chin, M.H., Trinh, H., Patterson, M., Sherman, S.P., Aimiuwu, O., Lindgren, A., Hakimian, S., Zack, J.A.*, et al.* (2010). Female human iPSCs retain an inactive X chromosome. Cell Stem Cell *7*, 329-342

Utikal, J., Polo, J.M., Stadtfeld, M., Maherali, N., Kulalert, W., Walsh, R.M., Khalil, A., Rheinwald, J.G., and Hochedlinger, K. (2009). Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. Nature *460*, 1145-1148

Wang, W., Yang, J., Liu, H., Lu, D., Chen, X., Zenonos, Z., Campos, L.S., Rad, R., Guo, G., Zhang, S.*, et al.* (2011). Rapid and efficient reprogramming of somatic cells to induced pluripotent stem cells by retinoic acid receptor gamma and liver receptor homolog 1. Proceedings of the National Academy of Sciences of the United States of America

Ware, C.B., Wang, L., Mecham, B.H., Shen, L., Nelson, A.M., Bar, M., Lamba, D.A., Dauphin, D.S., Buckingham, B., Askari, B.*, et al.* (2009). Histone deacetylase inhibition

elicits an evolutionarily conserved self-renewal program in embryonic stem cells. Cell Stem Cell *4*, 359-369

Warren, L., Manos, P.D., Ahfeldt, T., Loh, Y.H., Li, H., Lau, F., Ebina, W., Mandal, P.K., Smith, Z.D., Meissner, A.*, et al.* (2010). Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. Cell Stem Cell *7*, 618-630

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature *448*, 318-324

Westfall, S.D., Sachdev, S., Das, P., Hearne, L.B., Hannink, M., Roberts, R.M., and Ezashi, T. (2008). Identification of oxygen-sensitive transcriptional programs in human embryonic stem cells. Stem Cells Dev *17*, 869-881

Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol Cell *5*, 695-705

Yoshida, Y., Takahashi, K., Okita, K., Ichisaka, T., and Yamanaka, S. (2009). Hypoxia enhances the generation of induced pluripotent stem cells. Cell Stem Cell *5*, 237-241

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R.*, et al.* (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917-1920

Zhang, L.F., Huynh, K.D., and Lee, J.T. (2007). Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. Cell *129*, 693-706

FIGURE LEGENDS

Figure 1. Partial X-reactivation and high frequency Class III conversion in female hiPSCs.

(A) RNA FISH of IMR-90 and undifferentiated hiPS-10. XIST RNA, red. Cot-1 RNA, green. Asterisk, XIST cloud. Arrow, COT-1 hole.

(B) RNA FISH for XIST and Cot-1, followed by X-paint DNA FISH. Arrows, Cot-1 holes. Asterisk, XIST cloud. Double arrowheads, X-chromosomes. Shown is hiPS-1 p. 6.

(C) Immunostaining for H3K27me3 (red) followed by DNA FISH (green) for X-chromosomes in differentiated (d16) hiPSCs.

(D) Real-time PCR of XIST expression. Ct values were normalized to IMR-90 cells (set to 1) and GAPDH, and values represent averages of triplicates. Error bars, standard deviations (SD) of the mean. P-values were calculated using one-tailed Student's t-test assuming equal variance; * P = 0.04; ** P = 0.004. See also Figure S4.

(E) Summary of XIST RNA FISH. n, sample size.

(F) Three classes of XXY hiPSCs (d0, p.4). Arrows, Cot-1 holes. Asterisk, XIST cloud. Double arrowheads, X-chromosomes. See also Figure S5.

Figure 2. Class III female hiPSCs have unique global gene expression patterns.

(A) Pearson correlation coefficients between whole sets of gene expression levels (RMA normalization) in the 10 female hiPSCs samples. hiPS-2 p.9; hiPS-9 p.7; hiPS-10 p.24; hiPS-3 p.14; hiPS-12 p.23; hiPS-11 p.16 high O2; hiPS-11 p.16 low O2; hiPS-9 p.19 c.III; hiPS-12 p.30 c.III.

(B,C) PCA of gene expression patterns in indicated samples. Plot of component loadings shows relations of each microarray sample (RMA normalization) in PC1 versus PC2 (B), and PC2 versus PC3 (C). Class II to III conversion indicated by arrows.

(D,E) Expression levels for genes downregulated (D) and upregulated (E) in Class III samples. Shown are top genes with highest correlation (D) or anticorrelation (E) to XIST

expression, among those that are differentially expressed in at least 6 out of 8 Class II versus Class III lines. See also Figure S6.

Figure 3. Microarray analyses of male vs female hiPSCs and hESCs.

(A,B) PCA shown in two dimensions for ComBat-corrected (A) and uncorrected (B) samples. See Table S-IV for list of samples, GEO numbers, PubMed ID, and abbreviations. Class III hiPSCs from this study (L3) in green; L3 LO = hiPS-2 c.III in low oxygen p.50, L3-6 = hiPS-6C-1 c.III p. 28, L3-12 = hiPS-12D-1 c.III p.28. Blue, male hiPSC; pink, female hiPSC; black, female hESC).

(C) Expression heatmaps normalized to hiPS-9 and -12 c.III average (expression set as 0). Shown are genes up- and downregulated in Class III hiPSCs (Table III). L3 (cIII) = hiPS-9, -12 c.III; L3 (cIII) LO = hiPS-2 c.III; L3 (cIII) Dis. = hiPS 6C-1, 12D-1 c.III. Averages shown for duplicate and triplicate samples.

(D) XIST expression in indicated lines plotted against correlation of expression pattern across differentially expressed genes.

Figure 4. Comparative *in vitro* growth rates and *in vivo* differentiation.

(A) Growth profiles for indicated hiPSC lines in ambient oxygen. Doubling times calculated from line equations. One colony for each line was mechanically passaged and plated in 10 replicate, MEF-coated wells. Cells were trypsinized and counted. Averages shown. Percentages of $XIST^+$ nuclei at the end of the experiment shown. n, sample size. N.D., not determined. Two biological replicates performed; similar results; one shown.

(B) Growth profiles for indicated lines in physiological oxygen. Three colonies for each line were plated in quadruplicate on MEF-coated plates, then processed as in (A).

(C) Growth differences as a function of passage number at ambient or physiological oxygen. Average values shown.

(D) Teratomas from matched Class II-III sublines of hiPS-2 and hiPS-12.

(E) Representative histologic sections of Class II and III teratomas.

Figure 1

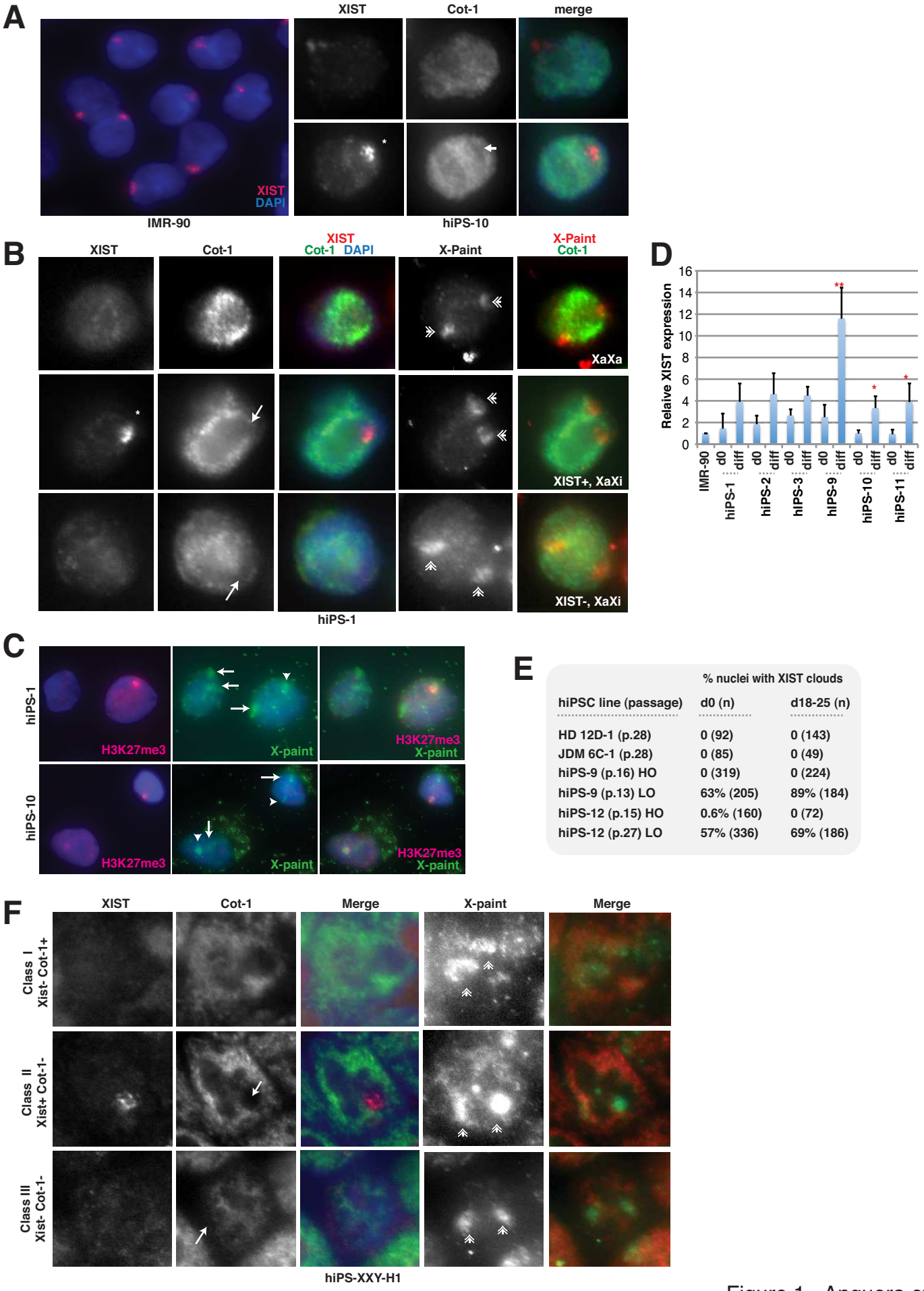


Figure 1 - Anguera et al.

Figure 2

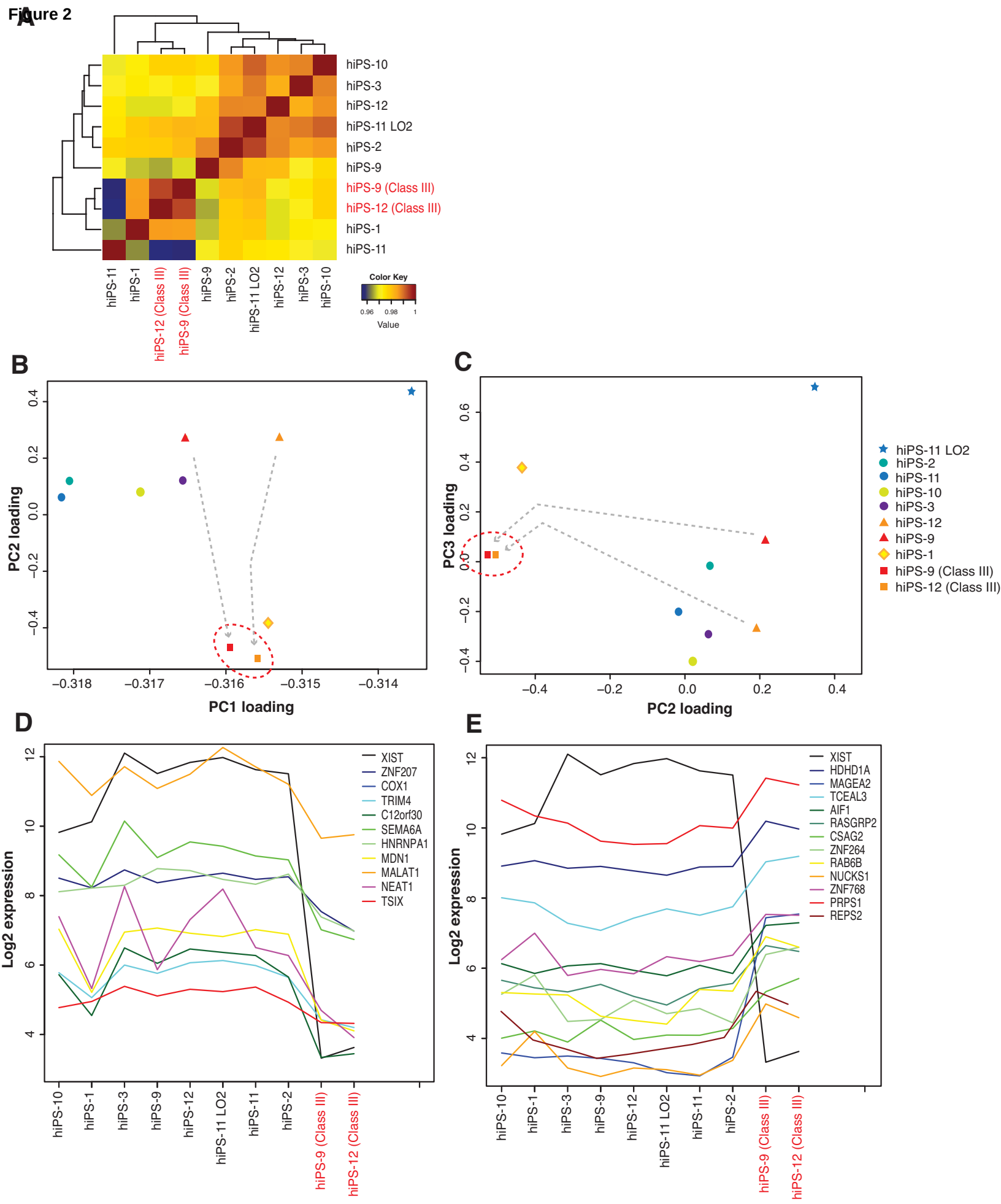


Figure 2 - Anguera et al.

Figure 3

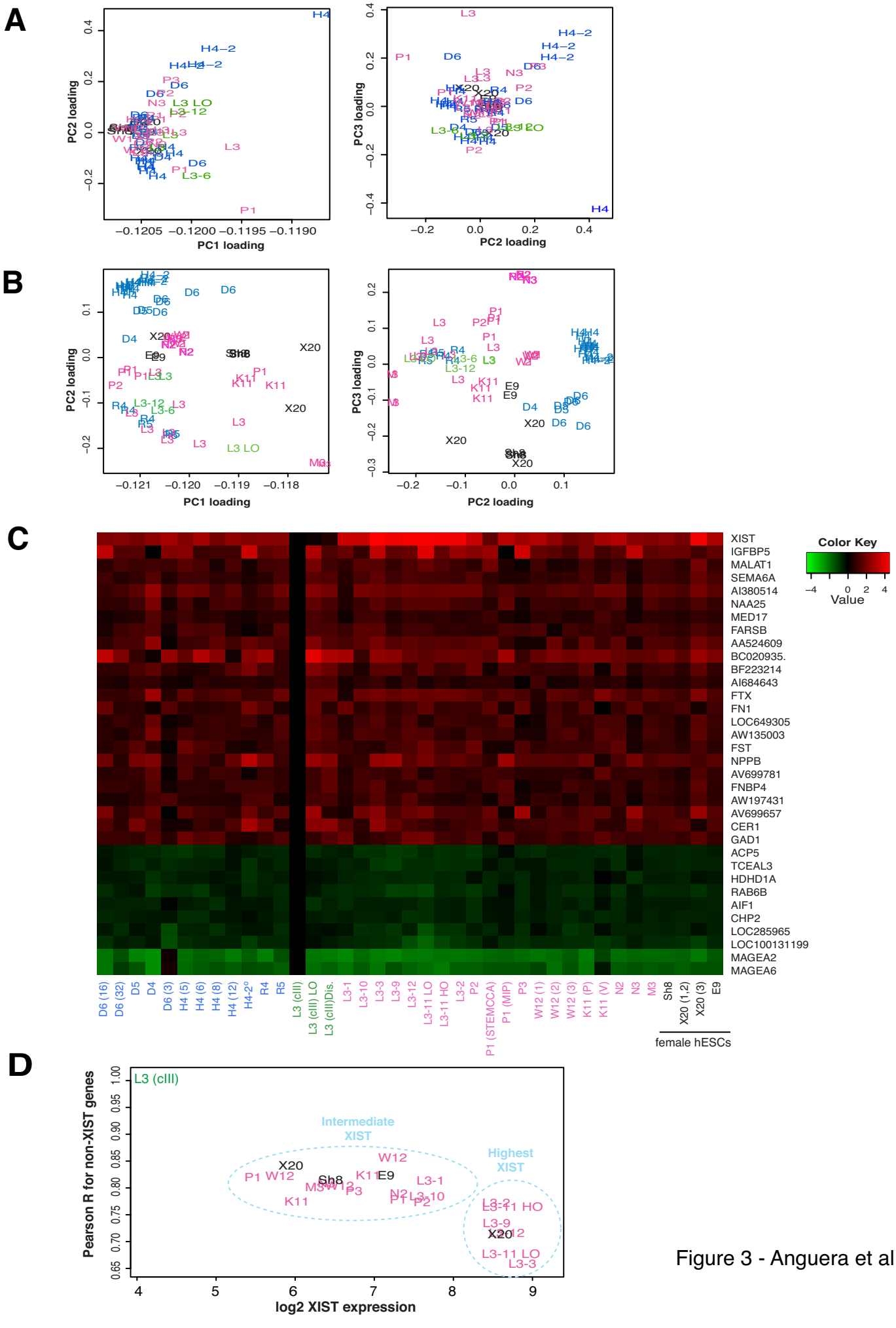
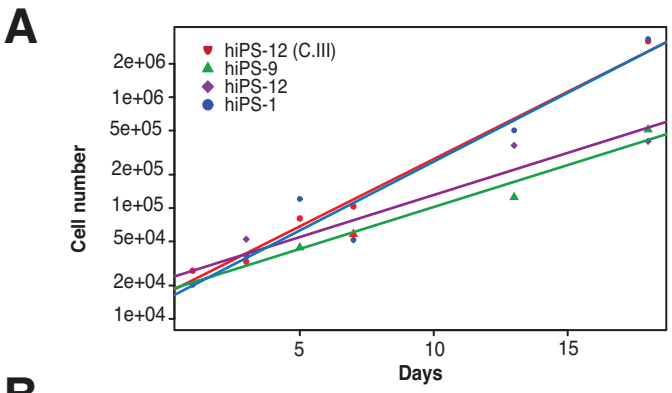


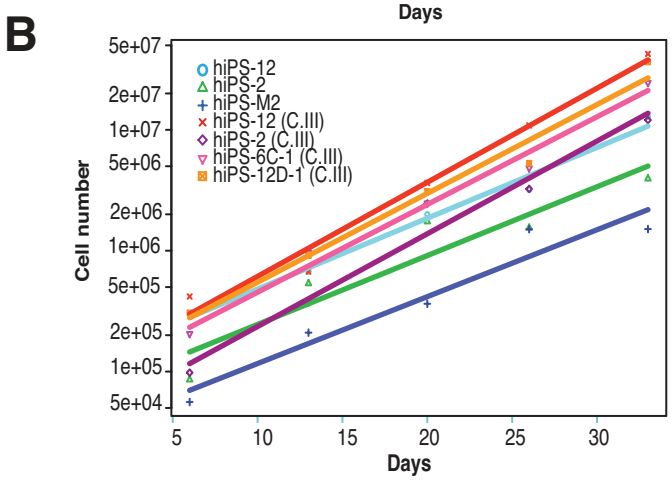
Figure 3 - Anguera et al

Figure 4



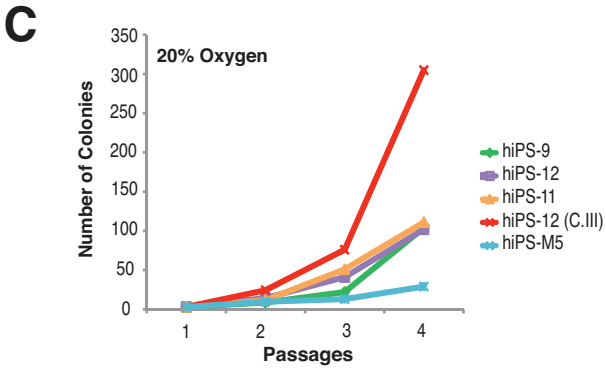
20% Oxygen

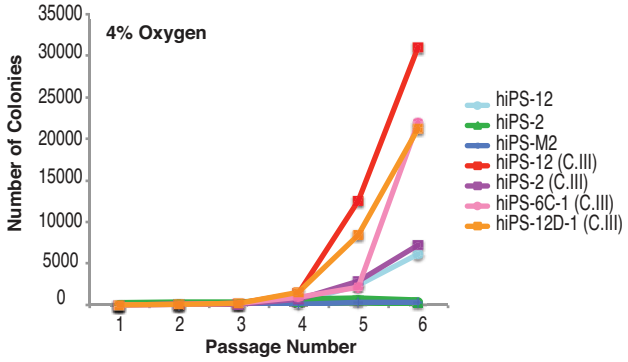
Sample	Doubling Time (hours)	% XIST (n)
hiPS-12	95.1	83% (81)
hiPS-9	95.1	85% (65)
hiPS-12 (C.III)	58.5	0
hiPS-1	58.2	N.D.



4% Oxygen

Sample	Doubling Time (hours)	% XIST (n)
hiPS-2 p.14	126.8	80% (86)
hiPS-2 p.32	104.4	79% (141)
hiPS-2 p.52 (C.III)	98.3	0
hiPS-9 p.30	92.7	35% (115)
hiPS-9 p.29 (C.III)	97.6	0
hiPS-11 p.34	103.9	93% (55)
hiPS-12 p.28	120.5	83% (116)
hiPS-12 p.32	94.7	29% (59)
hiPS-12 p.48 (C.III)	92.7	0 (98)
hiPS-6C-1 p.33 (C.III)	99.6	0 (109)
hiPS-12D-1 p.33 (C.III)	98.3	0 (127)
hiPS-M2 p.28	130.5	
hiPS-M6 p.23	105.3	





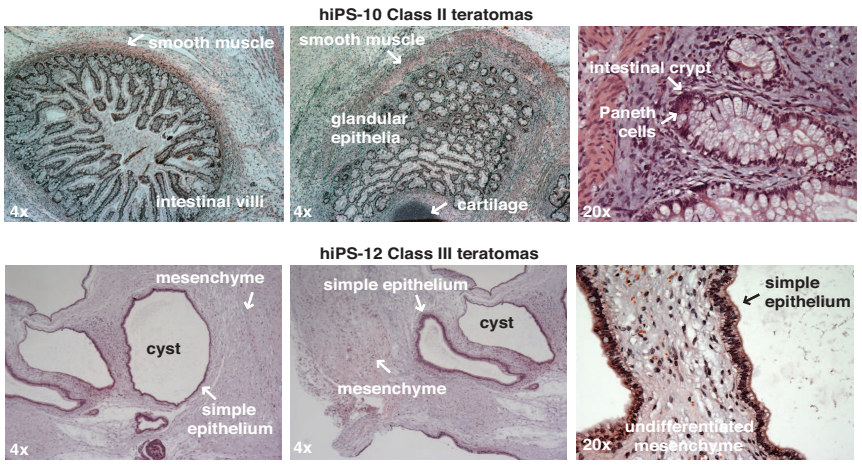


Figure 4 - Anguera et al.

Table I: X-chromosome states in hiPSC lines

	Undifferentiated d0				Differentiated d14-20		
	Sample size (n)	XIST+ cells	XIST- cells	XIST- cells	Sample size (n)	XIST+ cells	Class designation
hiPSC line (passage)		**XaXi**	**XaXa**	**XaXi* (Xi* in Cot-1 hole)**			
hiPS-1 (p.6)	78	60%	6.4%	33%	95	55%	Class II predominant
hiPS-2 (p.32)	100	68%	8%	24%	153	84%	Class II predominant
hiPS-3 (p.14)	117	58%	7%	35%	273	54%	Class II predominant
hiPS-9 (p.7)	111	78%	4.5%	17%	161	91%	Class II predominant
hiPS-10 (p.29)	154	78%	3.2%	20%	139	80%	Class II predominant
hiPS-11 (p.6)	106	67%	8%	24%	287	80%	Class II predominant
hiPS-12 (p.6)	79	65%	14%	22%	71	72%	Class II predominant
hiPS-2 Class III (p. 50)	145	0	ND	ND	80	0%	Class III
hiPS-9 Class III (p.15)	113	0	0	100%	224	0%	Class III
hiPS-12 Class III (p. 27)	105	0	0	100%	72	0%	Class III
hiPS-XXY-L1 (p.7)	127	75%	7%	18%	149	91%	Class II predominant
hiPS-XXY-L3 (p.7)	279	71%			232	90%	
hiPS-XXY-L4 (p.7)	357	66%			61	84%	
hiPS-XXY-H1 (p.7)	109	81%	2.80%	17%	148	91%	Class II predominant
hiPS-XXY-H2 (p.7)	194	75%			285	94%	
hiPS-XXY-H3 (p.7)	192	89%			205	93%	
hiPS-XXY-H4 (p.7)	332	72%			134	93%	
hiPS-MM 11 (p.0)	121	84%	4%	11%	ND		
hiPS-MM 13 (p.0)	112	76%	8%	18%	ND		
hiPS-TA 11 (p.0)	116	71%	6%	23%	ND		
hiPS-TA 12 (p.0)	138	76%	4%	20%	ND		
hiPS-TA 13 (p.0)	77	68%	8%	25%	ND		
hiPS-TB 10 (p.0)	83	74%	8%	18%	ND		
hiPS-TB 12 (p.0)	137	80%	5%	15%	ND		
hiPS-TB 13 (p.0)	85	73%	5%	22%	ND		

Table II: Genes with greatest expression change in Class III hiPSC.

A. Genes showing greatest upregulation

Gene Name	Description	Chr.	FDR
MAGEA6	melanoma antigen family A, 6	X	0
MAGEA2 /// MAGEA2B	melanoma antigen family A, 2 /// melanoma antigen family A,	X	0
ACP5	acid phosphatase 5, tartrate resistant	19	0
TCEAL3	transcription elongation factor A (SII)-like 3	X	0.0001
HDHD1A	haloacid dehalogenase-like hydrolase domain containing 1A	X	0.0007
RAB6B	RAB6B, member RAS oncogene family	3	0.0017
AIF1	allograft inflammatory factor 1	6	0.0082
CHP2	calcineurin B homologous protein 2	16	0.0111
LOC285965	hypothetical protein LOC285965	7	0.0231
LOC100131199	LOC100131199 transmembrane protein 178-like [*Homo sapie*	7	0.0475

B. Genes showing greatest downregulation

Gene name	Description	Chr.	FDR
MALAT1	metastasis associated lung adenocarcinoma transcript 1 (non-	11	0
XIST	X (inactive)-specific transcript (non-protein coding)	X	0
SEMA6A	sema domain, transmembrane domain (TM), and cytoplasmic	5	0
AI380514.1	tg01e02.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE	2	0
NAA25	N(alpha)-acetyltransferase 25, NatB auxiliary subunit	12	0
MED17	mediator complex subunit 17	11	0
FARSB	phenylalanyl-tRNA synthetase, beta subunit	2	0
AA524609.1	nh34c11.s1 NCI_CGAP_Pr3 Homo sapiens cDNA clone IMAGE:	/	0
BC020935.1	Similar to otoconin 90, clone IMAGE:4277593	13	0.0001
BF223214.1	7q30f03.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE	6	0.0003
AI684643.1	wa84h10.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clon	12	0.0005
FTX	FTX, NCRNA00182 non-protein coding RNA 182 [H*omo sapien*	X	0.0006
FN1	fibronectin 1	2	0.0007
LOC649305	hypothetical LOC649305	8	0.0008
AW135003.1	UI-H-BI1-abt-c-08-0-UI.s1 NCI_CGAP_Sub3 Homo sapiens cD	11	0.0028
FST	follistatin	5	0.0033
NPPB	natriuretic peptide precursor B	1	0.0058
AV699781.1	AV699781 GKC Homo sapiens cDNA clone GKCEKC01 3-, mRN	/	0.0061
FNBP4	formin binding protein 4	11	0.0063
AW197431.1	xm39b03.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAG	12	0.0064
IGFBP5	insulin-like growth factor binding protein 5	2	0.0167
NEAT1	NEAT1 nuclear paraspeckle assembly transcript 1 (non-proteir	11	0.0211
CER1	cerberus 1, cysteine knot superfamily, homolog (Xenopus laev	9	0.0265
GAD1	glutamate decarboxylase 1 (brain, 67kDa)	2	0.0484

Table III: Genes with greatest expression correlation with XIST.

A. Positive correlation

Gene	Description	Chr.	FDR	Corr.	# chip
XIST	X (inactive)-specific transcript (non-protein coding)	X	0	1	8
ZNF207	zinc finger protein 207	17	0.0034	0.9446	8
COX1	cytochrome c oxidase I	MT	0	0.9444	8
NAA25	N(alpha)-acetyltransferase 25, NatB auxiliary subunit	12	0	0.9423	8
TRIM4	tripartite motif-containing 4	7	0.0001	0.9416	7
SEMA6A	sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6A	5	0	0.9415	8
AV699781.1	AV738585 CB Homo sapiens cDNA clone CBFAWD05 5-, mRNA sequence	/	0.0258	0.9396	8
AA524609.1	nh34c11.s1 NCI_CGAP_Pr3 Homo sapiens cDNA clone IMAGE:954260 similar to contains Alu repetiti	/	0	0.9384	8
HNRNPA1 /// LOC728	heterogeneous nuclear ribonucleoprotein A1 /// hypothetical LOC728844	12	0.0044	0.9370	8
N72610	za46h03.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:295637 3-, mRNA seq	/	0.0287	0.9270	6
AI247478	qh56c08.x1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:1848686 3- simi	8	0.0007	0.9261	4
BE503070	hz83b02.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3214539 3-, mRNA sequence	5	0.0001	0.9204	8
MDN1	MDN1, midasin homolog (yeast)	6	0	0.9120	8
AI806781	wf15b12.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2350655 3-, mRNA sequence	17	0	0.9092	8
BG281679	602402364F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4544871 5-, mRNA sequence	/	0.0056	0.9090	4
BF223214	7q30f03.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:3699965 3-, mRNA sequence	6	0.0003	0.9081	8
AI539426	te46d04.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:2089735 3-, mRNA sequence	12	0.0039	0.9035	8
AI380514	tg01e02.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2107514 3-, mRNA sequence	2	0	0.9027	8
W86781	zh64a03.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:416812 3-, mRN	20	0.0003	0.9020	8
AI056872	oz03e12.x1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:1674286 3-, mR	6	0.0007	0.9018	1
MLL2	myeloid/lymphoid or mixed-lineage leukemia 2	12	0.0035	0.9016	8
FUS	fusion (involved in t(12;16) in malignant liposarcoma)	16	0.0012	0.9016	8
AA398740	zt75f06.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:728195 3-, mRNA sequence	1	0.0003	0.9016	7
CHD1L	Chromodomain helicase DNA binding protein 1-like	1	0.0006	0.9006	8
FTX	NCRNA00182 non-protein coding RNA 182 [*Homo sapiens*]	X	0.0006	0.9005	8
CHPT1	Choline phosphotransferase 1	12	0.0023	0.8984	3
AI367034.1	qq40f01.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1935001 3- similar to contains Al	2	0.0081	0.8958	8
MALAT1	metastasis associated lung adenocarcinoma transcript 1 (non-protein coding)	11	0	0.8948	8
AA436194.1	zv22f03.s1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:754397 3- similar to contains Alu	4	0.0434	0.8948	8
MED17	mediator complex subunit 17	11	0	0.8939	8

B. Negative correlation

Gene	Description	Chr.	FDR	Xist_cor	# Chip
HDHD1A	haloacid dehalogenase-like hydrolase domain containing 1A	X	0.0007	-0.9824	8
MAGEA2 /// MAGEA2B	melanoma antigen family A, 2 /// melanoma antigen family A, 2B	X	0	-0.9801	3
TCEAL3	transcription elongation factor A (SII)-like 3	X	0.0001	-0.9560	8
AIF1	allograft inflammatory factor 1	6	0.0082	-0.9557	0
RASGRP2	RAS guanyl releasing protein 2 (calcium and DAG-regulated)	11	0.0211	-0.9533	7
ACP5	acid phosphatase 5, tartrate resistant	19	0	-0.9490	0
SDCCAG8	serologically defined colon cancer antigen 8	1	0.0017	-0.9474	1
NXT2	nuclear transport factor 2-like export factor 2	X	0.0014	-0.9461	0
MAGEA6	melanoma antigen family A, 6	X	0	-0.9437	0
CHP2	calcineurin B homologous protein 2	16	0.0111	-0.9284	3
CSAG2 /// CSAG3	CSAG family, member 2 /// CSAG family, member 3	X	0	-0.9256	0
C20orf94	chromosome 20 open reading frame 94	20	0.0178	-0.9185	3
ZNF264	zinc finger protein 264	19	0.0078	-0.9169	7
RAB6B	RAB6B, member RAS oncogene family	3	0.0017	-0.9138	2
C1orf77	chromosome 1 open reading frame 77	1	0.0035	-0.9127	8
NUCKS1	Nuclear casein kinase and cyclin-dependent kinase substrate 1	1	0.003	-0.9112	3
MBNL3	muscleblind-like 3 (Drosophila) [*Homo sapiens*]	X	0.0035	-0.9096	2
ZNF768	zinc finger protein 768	16	0.0064	-0.9038	7
LOC284242	hypothetical protein LOC284242	18	0.0191	-0.9036	7
PRPS1	phosphoribosyl pyrophosphate synthetase 1	X	0.0173	-0.9019	8
REPS2	RALBP1 associated Eps domain containing 2	X	0.0197	-0.9009	6

INVENTORY OF SUPPLEMENTAL INFORMATION

Molecular signatures of human induced pluripotent stem cells (hiPSCs) highlight sex differences and cancer genes

Montserrat C. Anguera[1,2,3], Ruslan Sadreyev [1,2,3], Zhaoqing Zhang[4], Attila Szanto[1,2,3], Bernhard Payer[1,2,3], Steven D. Sheridan[5,6], Showming Kwok[5], Stephen J. Haggarty[6], Mriganka Sur[5], Jason Alvarez[3,7], Alexander Gimelbrant[3,7], Maisam Mitalipova[8], James E. Kirby[9], and Jeannie T. Lee[1,2,3]

[1]Howard Hughes Medical Institute
[2]Department of Molecular Biology, Massachusetts General Hospital, and
[3]Department of Genetics, Harvard Medical School, Boston, MA 02114, USA.
[4]SAB Biosciences, Qiagen, 6951 Executive Way, Suite 100, Frederick, MD 21703, USA.
[5]Department of Brain and Cognitive Sciences, Picower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[6]Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston MA 02114, USA
[7]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA
[8]Whitehead Institute for Biomedical Sciences, 9 Cambridge Center, Cambridge, MA 02142, USA
[9]Department of Pathology, Beth Israel Deaconess Medical Center, 330 Brookline Ave, Boston, MA 02215, USA.
*Corresponding Author: lee@molbio.mgh.harvard.edu
Running Title: Sex differences in hiPSCs

Figure S1. Generation of human female iPSC lines.

This figure contains a graphical representation of how the reprogramming of female IMR-90 fibroblasts was performed (generating hiPS-1,2,3,9,10,11,12). Also shown is data demonstrating pluripotency of all hiPSCs used in subsequent experiments (IF staining, qPCR for pluripotency genes, DNA methylation of OCT4 & NANOG promoters, and silencing of retroviral reprogramming factors). This figure can't be directly linked to a particular main figure (serves as basis for all cell lines examined here), but is linked to Figure S2 and S3.

Figure S2. All female hiPSCs can differentiate into the three germ lineages.

This figure demonstrates that female hiPSCs **(**hiPS-1,2,3,9,10,11,12) can be differentiated either *in vivo* or *in vitro* forming all 3 germ lineages. Similar to Figure S1, it is independent yet also a basis for all main figures, and is linked to Figures S1 and S3.

Figure S3. Female hiPSCs are karyotypically normal.

This figure demonstrates that 3 lines (hiPS-3, -11, -12) have a normal karyotype. Similar to S1 and S2, it functions to demonstrate reprogramming quality and is linked to Figures S1 and S2.

Figure S4. Allele-specific expression analysis of X-linked genes for hiPS-11 (A) and hiPS-12 (B) at day 0 and day 50 of differentiation.

This figure shows the allele-specific expression pattern for 2 hiPSC lines for a set of X-linked genes, for both undifferentiated and differentiated cells. It is linked to Figure 1.

Figure S5. hiPSCs derived from Klinefelter Syndrome fibroblasts.

This figure demonstrates that Klinefelter Syndrome hiPSCs have been successfully reprogrammed from fibroblasts (similar to figure S1) with IF staining for pluripotency markers. It also demonstrates that oxygen concentrations influence the expression profile of stress response genes. It is associated with Figure 1F and Table 1, containing XIST RNA-FISH results for these cell lines when reprogrammed in different oxygen concentrations.

Figure S6. qRT-PCR validations for microarray results of genes up- and down-regulated in Class III hiPSCs.

This figure contains qPCR results for 6 genes that are differentially expressed in Class III hiPSCs. It is linked to Figure 2 and Tables 2, S2, S3.

Figure S7. Array comparative genomic hybridization (aCGH) analyses of matched Class II and Class III hiPSCs.

This figure contains CGH results for 2 sets of paired hiPSCs (Class II vs Class III) indicating that there are no major insertion/deletion events with XIST loss. It is linked to Figure 4 and Figure S3.

Supplemental Table S1.

This table demonstrates that HDAC inhibitor treatment has no effect on XIST expression in both Class II and III hiPSCs (hESCs are included as a control, indicating that H9 is responsive).

Supplemental Table S2.

This Table lists all genes that are positively correlated with XIST expression, using a relaxed criterion (6/8 samples must have 2-fold or greater change). It is linked to Table 2 (listing differential expression defined with strict criterion of 8/8 samples).

Supplemental Table S3.

This Table lists all genes that are negatively correlated with XIST expression, using a relaxed criterion (6/8 samples must have 2-fold or greater change). This is table is similar to Supplemental Table S2 (but indicates negative correlation) and is also linked to Table 2.

Supplemental Table S4.

This Table lists all of the microarray data sets (Sample ID, Collection ID, Cell line Description, PubMed ID for reference) used for Figure 3. It also contains the abbreviations used for each sample. It is linked to Figure 3.

SUPPLEMENTAL ONLINE INFORMATION

Molecular signatures of human induced pluripotent stem cells (hiPSCs) highlight sex differences and cancer genes

Montserrat C. Anguera[1,2,3], Ruslan Sadreyev [1,2,3], Zhaoqing Zhang[4], Attila Szanto[1,2,3], Bernhard Payer[1,2,3], Steven D. Sheridan[5,6], Showming Kwok[5], Stephen J. Haggarty[6], Mriganka Sur[5], Jason Alvarez[3,7], Alexander Gimelbrant[3,7], Maisam Mitalipova[8], James E. Kirby[9], and Jeannie T. Lee[1,2,3]

[1]Howard Hughes Medical Institute

[2]Department of Molecular Biology, Massachusetts General Hospital, and

[3]Department of Genetics, Harvard Medical School, Boston, MA 02114, USA.

[4]SAB Biosciences, Qiagen, 6951 Executive Way, Suite 100, Frederick, MD 21703, USA.

[5]Department of Brain and Cognitive Sciences, Picower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[6]Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston MA 02114, USA

[7]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA

[8]Whitehead Institute for Biomedical Sciences, 9 Cambridge Center, Cambridge, MA 02142, USA

[9]Department of Pathology, Beth Israel Deaconess Medical Center, 330 Brookline Ave, Boston, MA 02215, USA.

*Corresponding Author: lee@molbio.mgh.harvard.edu

Running Title: Sex differences in hiPSCs

SUPPLEMENTAL FIGURES & LEGENDS

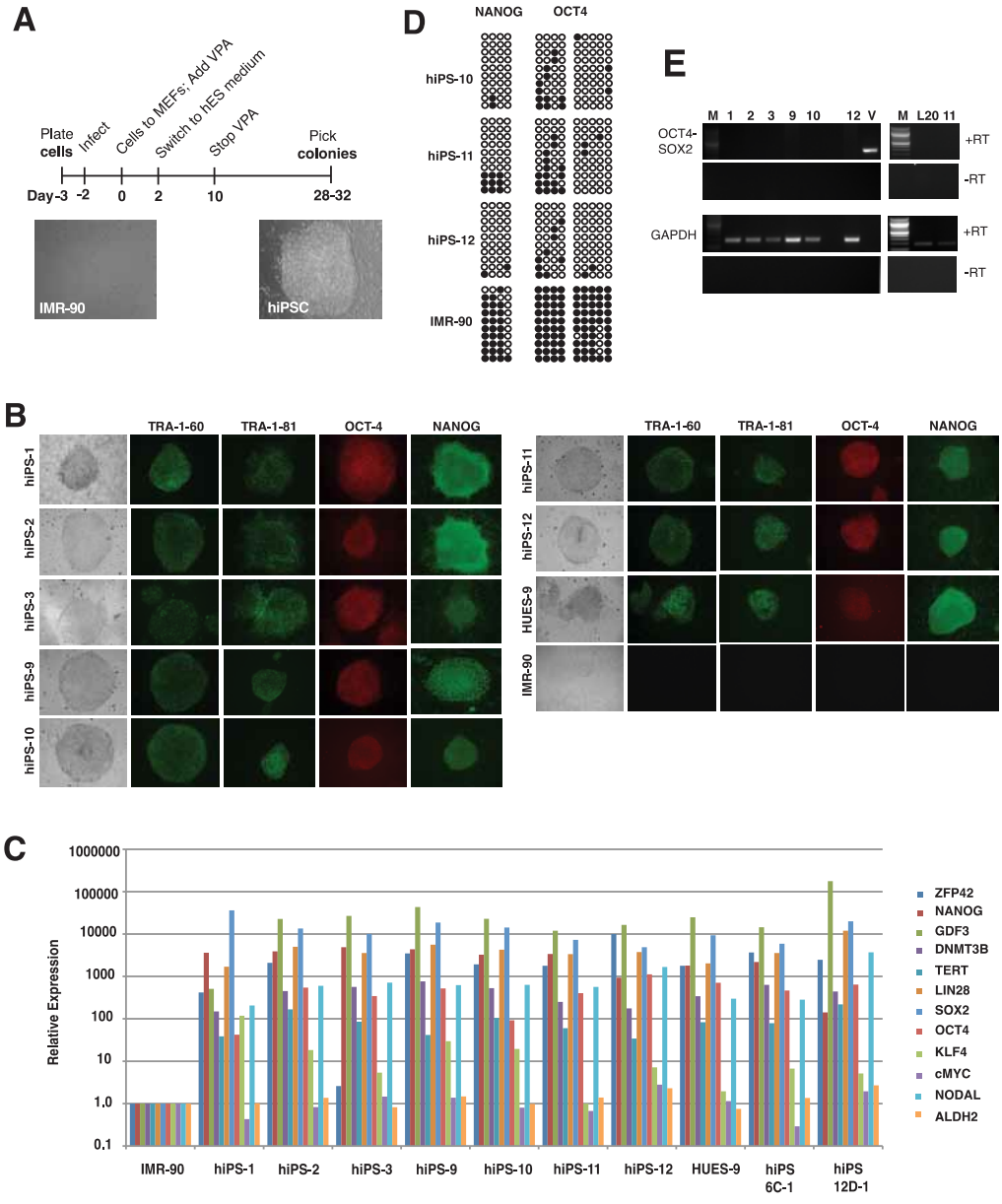


Figure S1

Figure S1. Generation of human female iPSCs.

(A) Schematic diagram showing the reprogramming strategy. The parental line IMR-90 and a representative hiPSCs colony are shown.

(B) Immunostaining for pluripotency markers Tra-1-60, Tra-1-81, OCT-4, and NANOG for all hiPSCs lines, HUES-9 (hESCs positive control), and the parental IMR-90 line (negative control).

(C) Real-time PCR analysis of endogenously expressed pluripotency genes (ZFP42, NANOG, GDF3, DNMT3B, TERT, LIN28, SOX-2, OCT4, KLF-4) and control genes (c-MYC, NODAL, ALDH2). Ct values were normalized to GAPDH expression and levels present in IMR-90 cell (set to 1). Positive control samples are HUES-9 (hESCs), and two female disease model hiPSCs lines (12D-1, 6C-1) previously characterized (Park et al., 2008a). The hiPS-12 sample is from a late passage that lost XIST expression (Class III).

(D) Bisulfite sequencing of promoter regions of NANOG and OCT4 for selected hiPSCs lines.

(E) RT-PCR analysis indicates silencing of viral reprogramming factors in hiPSCs lines. Positive control is DNA from the pEYK construct (V). GAPDH was used as control.

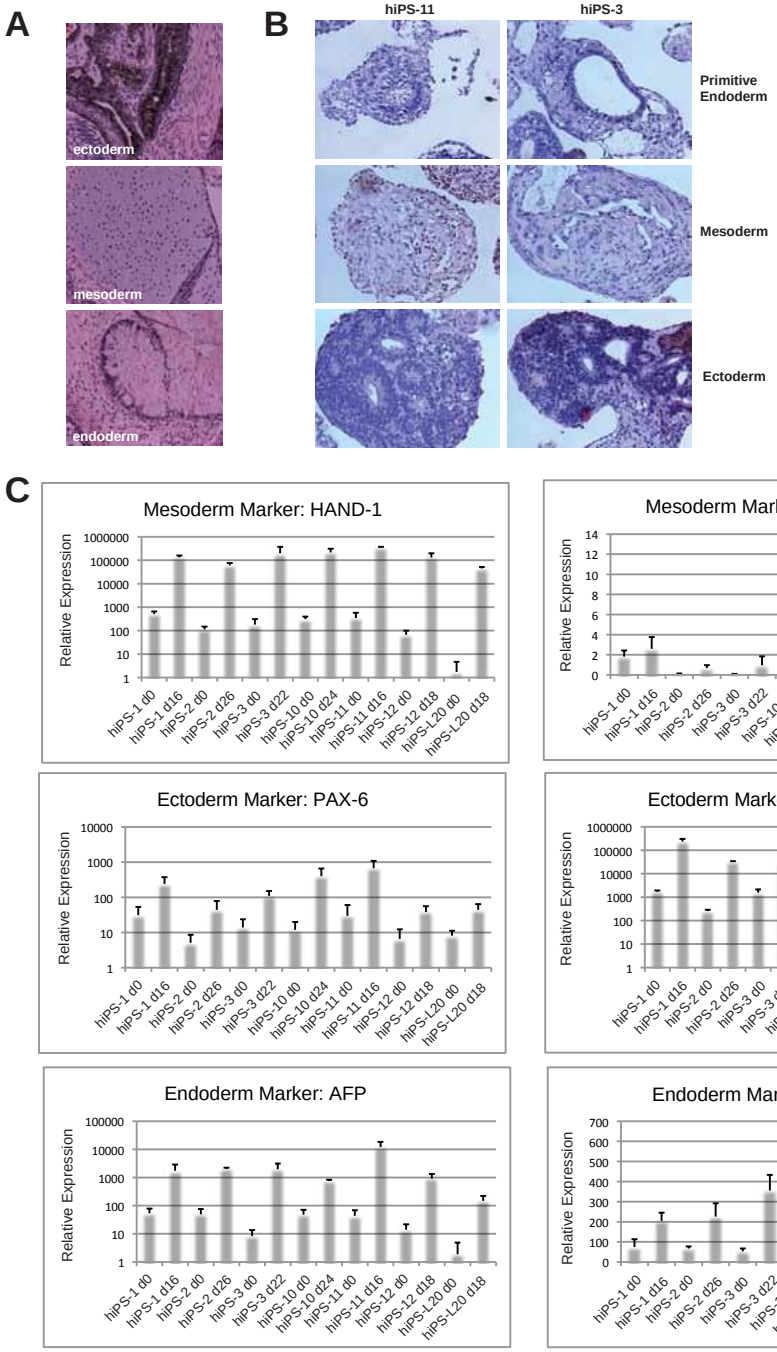

A
ectoderm
mesoderm
endoderm
B
hiPS-11
hiPS-3
Primitive Endoderm
Mesoderm
Ectoderm
C
Mesoderm Marker: HAND-1
Mesoderm Marker: GATA-2
Ectoderm Marker: PAX-6
Ectoderm Marker: NCAM-1
Endoderm Marker: AFP
Endoderm Marker: GATA-4
Relative Expression
hiPS-1 d0
hiPS-1 d16
hiPS-2 d0
hiPS-2 d26
hiPS-3 d0
hiPS-3 d22
hiPS-10 d0
hiPS-10 d24
hiPS-11 d0
hiPS-11 d16
hiPS-12 d0
hiPS-12 d18
hiPS-L20 d0
hiPS-L20 d18

Figure S2

Figure S2. All female hiPSCs can differentiate into the three germ lineages.

(A) Teratoma analysis for hiPS-10 (Class II), indicating differentiation into ectoderm (pigmented retinal epithelium), endoderm (gut-like epithelium), and mesoderm (cartilage) lineages. The sections were stained with H & E.

(B) Embryoid body (EBs) differentiation into 3 germ lineages. hiPSCs (hiPS-3, 11) were differentiated for 19 days, embedded in paraffin, sectioned (5uM) and stained with H & E.

(C) Real-time PCR analysis for two markers for each of the three germ lineages for undifferentiated and differentiated female hiPSCs samples. Error bars are for the standard deviation of the mean for triplicate measures.

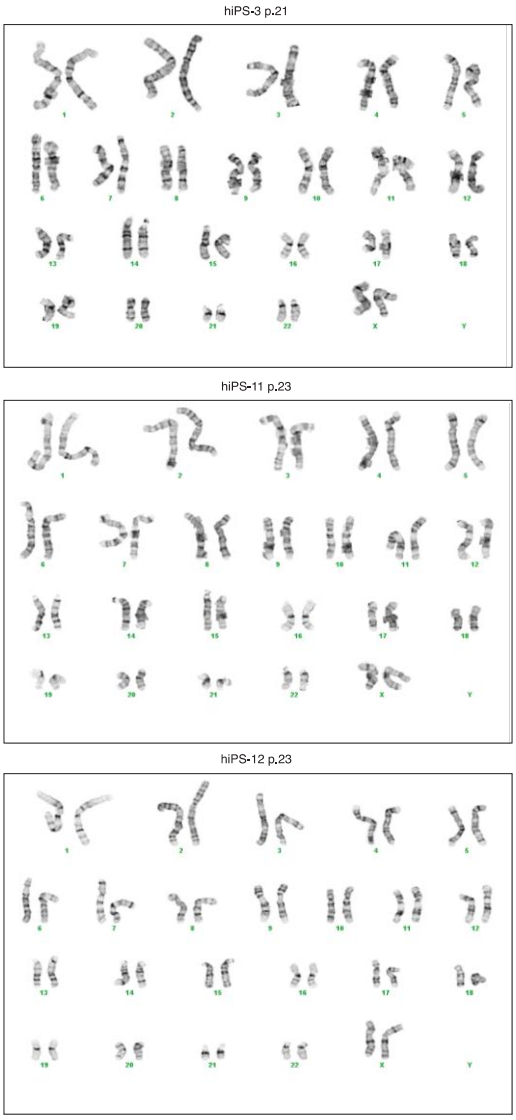


Figure S3

Figure S3. Female hiPSCs are karyotypically normal.

Cytogenetic analysis for various hiPSCs lines (hiPS-3, 11, 12) reprogrammed from IMR-90 fibroblasts at later passages (~20) indicate normal karyotypes. Karyotype analyses were performed by Cell Line Genetics Laboratory (Madison, WI).

Figure S4

A

SNP location	Gene Symbol	SNP for A	SNP for B	hiPS-11 d0	hiPS-11 d50
3587020	PRKX	C	G	AA	AA
5917485	NLGN4X	A	C	BB	BB
8367345	KAL1	C	T	AA	AA
9249803	TBL1X	C	G	AA	AA
9304346	TBL1X	C	T	AA	AA
9305195	TBL1X	C	T	AA	AA
9344256	TBL1X	A	G	AA	AA
9407521	TBL1X	A	G	AA	AA
9710791	APXL	A	G	BB	BB
11935310	PDZK10	C	T	AA	AA
16459882	CTPS2	C	T	AA	AA
16619285	CXorf15	A	G	BB	AB
17319709	NHS	A	G	AA	AA
17478284	NHS	G	T	AA	AA
18826469	GPR64	A	C	BB	BB
19943694	RPS6KA3	A	G	BB	BB
28680456	IL1RAPL1	C	T	BB	BB
30944220	DMD	C	G	AA	AA
31014220	DMD	A	C	BB	BB
37066977	PRRG1	A	G	BB	BB
37648565	SYTL5	C	T	AA	AA
37666283	SYTL5	A	G	BB	BB
38257705	TSPAN7	A	C	BB	BB
40955880	DDX3X	A	C	BB	BB
44687951	UTX	C	G	BB	BB
46072828	ZNF673	C	G	AA	AA
46092073	ENST00000344153	A	G	BB	BB
46319179	SLC9A7	G	T	BB	BB
48189753	RBM3	C	G	AA	AA
53110364	SMCX	C	T	AA	AA
54081339	CXorf17	A	G	BB	BB
54081723	CXorf17	C	T	BB	BB
54400117	FGD1	A	G	BB	BB
65197638	HEPH	A	T	AA	AA
67060668	OPHN1	G	T	BB	BB
83432376	CXorf43	A	G	BB	BB
95820493	ENST00000324765	A	G	BB	BB
95822662	ENST00000324765	A	G	BB	BB
95825925	ENST00000324765	C	G	AA	AA
109277505	AMMECR1	G	T	AA	AA
117289500	WDR44	A	G	AA	AA
122235467	GRIA3	G	T	AA	AA
122765447	BIRC4	A	T	AA	AA
132190020	GPC4	A	T	BB	BB
134201782	ZNF75	C	G	BB	BB
135007255	FHL1	A	G	BB	BB
135018602	FHL1	A	G	AA	AA
147620149	AFF2	C	T	AA	AA
147625081	AFF2	C	T	AA	AA
148271977	IDS	A	G	BB	BB
149547780	MTMR1	A	G	BB	BB
151688320	NSDHL	A	G	AA	AA
152250923	TREX2	C	T	AA	AA

B

SNP location	Gene Symbol	SNP for A	SNP for B	hiPS-12 d0	hiPS-12 d50
9249803	TBL1X	C	G	AA	AA
9304346	TBL1X	C	T	AA	AA
9644975	APXL	A	G	BB	BB
9871899	KIAA1280	C	G	AA	AA
11935310	PDZK10	C	T	AA	AA
30944220	DMD	C	G	AA	AA
31014220	DMD	A	C	BB	BB
40955880	DDX3X	A	C	BB	BB
46072828	ZNF673	C	G	AA	AA
48189753	RBM3	C	G	AA	AA
54081339	CXorf17	A	G	BB	BB
71602712	PHKA1	C	T	BB	AA
117289500	WDR44	A	G	AA	AA
122720396	BIRC4	A	T	BB	BB
134201782	ZNF75	C	G	BB	BB
135018602	FHL1	A	G	AA	AA
149547780	MTMR1	A	G	BB	BB
152250923	TREX2	C	T	AA	AA

Color Code:		
	AA	monoallelic for A
	BB	monoallelic for B
	AB	Biallelic

Figure S4. Allele-specific expression analysis of X-linked genes for hiPS-11 (A) and hiPS-12 (B) at day 0 and day 50 of differentiation.

SNP-Chip analysis performed using X-linked SNPs identified in hiPS-11 (p. 34) and hiPS-12 (p.33). Nuclear RNA isolated from undifferentiated (d0) and differentiated (d50) cells and cDNA was converted to double-stranded cDNA and analyzed using Affymetrix 250K Nsp arrays. SNP position (annotated according to hg18 assembly), the corresponding gene for the SNP ('Gene Symbol'), and the actual nucleotide are shown. Only SNPs within genes that were assigned allelic expression call in matched undifferentiated and differentiated samples are shown. *Gold* denotes biallelic expression, and green colors denote monoallelic expression (*dark green* corresponds to allele 'A'; *light green* corresponds to allele 'B'). Related to Figure 1 (characterization of XIST expression in hiPSCs).

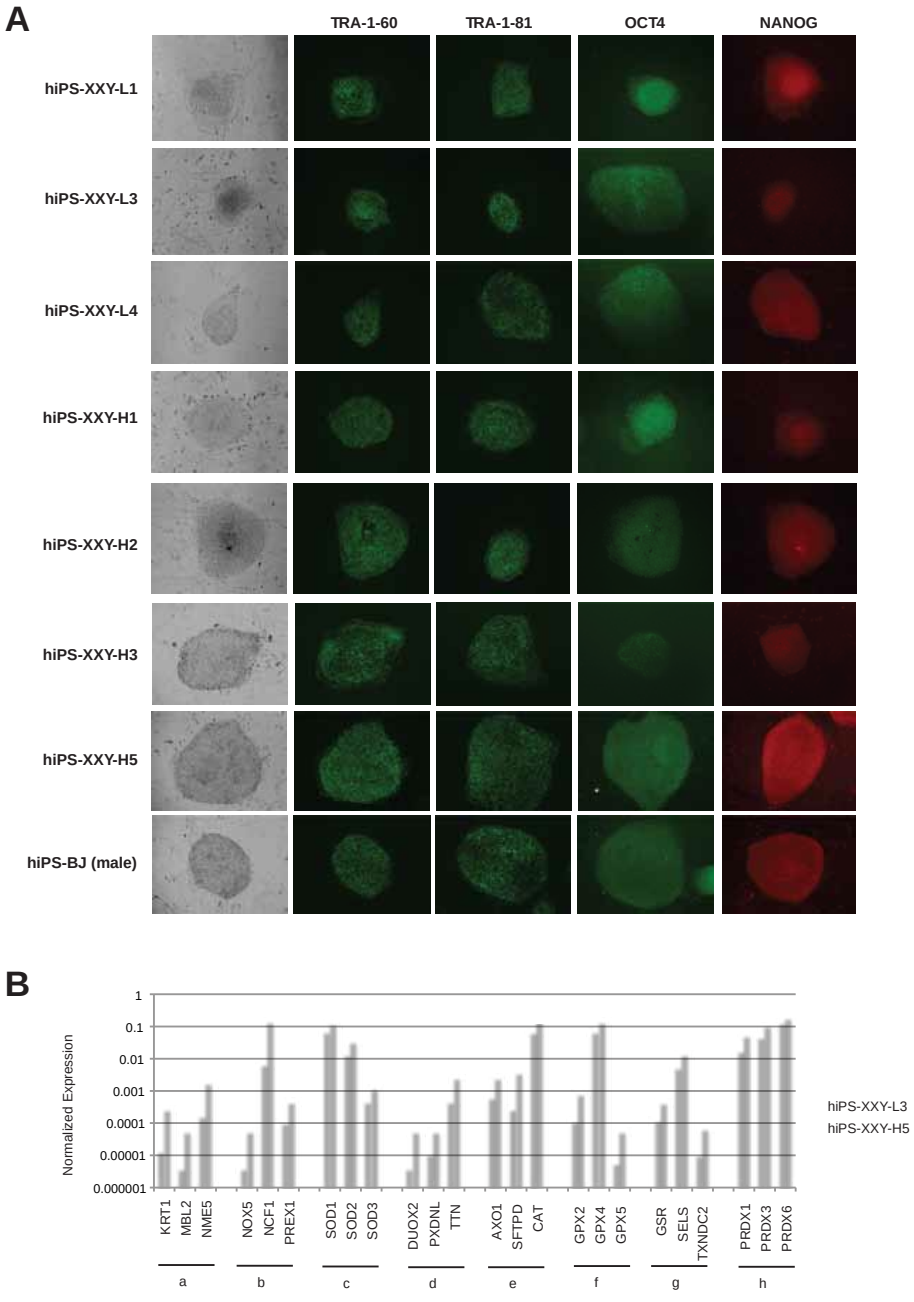

A
TRA-1-60
TRA-1-81
OCT4
NANOG
hiPS-XXY-L1
hiPS-XXY-L3
hiPS-XXY-L4
hiPS-XXY-H1
hiPS-XXY-H2
hiPS-XXY-H3
hiPS-XXY-H5
hiPS-BJ (male)
B
Normalized Expression
1
0.1
0.01
0.001
0.0001
0.00001
0.000001
KRT1
MBL2
NME5
NOX5
NCF1
PREX1
SOD1
SOD2
SOD3
DUOX2
PXDNL
TTN
AXO1
SFTPD
CAT
GPX2
GPX4
GPX5
GSR
SELS
TXNDC2
PRDX1
PRDX3
PRDX6
a
b
c
d
e
f
g
h
hiPS-XXY-L3
hiPS-XXY-H5

Figure S5

Figure S5. hiPSCs derived from Klinefelter Syndrome fibroblasts.

(A) Immunostaining for pluripotency markers TRA-1-60, TRA-1-81, OCT-4, and NANOG for the Klinefelter Syndrome hiPSCs lines reprogrammed in high (H) or low (L) oxygen. A male hiPSCs line derived from BJ fibroblasts was used as a positive control. Related to Figure 1F and Table I.

(B) qPCR analysis for early passage (p.6) lines reprogrammed in either high (hiPS-XXY-H5) or low oxygen (hiPS-XXY-L3) concentrations. Ct values were normalized to GAPDH, and y-axis values are in log scale. Shown are examples of genes with greatest expression difference based on gene groupings: oxidative stress response pathways (a), superoxide metabolism genes (b), superoxide dismutases (c), peroxidases (d), ROS metabolism (e), glutathione peroxidases (f), antioxidants (g), and peroxiredoxins (h).

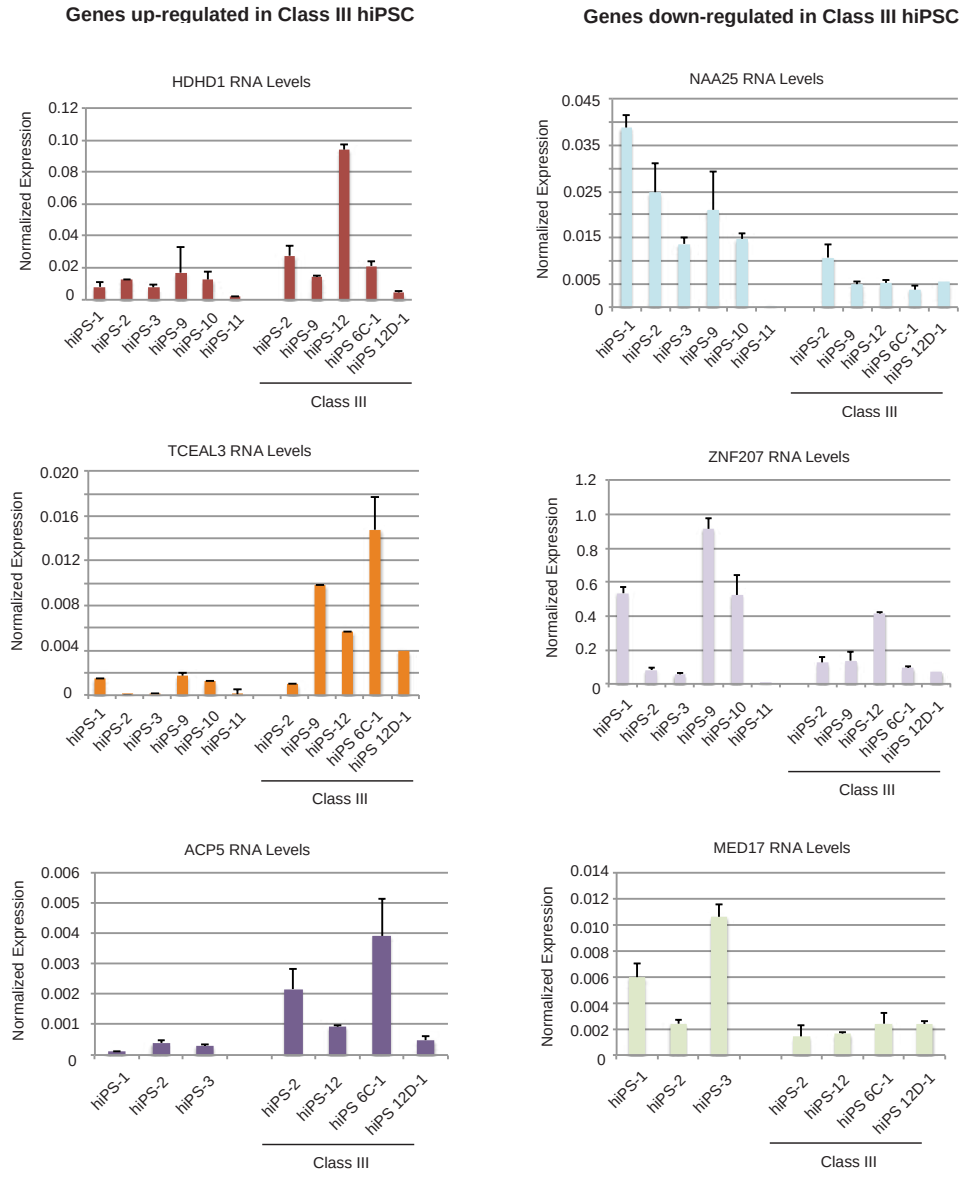

Genes up-regulated in Class III hiPSC
Genes down-regulated in Class III hiPSC
HDHD1 RNA Levels
NAA25 RNA Levels
TCEAL3 RNA Levels
ZNF207 RNA Levels
ACP5 RNA Levels
MED17 RNA Levels
Normalized Expression
hiPS-1
hiPS-2
hiPS-3
hiPS-9
hiPS-10
hiPS-11
hiPS-12
hiPS 6C-1
hiPS 12D-1
Class III

Figure S6

Figure S6. qRT-PCR validations for microarray results of genes up- and down-regulated in Class III hiPSCs.

qRT-PCR validations for Class II and Class III hiPSCs for three genes that were upregulated in Class III (HDHD1, TCEAL3, ACP5) and downregulated in Class III (NAA25, ZNF207, MED17; see Table III for complete list). Shown are representative triplicate measures (from 3-5 independent experiments) of Ct values that were normalized to GAPDH. Related to Figure 2.

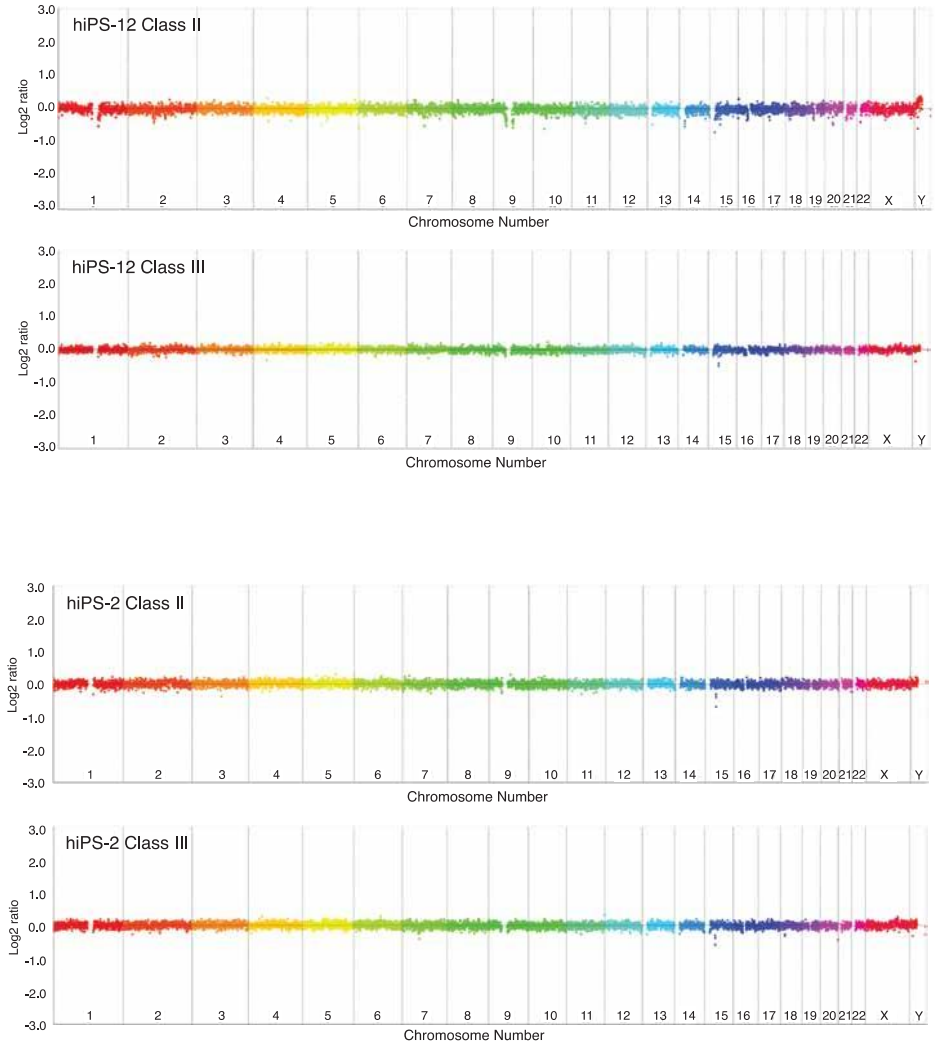

hiPS-12 Class II
hiPS-12 Class III
hiPS-2 Class II
hiPS-2 Class III
Log2 ratio
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y
Chromosome Number

Figure S7

Figure S7. Array comparative genomic hybridization (aCGH) analyses of matched Class II and Class III hiPSCs.

Genomic DNA from two sets of hiPSCss were analyzed (hiPS-2 p.41 and hiPS-2 c.III p.57; hiPS-12 p. 37 and hiPS-12 c.III p. 58) using the Roche Nimbelgen 4x72K Human Gene Expression array, with a resolution of 40 kb. The parental IMR-90 fibroblast cell line (genomic DNA) was used as a reference. Shown are the averaged values of the Log2 ratio for the sample divided by the reference (IMR-90) for all chromosomes. Related to Figure 4, indicating that increased cell growth rate is not due to genomic aberrations (insertion/deletions).

Table S-I: Effect of HDAC inhibition on XIST expression

	HDAC inhibitor	% nuclei with XIST clouds	
hiPS or hESC cell line		Day 0 (n)	Differentiation d17-21 (n)
hiPS-3	-	61% (386)	54% (273)
hiPS-3	+	65% (186)	64% (190)
hiPS-10	-	39% (258)	69% (97)
hiPS-10	+	51% (47)	60% (297)
hiPS-11	-	53% (81)	80% (287)
hiPS-11	+	56% (199)	78% (216)
hiPS-12	-	0.6% (160)	0 (72)
hiPS-12	+	0 (221)	0 (36)
hiPS 6C-1	-	0 (109)	0 (49)
hiPS 6C-1	+	0 (96)	0 (29)
hiPS 12D-1	-	0 (127)	0 (143)
hiPS 12D-1	+	0 (133)	0 (70)
HUES-9 (hESC)	-	15% (805)	58% (147)
HUES-9 (hESC)	+	37% (163)	86% (138)
H9 p.33 (hESC)	-	33% (144)	6% (235)
H9 p.36 (hESC)	+	8% (184)	32% (93)

Table S-IV: Complete list of microarray data analyzed from male and female hiPSC and female hESC lines.

hiPSCs (female)

Sample ID (GSM)	Collection (GSE)	Description	PubMed ID	Abbreviation used
GSM402717	GSE16093	Human Induced Pluripotent Stem Cells by Direct Delivery of Reprogramming Proteins (Line 1)	19481515	K11
GSM402752	GSE16093	Human Induced Pluripotent Stem Cells by Direct Delivery of Reprogramming Proteins (Line 2)	19481515	K11
GSM402806	GSE16093	Human Induced Pluripotent Stem Cells by retroviral transfection of Reprogramming genes	19481515	K11
GSM501007	GSE20033	minicircle-derived iPSC subclone 1 (from human adipose stem cells)	20139967	W12
GSM501008	GSE20033	minicircle-derived iPSC subclone 2	20139967	W12
GSM501009	GSE20033	minicircle-derived iPSC subclone 3	20139967	W12
GSM531015	GSE21262	undifferentiated induced pluripotent stem cell clone 3975.2	20887957	N2
GSM531016	GSE21263	undifferentiated induced pluripotent stem cell clone 3975.4	20887957	N2
GSM531017	GSE21264	undifferentiated induced pluripotent stem cell clone IMR9.4	20887957	N3
GSM649332	GSE26453	ESIMR90, biological rep1	unpublished	M3
GSM649333	GSE26453	ESIMR90, biological rep2	unpublished	M3
GSM553720	GSE22246	Reprogrammed iPSC line hiPSC pMIP2	20727844	P1
GSM553717	GSE22246	Reprogrammed iPSC line hiPSC 3975 (hiPSC 3975 line p24)	20727844	P2
GSM553719	GSE22246	Reprogrammed iPSC line hiPSC G	20727844	P1
GSM553718	GSE22246	Reprogrammed iPSC line hiPSC E	20727844	P1
GSM556998	GSE22392	hiPSC line IMR90 (p9)	20727844	P3
GSM556996	GSE22392	Reprogrammed iPSC line (hiPSC 3975 p6)	20727844	P2
GSM556997	GSE22392	Reprogrammed iPSC line (hiPSC 3975 p9)	20727844	P2

hESCs (female)

GSM239824	GSE14503	Human embryonic stem cell line T3ES	20735361	Sh8
GSM239825	GSE14503	Human embryonic stem cell line T3ES, biological rep2	20735361	Sh8
GSM239826	GSE14503	Human embryonic stem cell line T3ES, biological rep3	20735361	Sh8
GSM172579	GSE7234	human embryonic stem cells 1 (HES-20)	unpublished	X20
GSM172580	GSE7234	human embryonic stem cells 2; simple duplication karyotype	unpublished	X20
GSM172581	GSE7234	human embryonic stem cells 3; complex karytoype	unpublished	X20
GSM241167	GSE9510	H9 hESC_4% O2 rep1	18811242	E9
GSM241168	GSE9510	H9 hESC_20% O2 rep1	18811242	E9

hiPSCs (male)					
	GSM310838	GSE12390	BJ hIPS #5p9 sample 1	18786420	H4
	GSM310839	GSE12390	BJ hIPS #5p9 sample 2	18786420	H4
	GSM310844	GSE12390	BJ hIPS #5p9 sample 3	18786420	H4
	GSM310845	GSE12390	BJ hIPS #6p9 sample 1	18786420	H4
	GSM310846	GSE12390	BJ hIPS #6p9 sample 2	18786420	H4
	GSM310847	GSE12390	BJ hIPS #6p9 sample 3	18786420	H4
	GSM310848	GSE12390	BJ hIPS #8p10 sample 1	18786420	H4
	GSM310849	GSE12390	BJ hIPS #8p10 sample 2	18786420	H4
	GSM310850	GSE12390	BJ hIPS #8p10 sample 3	18786420	H4
	GSM310851	GSE12390	BJ hIPS #12p5 sample 1	18786420	H4
	GSM310852	GSE12390	BJ hIPS #12p5 sample 2	18786420	H4
	GSM310853	GSE12390	BJ hIPS #12p5 sample 3	18786420	H4
	GSM310857	GSE12390	BJ hIPS #12p6 afp 4 #12 p 7 sample 1	18786420	H4
	GSM310858	GSE12390	BJ hIPS #12p6 afp 4 #12 p 7 sample 2	18786420	H4
	GSM310859	GSE12390	BJ hIPS #12p6 afp 4 #12 p 7 sample 3	18786420	H4
	GSM248203	GSE9832	dH1f-iPS3-3 iPS cells	18157115	D6
	GSM248205	GSE9832	dH1cf16-iPS5 iPS cells_30	18157115	D6
	GSM248206	GSE9832	dH1cf16-iPS5 iPS cells_32	18157115	D6
	GSM248207	GSE9832	dH1cf32-iPS2 iPS cells_10	18157115	D6
	GSM248208	GSE9832	dH1cf32-iPS2 iPS cells_20	18157115	D6
	GSM248211	GSE9832	MRC5-iPS2 iPS cells_2	18157115	D5
	GSM248212	GSE9832	MRC5-iPS2 iPS cells_22	18157115	D5
	GSM248215	GSE9832	BJ1-iPS1 iPS cells	18157115	D4
	GSM579907	GSE23583	BJ_RiPS_1.1	20888316	R4
	GSM579908	GSE23583	BJ_RiPS_1.2	20888316	R4
	GSM579909	GSE23583	BJ_RiPS_1.3	20888316	R4
	GSM579913	GSE23583	MRC5_RiPS_1.8	20888316	R5
	GSM579914	GSE23583	MRC5_RiPS_1.9	20888316	R5
	GSM579915	GSE23583	MRC5_RiPS_1.11	20888316	R5

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

RNA/DNA FISH and Immunostaining

RNA and DNA FISH were carried out as described previously (Silva et al., 2008)(Erwin and Lee, 2010). Human XIST probe (exon 1)was labeled by nick translation with Cy3-dUTP, and Cot-1 DNA was labeled with fluoroscein-12-dUTP. Images were collected using a Nikon eclipse 90i microscope, and 0.2 um z-sections merged. StarFISH X chromosome paints (Cambio, Cambridge UK) were hybridized per manufacturer's instructions. For sequential RNA/DNA FISH, the RNA FISH was performed first, 0.2 um z-section images taken, and x-y coordinates marked. Slides were then denatured for DNA FISH and images merged with RNA FISH signals using DAPI nuclear staining as reference.

For immunofluorescence staining, hiPSCs were grown in 12-well plates, fixed with 4% PFA, permeabilized with 0.05% Trition X-100, blocked with 4% goat serum, and then incubated with primary antibody. Antibodies: DyLight 488 Human/mouse Oct4 (1:100, StemGent), StainAlive DyLight 488 Human Tra-1-60 (1:100, StemGent), StainAlive DyLight 488 Human Tra-1-81 (1:100, StemGent), and NANOG (1:750, Novus Biologicals).

Allele-Specific Expression Analyses

Allele-specific expression analysis was performed as described (Gimelbrant et al., 2007; Lengner et al., 2010). Nuclei from hiPS-11 p.34 (days 0 and 50) and hiPS-12 p.23 were isolated from 2-10 million cells. Total RNA isolated with Trizol and treated with Turbo DNAfree (Ambion). Genomic DNA was prepared with QIAamp DNA Blood Mini Kit (Qiagen). DNA and cDNA samples were further processed using standard protocols for the 250K Nsp Human Mapping Array (Affymetrix). After hybridization of one array with DNA and two

duplicates with cDNA, genotypes were called by the Affymetrix GTYPE software using Dynamic Model Mapping algorithm (DM). For DNA samples, all genotyping calls by DM were accepted. For cDNA samples, we accepted only genotypes that were replicated with DM confidence scores < 0.10. SNP coordinates for hg18 assembly.

Real-time PCR

Total RNA was purified with RNeasy Plus Micro Kit (QIAGEN) or cleaned with RNeasy MinElute Cleanup Kit (QIAGEN) after TRIZol purification. RNA was reverse transcribed into cDNA using RT^2 First Strand cDNA Synthesis Kit (SABiosciences). cDNA was applied to qBiomarker iPSC PCR Arrays or RT^2 Profiler PCR Arrays (SABiosciences) at 2 ng cDNA per well of the 384-well plate. Real-time PCR was performed on Roche LightCycler 480 using SYBR Green w/Fluorescein Master Mix (SABiosciences). Each sample was assayed three times and raw Ct values normalized to GAPDH. XIST primers: 5' - TTG CCC TAC TAG CTC CTC GGA C (exon 1 F), 5' - TTC TCC AGA TAG CTG GCA ACC (exon 3 R).

Bisulfite Sequencing

Genomic DNA was purified with DNeasy Blood & Tissue Kit (QIAGEN) and treated using the EpiTect Bisulfite Kit (QIAGEN) to convert unmethylated cytocine to uracil. The promoter regions of NANOG (NM_024865; -302 to -138), OCT4 (NM_002701; -2331 to -2111 and -6239 to -369) were amplified using specific primers, and treated templates amplified using HotStarTaq Plus Master Mix Kit (QIAGEN). Specific PCR products were gel-purified using the QIAquick Gel Extraction Kit (QIAGEN) and cloned using the TOPO TA Cloning Kit for

Sequencing (Invitrogen). At least 12 clones for each were sequenced by Mclab (http://www.mclab.com/home.php).

Supplemental Movies and Spreadsheets

Click here to download Supplemental Movies and Spreadsheets: Table S-II.xlsx

Supplemental Movies and Spreadsheets

Click here to download Supplemental Movies and Spreadsheets: Table S-III.xlsx