

תוכנית מחקר המוגשת לאישור כתכנית לעבודת דוקטור

Student's name: Uri Ben-David

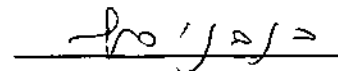
שם התלמיד: אורי בן-דוד

שם המדריך: פרופ' נסים בנבניסטי

יציבות גנומית וגידוליות של תאי גזע פלוריפוטנטיים

Genomic stability and tumorigenicity
of pluripotent stem cells

הריני מאשר את הנושא והתוכנית, ומסכים להדריך את המועמד בביצוע עבודה זו.



פרופ' נסים בנבניסטי

Abstract

The unique abilities of human embryonic stem cells (HESCs) to self-renew and to differentiate into cells of the three germ layers make them an invaluable tool for the future of regenerative medicine. However, the same properties also make them tumorigenic, and therefore hinder their clinical application. The tumorigenicity of HESCs may be increased by their tendency to acquire chromosomal aberrations during their culture propagation. The reprogramming of human somatic cells into induced pluripotent stem cells (HiPSCs) has revolutionized the field of pluripotent stem cells, as these cells may possibly solve the problem of immune rejection in cell replacement therapy. To date, the tumorigenicity of HiPSCs, their genomic stability in culture, and their similarity to HESCs in these regards, have not been sufficiently addressed. In addition, practical and efficient techniques to eliminate residual HESCs and HiPSCs from their differentiated progeny must be developed in order to decrease the tumorigenicity risk of these cells. In my work, I have developed a bioinformatic methodology to comprehensively examine the genomic integrity of stem cells. Using this novel methodology, I have studied the genomic integrity of HiPSCs and compared it to that of HESCs, in order to identify similarities and differences between these cells types; I have further examined the genomic integrity of human adult stem cells, and thus compared the genomic stability of stem cells at distinct developmental stages. I am also comparing the genomic integrity of mouse and human pluripotent stem cells, a comparison through which novel conclusions could potentially be drawn with regard to the human cells. Next, I would like to examine whether the characteristic chromosomal aberrations that arise in HESCs and HiPSCs make them more tumorigenic, and whether these aberrations shift the global gene expression pattern of these cells to a more cancer-like expression pattern. Lastly, I would like to develop new strategies to robustly eliminate residual undifferentiated cells from cultures of differentiated cells. I intend to achieve the last goal by implementing two strategies: 1. Identifying cell surface molecules that are unique to pluripotent stem cells, and targeting the cells that express them. 2. Performing a high-throughput screen to identify small molecules that are selectively cytotoxic towards human pluripotent stem cells. This work could have broad implications for the future use of HESCs and HiPSCs, both in basic science and in regenerative medicine.

Scientific background

Human embryonic stem cells (HESCs) are endowed with two unique properties: self-renewal, that is the ability to proliferate indefinitely while maintaining their cellular identity; and pluripotency, that is the ability to differentiate into all of the cell types that comprise the embryo proper. These traits make HESCs promising for future regenerative medicine, but the very same traits also make them tumorigenic, and consequently hinder the clinical fulfillment of their potential. Thus, HESCs could be aptly described using the metaphor of a "double-edged sword", as their defining characteristics make them both powerful and dangerous at the same time.

HESCs share cellular and molecular phenotypes with tumor cells and cancer cell lines ¹⁻³. Among these are rapid proliferation rate ⁴, lack of contact inhibition ⁵, a propensity for genomic instability ^{6,7}, as well as high activity of telomerase ⁸, high expression of onco-fetal genes such as *MYC* ⁹ and *KLF4* ¹⁰, and remarkable similarities in their overall gene expression patterns ¹¹⁻¹³, miRNA signatures ¹⁴, and epigenetic status ¹⁵. When injected into immunodeficient mice, HESCs form teratomas ⁵ — benign tumors which are composed of differentiated tissues from all three germ layers. These tumors are so characteristic of HESCs that they have become the most stringent test for pluripotency in human cells. Indeed, treatment attempts with embryonic stem cells in animal models were shown to be fatal due to the formation of teratoma-like tumors ¹⁶.

The above-mentioned features apply to normal diploid HESCs, however HESCs are also known to acquire chromosomal aberrations in culture ⁶. Chromosomal aberrations are a hallmark of cancer, and several studies have demonstrated that culture-adapted HESCs can generate more aggressive tumors, which might be classified as teratocarcinomas, the malignant counterparts of teratomas ¹⁷⁻²⁰. Teratocarcinomas are also composed of a differentiated tissue mixture of the three germ layers, but they contain foci of completely undifferentiated cells, called embryonal carcinoma (EC) cells, and are thus highly malignant.

Regardless of the malignancy risk, even the formation of benign teratomas on transplantation of HESC-derived cells into human patients, is highly alarming and would be unacceptable. Therefore, the tumorigenicity of HESCs is a major hurdle, which must be confronted before achievements of this field of research can be safely translated into the clinic.

The discovery of human induced pluripotent stem cells (HiPSCs) ²³⁻²⁵, has revolutionized the field of pluripotent stem cells (PSCs) for two main reasons: they changed the perception of cellular reprogramming, showing the plasticity of somatic cells is much greater than had previously been thought; and they offered a very appealing solution to the likely immune-rejection of HESC-derived cells on their transplantation into an un-matched patient, thus providing new and exciting avenues for patient-specific cell therapy.

While constituting a huge leap towards overcoming the immunogenicity obstacle, the translation of HiPSCs into the clinic faces the same substantial tumorigenicity problem as that of HESCs. Sharing with HESCs their basic properties of self-renewal and pluripotency, HiPSCs are doomed to share with them the other "edge of the sword". Indeed, HiPSCs

exhibit the cellular and molecular phenotypes that make HESCs resemble cancer cells, and form benign teratomas on their injection into immunodeficient mice.

However, HESCs and HiPSCs (as well as their mouse counterparts) are not identical, and a rapidly accumulating body of work suggests considerable differences between these two pluripotent cell types, encompassing crucial aspects such as global gene expression ²⁶, epigenetic landscape ²⁷⁻²⁹ and genomic imprinting ³⁰. Special attention should therefore be paid to exploring the genomic stability and the tumorigenicity of HiPSCs., and to developing new strategies to efficiently eliminate both HESCs and HiPSCs from their differentiated progeny.

Research aims

A. Identification and characterization of chromosomal aberrations in HiPSCs:

- 1) Developing a novel methodology for detection of chromosomal aberrations by gene expression profiling.
- 2) Validating the methodology using known aberrations in HESCs.
- 3) Identifying, quantifying and classifying the chromosomal aberrations in HiPSCs.
- 4) Assessing the rate of culture adaptation in HiPSCs.
- 5) Identifying specific genes that confer selective advantage to recurrent aberrations.

B. Large-scale analysis of chromosomal aberrations in human adult stem cells:

- 1) Adjusting the abovementioned methodology to various types of multipotent stem cells.
- 2) Analyzing chromosomal aberrations in human neural stem cells (NSCs), mesenchymal stem cells (MSCs) and hematopoietic stem cells (HSCs).
- 3) Comparing the aberrations that arise in stem cell cultures to those that occur in-vivo during tumorigenesis.

C. Comparison of chromosomal aberrations between mouse and human pluripotent stem cells:

- 1) Adjusting the abovementioned methodology to mouse pluripotent stem cells.
- 2) Analyzing chromosomal aberrations in mouse ESCs and iPSCs.
- 3) Comparing the common aberrations between mouse and human PSCs.

4) Gene expression analysis in recurrent conserved aberrations, in order to identify putative genes that confer selective advantage to these aberrations.

D. Studying the functional implications of characteristic aberrations to the tumorigenicity of human HESCs and HiPSCs:

1) Comparing aneuploid PSCs with typical aberrations to normal diploid PSCs, both in-vitro and in-vivo.

2) Gene expression analysis of diploid and aneuploid PSCs and their comparison to embryonic carcinoma (EC) cells.

E. Elimination of residual undifferentiated cells from differentiated cultures by targeting pluripotent-specific cell surface proteins:

1) Identifying cell-surface proteins unique to PSCs, compared to all of the cell types in the adult body.

2) Characterizing of the role of the identified protein/s in PSCs.

3) Developing a strategy to specifically eliminate PSCs from differentiated cultures based on the expression of the identified protein/s.

4) In-vivo studies to demonstrate the usefulness of this strategy for teratoma prevention.

F. Identification and characterization of highly-selective inhibitors of HESCs and HiPSCs by high-throughput screening:

1) Assay development for screening HESCs and HiPSCs in a 384-well format.

2) Primary screen: measuring the effect of ~52,500 small molecules on the viability of HESCs.

3) Confirmation and validation screens: verifying the positive “hits” from the primary screen with various HESC and HiPSC lines at multiple concentrations.

4) Counter-screen with various PSC-derived cell types of all germ layers and developmental stages in order to identify highly-selective inhibitors of PSCs.

5) Attempts to identify the pathways and/or molecular targets of highly-selective inhibitors of HESCs and HiPSCs.

6) In-vitro and in-vivo studies to demonstrate the usefulness of the selective inhibitors for teratoma prevention.

Preliminary results:

A. Identification and characterization of chromosomal aberrations in HiPSCs:

1) Developing a novel methodology for detection of chromosomal aberrations by gene expression profiling: Two bioinformatic tools were applied. In the first test, the list of upregulated (>1.5 fold) or downregulated (<0.5) genes was submitted to a location enrichment analysis (using available gene expression analysis tools, such as Expander and EASE). In the second test, gene expression data is applied to a CGH analysis software (CGH-explorer), enabling recognition of regionally-biased over- or under-expression. The bioinformatic tests were selected, scaled and adjusted, in order to minimize the false positive and false negative rates of the analysis, when compared to high-density SNP-array copy number analysis and/or karyotyping conducted on the same cells (see Figure 1).

2) Validating of the methodology with known aberrations in HESCs: 38 HESC lines were analyzed, and 12 of them were found to harbor chromosomal aberrations. Known aberrations were correctly detected, and novel ones were identified. In line with previous reports, the most recurrent aberrations were duplications in chromosomes 12 and 17 (see Figure 1 and Figure 4a).

3) Identifying, quantifying and classifying the chromosomal aberrations in HiPSCs: 66 HiPSC lines were analyzed, and 13 of them were found to be abnormal (~20%), of which 6 carried at least one full trisomy. Aberrations were also found in lines previously reported to be normal. Trisomy 12 was found to be the most prevalent in HiPSCs (see Figure 2 and Figure 4a). Three distinct sources of aberrations were characterized: aberrations inherited from the somatic cell of origin; aberrations acquired during or immediately after the reprogramming process; and aberrations acquired during the prolonged passaging of the cells. Examples of the three categories were identified (see Figure 2 and Figure 4f).

4) Assessing the rate of culture adaptation in HiPSCs: In several cell lines, gene expression data were available from multiple passages. HiPSCs were karyotyped and their gene expression was profiled at multiple passages, revealing that the cells could acquire chromosomal aberrations rapidly – as few as five passages were needed for trisomy 12 to take over the culture (see Figure 3).

5) Identifying specific genes that confer selective advantage to recurrent aberrations: A minimal recurrent region was identified in chromosome 12. Inside this region reside the hallmark pluripotency genes, NANOG and GDF3. The expression of these genes was measured with qPCR at several time points, and was found to correlate with the establishment of the trisomy in the culture. The findings suggest that NANOG and GDF3 confer a selective growth advantage to the aberrant cells. (see Figure 4).

B. Large-scale analysis of chromosomal aberrations in human adult stem cells:

1) Adjusting the methodology to various types of multipotent stem cells: The bioinformatic methodology to detect chromosomal aberrations based on gene expression profiling, was adjusted using known aberrations in various stem cell cultures. Solid parameters and detection resolution were determined, so that the false positive rate of the methodology would be negligible.

2) Analyzing chromosomal aberrations in human neural stem cells (NSCs), mesenchymal stem cells (MSCs) and hematopoietic stem cells (HSCs): 208 samples of pluripotent stem cells (PSCs), 144 samples of mesenchymal stem cells (MSC), 97 samples of neural stem cells (NSCs), and 177 samples of hematopoietic stem cells (HSCs) were analyzed. All stem cell types that are expanded *in-vitro*, i.e. neural and mesenchymal stem cells, were found to acquire large chromosomal aberrations at a similar, or somewhat lower, frequency to that seen in PSCs. These aberrations were found to accumulate rapidly, within a few passages in culture. Importantly, the acquired chromosomal abnormalities are specific for each of the stem cell types, suggesting that different aberrations confer selection advantage to different stem cells (see Figure 5 and Figure 6a-c).

3) Comparing the aberrations that arise in stem cell cultures to those that occur *in-vivo* during tumorigenesis: Data were gathered from thousands of tumors of the same tissues as the various stem cells analyzed, using a well-established database of chromosomal aberrations in cancer. A calculation of the frequency of chromosomal aberrations in these tumors revealed a partial correlation between the lineage-specific aberrations that arise in stem cell cultures and the ones most common in tumors of the respective tissue. (Figure 6d-f). These results demonstrate that specific aberrations confer selective advantage in a cell lineage-specific manner, both to stem cells *in-vitro* and to tumors *in-vivo*

C. Comparison of chromosomal aberrations between mouse and human pluripotent stem cells:

1) Adjusting the abovementioned methodology to mouse pluripotent stem cells: The bioinformatic methodology to detect chromosomal aberrations based on gene expression profiling, was adjusted using known aberrations in mouse ESCs and iPSCs. Solid parameters and detection resolution were determined so that the false positive rate of the methodology would be negligible

2) Analyzing chromosomal aberrations in mouse PSCs: 129 samples of mouse ESCs, 127 samples of mouse iPSCs, 25 samples of mouse ESCs derived from iPSCs (through nuclear transfer), and 44 samples of epiblast stem cells (EpiSCs) were subjected to analysis. High prevalence of chromosomal aberrations was identified in both ESCs (~38%) and iPSCs (~23%). Similar rate of aberrations was found in multiple ESC lines of various mouse strains, and in iPSCs of various cellular origins and derivation techniques. Four hotspots of chromosomal aberrations were identified: full trisomy 11 (with a minimally recurrent gain in 11qE2), full trisomy 8, and deletions in chromosomes 10qB and 14qC-14qE (see Figures 7 and 8).

D. Studying the functional implications of characteristic aberrations to the tumorigenicity of human HESCs and HiPSCs:

1) Comparing aneuploid PSCs with typical aberrations to normal diploid PSCs, both in-vitro and in-vivo: The HiPSC cell line hiPS18 was karyotyped at early and late passages, and was found to harbor trisomy 12 at the later passage. Cells from both passages were injected to NOD-SCID mice, and teratomas were collected 4 weeks later. OCT4 expression was measured in the teratomas, using qPCR and immunostaining. Only teratomas from cells at the later passage (with trisomy 12) were found to express high levels of OCT4. These results suggest that trisomy 12 indeed makes the cells more tumorigenic (see Figure 9).

E. Elimination of residual undifferentiated cells from differentiated cultures by targeting pluripotent-specific cell surface proteins:

1) Identifying cell-surface proteins unique to PSCs, compared to all of the cell types in the adult body: Expression arrays were compared between HESCs, HiPSCs, teratomas, embryoid bodies (EBs) and 12 differentiated tissues – looking for genes that are highly expressed in PSCs, downregulated in teratomas and EBs, and not expressed at all in all differentiated tissues. Of the identified genes, only those that are known to be expressed on the surface of the cell (by GO annotation) were selected. Two such genes were identified: TDGF1 and CLDN6. TDGF1 is a known hallmark pluripotency gene, while CLDN6 has never been studied in PSCs. Thus, CLDN6 was selected for further study. Its high and selective expression was demonstrated at the RNA level (using qPCR), and at the protein level (using FACS and immunostaining) (see Figure 10).

F. Identification and characterization of highly-selective inhibitors of HESCs and HiPSCs by high-throughput screening:

1) Assay development for screening HESCs and HiPSCs in a 384-well format: CSES2 HESCs were grown on matrigel with mTeSR defined medium (“feeder-free”). Their pluripotency at these conditions was verified by morphology and by alkaline phosphatase (AP) staining. Cells were then grown at these conditions in 384-well plates. The cells remained undifferentiated, as was validated by immunostaining to OCT4 and NANOG, and by measurement of the red fluorescence of CSES2-SO2 cells. These cells were genetically-manipulated to carry mCherry under the promoter of OCT4, and are thus red when undifferentiated. A pilot screen of 50 known cytotoxic compounds was performed in order to select compounds that can use as positive controls (i.e. compounds that kill all cells in culture) and in order to determine the best time course for the primary screen. Following this pilot screen, positive controls were selected, and exposure time of 24 hours was determined (see Figure 11).

2) Primary screen: measuring the effect of ~52,500 small molecules on the viability of HESCs: HESCs were screened, using a “golden library” of ~52,500 small molecules, representing over 1 million compounds from the compound bank of the pharmaceutical company Roche. Screening was performed as following: CSES2 HESCs were grown on matrigel-coated plates with mTeSR defined medium. Before their transfer to 384-well plates, pluripotency was verified by morphology and alkaline phosphatase staining. Cells were harvested, counted, and dispensed in 384-well plates, using the Wellmate machine, so that 5,000 cells were plated in each well. Plates were incubated overnight (~24h) to allow the cells to settle down. The next day, compounds were added to the assay plates, together with positive and negative controls, using the Biomek FX liquid handling machine. 24 hours later, the number of living cells was quantified using the CellTiter-Glo viability assay. This is an ATP-luciferase-based assay, which generates luminescence that correlates accurately to the number of living cells in the well. The luminescence from each well was read using the Envision luminator. Raw data was normalized to internal controls using the Genedata software. The Z' factor, which is an accepted index for screen quality, was calculated for each plate. “Hits” were determined as compounds that caused over 60% inhibition. Using this cut-off, 2,031 compounds were selected to be re-tested in confirmation screens (see Figure 12).

3) Confirmation and validation screens: verifying the positive “hits” from the primary screen with various HESC lines at multiple concentrations: In order to confirm the “hits” from the primary screen, three HESC lines (CSES2, CSES2-SO2 and H9) and one iPSC line (iPS-BJ28) were screened in triplicates. The CSES2-SO2 cells were also subjected to a high-content screen (HCS), in which the level of red fluorescence was quantified in each well by microscopy imaging. 696 compounds were confirmed to be genuine inhibitors of pluripotent cells. These confirmed hits were screened again at multiple (6-8) concentrations in order to obtain their IC50 values (the concentration in which each compound causes 50% inhibition). Very high correlation values were observed within the HESC lines, as well as between the HESCs and HiPSCs (see Figure 13).

4) Counter-screen with various PSC-derived cell types of all germ layers and developmental stages: HeLa cells were screened with the 696 “hits”, as a first selectivity screen. ~200 compounds were found to have no significant effect on HeLa cells, although they dramatically reduce the viability of PSCs. These results are encouraging regarding the chances to identify selective inhibitors of PSCs (see Figure 13).

Research plan:

A. Identification and characterization of chromosomal aberrations in HiPSCs:

1) Developing a novel methodology for detection of chromosomal aberrations by gene expression profiling. This part was completed.

- 2) **Validating the methodology with known aberrations in HESCs.** This part was completed.
- 3) **Identifying, quantifying and classifying chromosomal aberrations in HiPSCs.** This part was completed.
- 4) **Assessing the rate of culture adaptation in HiPSCs.** This part was completed
- 5) **Identifying specific genes that confer selective advantage to recurrent aberrations.** This part was completed.

B. Large-scale analysis of chromosomal aberrations in human adult stem cells:

- 1) **Adjusting the methodology to various types of multipotent stem cells.** This part was completed.
- 2) **Analyzing chromosomal aberrations in human neural stem cells (NSCs), mesenchymal stem cells (MSCs) and hematopoietic stem cells (HSCs).** This part was completed.
- 3) **Comparing the aberrations that arise in stem cell cultures to those that occur in-vivo during tumorigenesis.** This part was completed.

C. Comparison of chromosomal aberrations between mouse and human pluripotent stem cells:

- 1) **Adjusting the abovementioned methodology to mouse pluripotent stem cells.** This part was completed.
- 2) **Analyzing chromosomal aberrations in mouse ESCs and iPSCs.** This part was completed.
- 3) **Comparing the common aberrations between mouse and human PSCs.** I intend to examine the synteny between the common chromosomal aberrations identified in mouse and human ESCs and iPSCs, in order to determine whether these aberrations are evolutionarily-conserved or species-specific.
- 4) **Gene expression analysis in recurrent conserved aberrations in order to identify putative genes that confer selective advantage to these aberrations.** I plan to compare gene expression between syntenic aberrations, in order to identify genes that are consistently over-expressed in the aberrant vs. the normal cell lines, both in mouse and in human. Such genes, if identified, are putative candidates to be the “driving force” that underlies the recurrence of the aberrations.

D. Studying the functional implications of characteristic aberrations to the tumorigenicity of human HESCs and HiPSCs:

1) Comparing aneuploid PSCs with typical aberrations to normal diploid PSCs, both in-vitro and in-vivo. I intend to compare HESCs and HiPSCs that acquired typical chromosomal aberrations (trisomy 12 and/or trisomy 17) to normal controls. I will compare the proliferation rate, differentiation propensity, apoptosis rate and cell cycle between diploid and aneuploid cells, in order to identify the mechanism/s that allow the aberrant cells to rapidly take over the culture. I will also inject the cells to immunodeficient mice and compare the formed teratomas, in order to examine whether the aberrant cells indeed generate more aggressive teratomas.

2) Gene expression analysis of diploid and aneuploid PSCs and their comparison to embryonic carcinoma (EC) cells. I intend to analyze the expression patterns of aberrant and normal PSCs, and compare them to those of EC cells. I would like to examine whether aberrant PSCs are more closely related (gene-expression wise) to EC cells than normal PSCs, and to identify specific genes that are typically expressed only in EC cells and not in PSCs, but are upregulated in aberrant PSCs.

E. Elimination of residual undifferentiated cells from differentiated cultures by targeting pluripotent-specific cell surface proteins:

1) Identifying cell-surface proteins unique to PSCs, compared to all of the cell types in the adult body. This part was completed.

2) Characterizing of the role of the identified protein/s in PSCs. The role of CLDN6 in PSCs will be examined by its knockdown, using siRNAs. Cells will be treated either with siRNA against CLDN6 or with mock siRNA. The proliferation rate, differentiation propensity and apoptosis of the treated cells will be compared. If CLDN-6 is required for the pluripotency or for the survival of PSCs, the knockdown will result in dramatic effects.

3) Developing a strategy to specifically eliminate PSCs from differentiated cultures based on the expression of the identified protein/s. Various techniques to target CLDN-6 expressing cells will be examined, such as: FACS-based cell sorting of CLDN-6 positive and negative cells from a mixed population; cytotoxic antibodies against CLDN-6; and chemical agents that specifically target CLDN-6.

4) In-vivo studies to demonstrate the usefulness of this strategy for teratoma prevention: Once a strategy to eliminate PSCs from differentiated cultures based on their CLDN6 expression will be established, its efficiency in preventing teratoma formation will be examined in an in-vivo setting. Mixed populations of undifferentiated and differentiated cells will be targeted, and subsequently injected into NOD-SCID mice. The generation of teratomas will be compared to that of untreated control cells.

F. Identification and characterization of highly-selective inhibitors of HESCs and HiPSCs by high-throughput screening:

1) Assay development for screening HESCs and HiPSCs in a 384-well format. This part was completed.

2) Primary screen: measuring the effect of ~52,500 small molecules on the viability of HESCs. This part was completed.

3) Confirmation and validation screens: verifying the positive “hits” from the primary screen with various HESC lines at multiple concentrations. This part was completed.

4) Counter-screen with various PSC-derived cell types of all germ layers and developmental stages in order to identify highly-selective inhibitors of PSCs: In order to find selective inhibitors of PSCs, the 696 compounds that had a dramatic cytotoxic effect on various HESCs and HiPSCs, will now be screened on various other cell types. Most of these cell types be differentiated from PSCs, and they should represent all the germ layers in multiple developmental stages. Among the cell lines I intend to screen: HESC-derived neural stem cells, HESC-derived mesenchymal stem cells, HESC-derived endodermal progenitor cells, HESC-derived hepatocytes, HiPSC-derived cardiomyocytes, BJ-fibroblasts, and various cancer cell lines. I will then look for compounds that are strong inhibitors of PSCs, but have a very mild effect – if any – on all other cell types.

5) Attempts to identify the pathways and/or molecular targets of highly-selective inhibitors of HESCs and HiPSCs: Once specific inhibitors of PSCs will be found, I will try to characterize the pathways in which they are involved and/or their molecular targets, using an approach called toxicogenomics. RNA will be derived from the cells after treatment with the selective inhibitors, but before their complete elimination. Using gene expression microarrays, I try to identify the pathway in which the molecule interfered. I will also examine the chemical structures of the compounds, try to identify them in-silico, and will gather information on them using bioinformatic databases of small molecules. If a prediction for the activity of a compound is found, it will be tested empirically.

6) In-vitro and in-vivo studies to demonstrate the usefulness of the selective inhibitors for teratoma prevention: The selective compounds will be assessed for their tumorigenicity prevention potential, both in-vitro and in-vivo. After elimination of undifferentiated cells from mixed populations using the compounds, the cultures will be FACS-analyzed for the presence of pluripotent cells, and will also be injected into immunodeficient mice for the purpose of teratoma assays.

References

1. Dreesen, O. & Brivanlou, A.H. Signaling pathways in cancer and embryonic stem cells. *Stem Cell Rev* **3**, 7-17 (2007).
2. Knoepfler, P.S. Deconstructing stem cell tumorigenicity: a roadmap to safe regenerative medicine. *Stem Cells* **27**, 1050-6 (2009).
3. Blum, B. & Benvenisty, N. The tumorigenicity of human embryonic stem cells. *Adv Cancer Res* **100**, 133-58 (2008).
4. Amit, M. et al. Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Dev Biol* **227**, 271-8 (2000).
5. Thomson, J.A. et al. Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145-7 (1998).
6. Baker, D.E. et al. Adaptation to culture of human embryonic stem cells and oncogenesis in vivo. *Nat Biotechnol* **25**, 207-15 (2007).
7. Harrison, N.J., Baker, D. & Andrews, P.W. Culture adaptation of embryonic stem cells echoes germ cell malignancy. *Int J Androl* **30**, 275-81; discussion 281 (2007).
8. Hiyama, E. & Hiyama, K. Telomere and telomerase in stem cells. *Br J Cancer* **96**, 1020-4 (2007).
9. Eilers, M. & Eisenman, R.N. Myc's broad reach. *Genes Dev* **22**, 2755-66 (2008).
10. Evans, P.M. & Liu, C. Roles of Krupel-like factor 4 in normal homeostasis, cancer and stem cells. *Acta Biochim Biophys Sin (Shanghai)* **40**, 554-64 (2008).
11. Sperger, J.M. et al. Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proc Natl Acad Sci U S A* **100**, 13350-5 (2003).
12. Ben-Porath, I. et al. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat Genet* **40**, 499-507 (2008).
13. Wong, D.J. et al. Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell Stem Cell* **2**, 333-44 (2008).
14. Neveu, P. et al. MicroRNA profiling reveals two distinct p53-related human pluripotent stem cell states. *Cell Stem Cell* **7**, 671-81 (2010).
15. Calvanese, V. et al. Cancer genes hypermethylated in human embryonic stem cells. *PLoS One* **3**, e3294 (2008).
16. Bjorklund, L.M. et al. Embryonic stem cells develop into functional dopaminergic neurons after transplantation in a Parkinson rat model. *Proc Natl Acad Sci U S A* **99**, 2344-9 (2002).
17. Herszfeld, D. et al. CD30 is a survival factor and a biomarker for transformed human pluripotent stem cells. *Nat Biotechnol* **24**, 351-7 (2006).
18. Yang, S. et al. Tumor progression of culture-adapted human embryonic stem cells during long-term culture. *Genes Chromosomes Cancer* **47**, 665-79 (2008).
19. Werbowetski-Ogilvie, T.E. et al. Characterization of human embryonic stem cells with features of neoplastic progression. *Nat Biotechnol* **27**, 91-7 (2009).
20. Blum, B. & Benvenisty, N. The tumorigenicity of diploid and aneuploid human pluripotent stem cells. *Cell Cycle* **8**, 3822-30 (2009).
21. Martin, G.R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* **78**, 7634-8 (1981).
22. Shih, C.C., Forman, S.J., Chu, P. & Slovak, M. Human embryonic stem cells are prone to generate primitive, undifferentiated tumors in engrafted human fetal tissues in severe combined immunodeficient mice. *Stem Cells Dev* **16**, 893-902 (2007).
23. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-76 (2006).
24. Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-72 (2007).

25. Yu, J. et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917-20 (2007).
26. Chin, M.H. et al. Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* **5**, 111-23 (2009).
27. Urbach, A., Bar-Nur, O., Daley, G.Q. & Benvenisty, N. Differential modeling of fragile X syndrome by human embryonic stem cells and induced pluripotent stem cells. *Cell Stem Cell* **6**, 407-11 (2010).
28. Polo, J.M. et al. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat Biotechnol* **28**, 848-55 (2010).
29. Kim, K. et al. Epigenetic memory in induced pluripotent stem cells. *Nature* **467**, 285-90 (2010).
30. Pick, M. et al. Clone- and gene-specific aberrations of parental imprinting in human induced pluripotent stem cells. *Stem Cells* **27**, 2686-90 (2009).

Figures

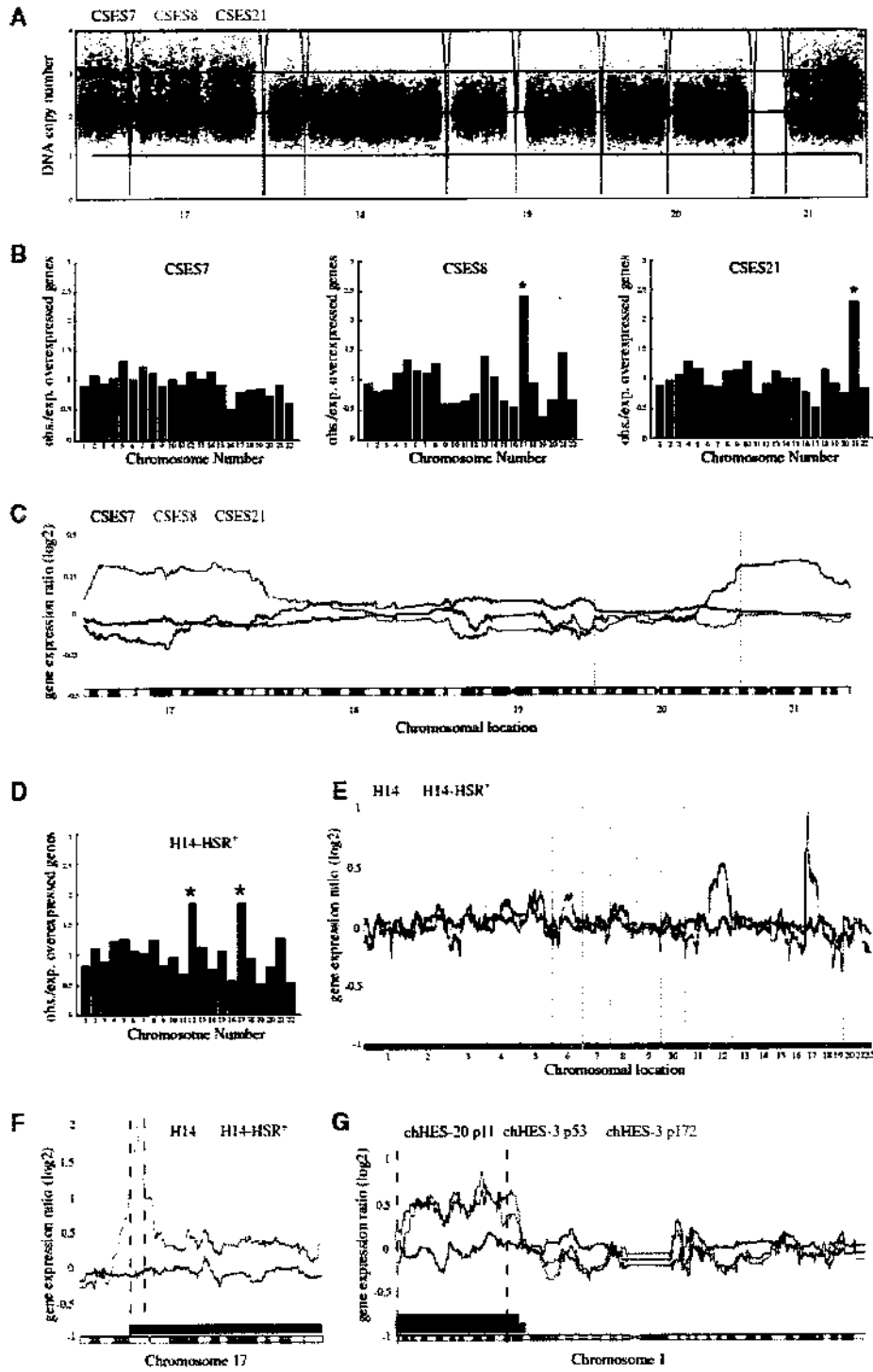


Figure 1: Identification of chromosomal aberrations in HESCs using gene expression analysis.

Identification of aneuploidy in pre-implantation genetic screening (PGS) derived HESC lines: (A) High density SNP-array copy number analysis of: CSES7 (blue – normal karyotype); CSES8 (green – 47XX+17); CSES21 (red – 47XY+21). (B) Whole chromosome gain analysis of over-expressed genes. Bars represent fold enrichment of over-expressed genes in each particular chromosome relative to the expected random frequency. CSES8 shows significant enrichment of over-expressed genes in chromosome 17 (Bonferroni corrected p-value = 1×10^{-7} by Expander and 2.1×10^{-9} by EASE), CSES21 shows significant enrichment in chromosome 21 (Bonferroni corrected p-value = 4.9×10^{-4} by Expander and 1.1×10^{-10} by EASE). (C) Gene expression profile moving average plot demonstrates homogenous over-expression of genes along the abnormal chromosomes.

Identification of aneuploidy in culture adapted HESCs: (D) Identification of trisomies 12 and 17 in the H14-HSR⁺ cell line (Baker et al., 2007) by whole chromosome analysis, showing a significant enrichment in chromosomes 12 and 17 (Bonferroni corrected p-value = 5.5×10^{-18} by Expander and 3.6×10^{-18} by EASE, for chromosome 12; 2.6×10^{-18} by Expander and 1.5×10^{-12} by EASE, for chromosome 17). This enrichment was not found in the parental H14 cell line, reported to be normal. (E) Identification of the same trisomies by moving average plot. (F) Tandem multiple copy number gain in proximal 17p in the adapted cell line H14-HSR⁺ relative to the normal control. (G) Single copy number gain in distal 1p in the chHES-3 cell line (Yang et al., 2008). Two samples of different passages have been shown to carry the same aberration 46XX,dup(1)(p32p36) relative to a cytogenetically verified normal cell line from the same study.

Asterisks indicate p-value < 1×10^{-4} , judged by Expander and EASE location analyses. All significance tests are presented following Bonferroni multiple test correction. Vertical dashed lines represent cytogenetic boundaries of chromosomal aberrations as previously published in the respective studies. Horizontal colored bars represent piecewise constant fit (PCF) abnormality detection as described in the methods section.

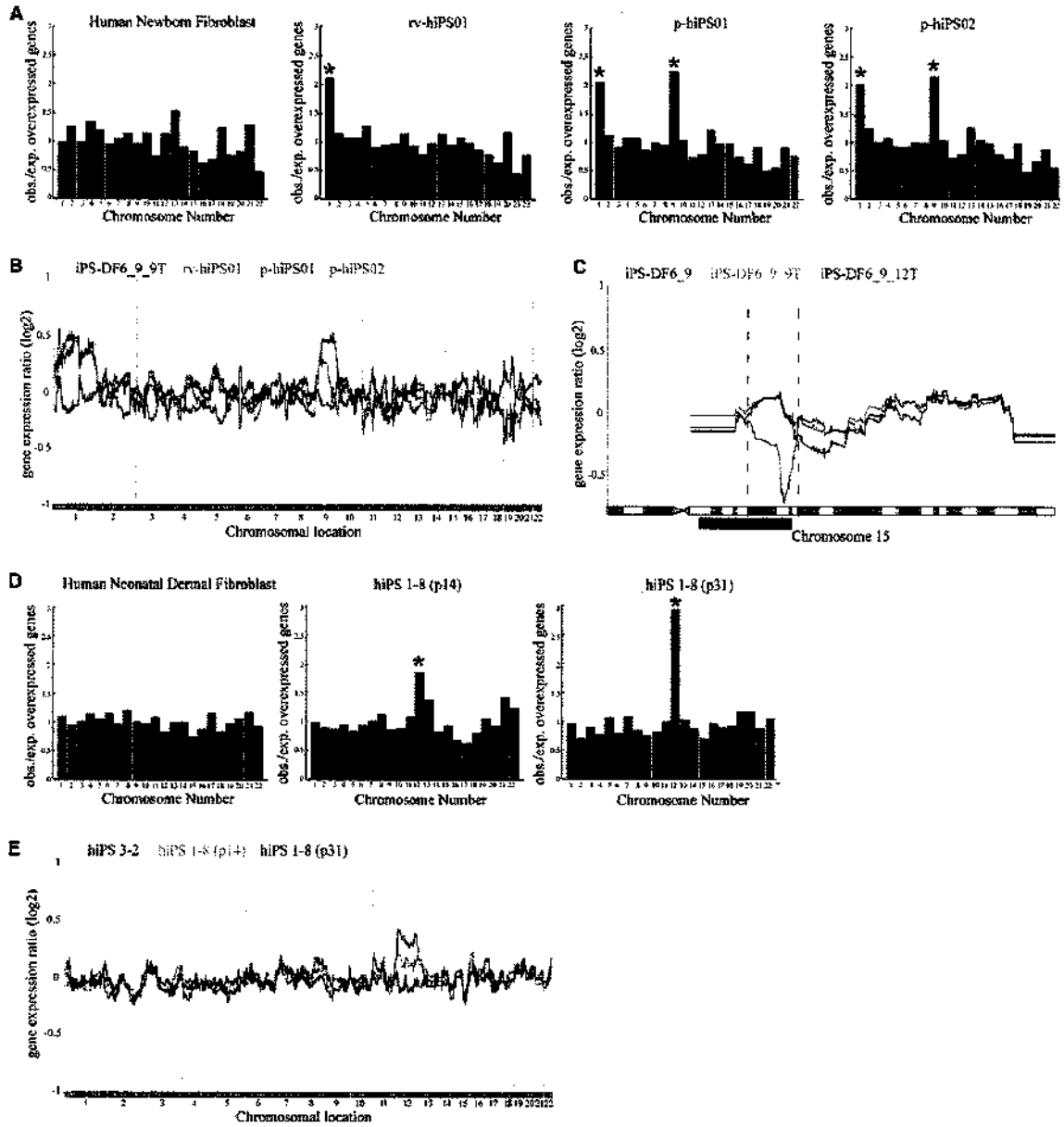


Figure 2: Identification of chromosomal aberrations in HiPSCs.

(A) Analysis of whole chromosome gains in HiPSCs derived by retroviral induction (rv-hiPS01) or direct delivery of reprogramming proteins (p-hiPS01 and p-hiPS02) (Kim et al., 2009). Chromosome 1 and chromosome 9 trisomies were detected in p-hiPS01 (Bonferroni corrected p-values = 5.0×10^{-32} by Expander and 1.1×10^{-30} by EASE, for chromosome 1; 6.0×10^{-13} by Expander and 3.2×10^{-18} by EASE, for chromosome 9) and p-hiPS02 (Bonferroni corrected p-values = 4.0×10^{-33} by Expander and 4.5×10^{-44} by EASE, for chromosome 1; 2.7×10^{-12} by Expander and 2.2×10^{-16} by EASE, for chromosome 9), while chromosome 1 and chromosome 9p trisomies were detected for rv-hiPS01 (Bonferroni corrected p-values = 1.3×10^{-30} by Expander and 5.0×10^{-46} by EASE, for chromosome 1; 1.4×10^{-7} by Expander and 8.7×10^{-13} by EASE, for chromosomal arm 9p). The original study described trisomy in chromosomes 1 and 9 for cell lines p-hiPS01 and p-hiPS02 as well as for their parental somatic cell line (Kim et al., 2009). (B) Moving average plot of the gene expression profile of these cell lines. The normal HiPSC line iPS-DF6_9 is displayed as reference.

Identification of aneuploidy acquired in culture: (C) Identification of a small deletion described by Yu et al. (2009) in the subclone iPS-DF6_9_12T. All the other clones described in this study were found in our analysis to be normal, congruent with the original report (Yu et al 2009). The parental HiPSC line iPS-DF6_9 and another subclone isolated from this line (iPS-DF6_9_9T) are presented as normal reference. (D, E), Acquired chromosome 12 trisomy in two separately grown samples of HiPSC line 1-8 from Masaki et al. (2008): (D) Parental somatic cell and HiPSC 1-8 from different passages (14 and 31), showing trisomy in chromosome 12 (Bonferroni corrected p-values = 6.0×10^{-10} by Expander and 1.0×10^{-18} by EASE, for passage 14; 5.0×10^{-36} and 1.3×10^{-26} , for passage 31). Two other clones, HiPSC 2-4 and HiPSC 3-2, from the same somatic origin as clone HiPSC 1-8 did not share this trisomy (Table S2). (E) Moving average plot of the gene expression profile of clone 1-8 at passages 14 and 31, clone 3-2 is presented as normal reference. These results suggest a trisomy of chromosome 12, which was acquired very early in culture and gradually took over.

Asterisks indicate $p\text{-value} < 1 \times 10^{-4}$, judged by Expander and EASE location analyses. All significance tests are presented following Bonferroni multiple test correction. Vertical dashed lines represent cytogenetic boundaries of chromosomal aberrations as previously published in the respective studies. Horizontal colored bars represent piecewise constant fit (PCF) abnormality detection as described in the methods section.

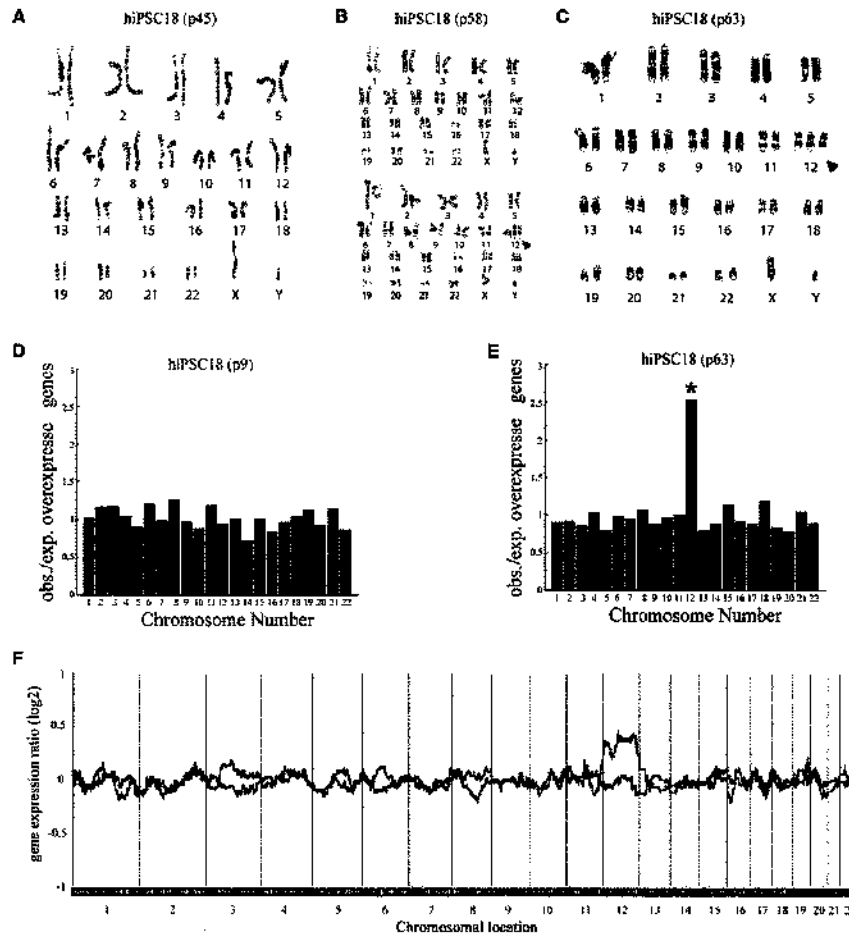


Figure 3: Generation of trisomy 12 in HiPSCs upon their growth in culture

(A) Karyotype analysis of hiPSC18 at passage 45, demonstrating this cell line was normal at that stage. Array CGH-analysis conducted at passage 48 showed this cell line still had a normal karyotype (Chin et al., 2009). (B) HiPSC18 cells were grown to passage 58, and another karyotype analysis was performed, demonstrating that about half of the cells have acquired a full trisomy of chromosome 12 (9/20 metaphases). Normal and trisomic karyotypes from the same analysis are presented. (C) The cells were further grown to passage 63, and a third karyotype analysis was performed, demonstrating that by this passage the trisomy had taken over the culture. (D) Whole chromosome gain analysis of hiPSC18 from passage 9, showing this cell line was indeed normal. (E) Whole chromosome gain analysis of hiPSC18 from passage 63 correctly detected the full trisomy of chromosome 12 (Bonferroni corrected p-value = 4.3×10^{-47} by Expander and 1.6×10^{-38} by EASE). (F) Moving average plot of the gene expression profile of hiPSC18 at passage 9 and 63, demonstrating that the cells acquired trisomy 12.

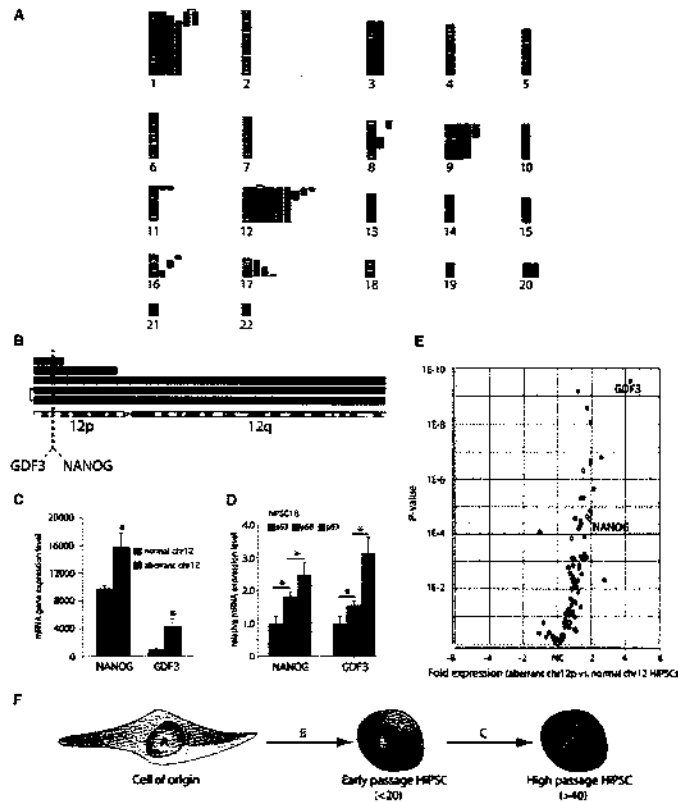


Figure 4: Functional analysis of recurring chromosomal aberrations in HiPSCs

(A) Ideogram representing gained chromosomal regions identified by PCF analysis. Each autosome is represented by its typical cytogenetic structure, and each red bar represents a gain of the respective chromosomal region in one line. Dark and light bars represent gains in HiPSCs and HESCs, respectively. (B) Representation of chromosome 12 gains identified by PCF analysis. Each bar represents a gain of the respective chromosomal region. The chromosomal location of NANOG and GDF3 is presented. (C) Comparison of the levels of expression of NANOG and GDF3 between the 5 aberrant HiPSCs and HiPSCs lines with normal chromosome 12 copy number, showing that both genes are significantly over-expressed in the lines with gains in chromosome 12 (p -values= 4.3×10^{-5} and 2.9×10^{-10} respectively; average fold changes $\times 1.6$ and $\times 4.4$, respectively; error bars represent the s.e.m.). (D) RT-PCR expression of NANOG and GDF3 at three different passages of hiPSC18, demonstrating their expression increase upon selection for trisomy 12 in the culture. Expression values were normalized according to the expression of beta-Actin. Asterisks indicate p -value <0.05 . (E) Volcano plot showing over-expression and under-expression of genes residing in the minimal gained region common to all 5 aberrant HiPSCs, in these aberrant lines relative to lines with normal chromosome 12 copy number. NANOG and GDF3 are highlighted in red. (F) Schematic model of the different types of chromosomal aberrations found in HiPSCs. The genomic aberrations identified in HiPSCs are divided to three general categories: aberrations with somatic cell origin (A); aberrations present in early passage but without apparent somatic cell origin (B); aberrations acquired during passaging of the cells (C).

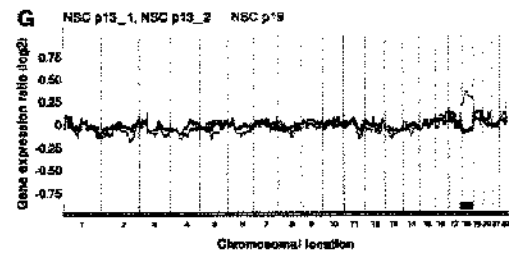
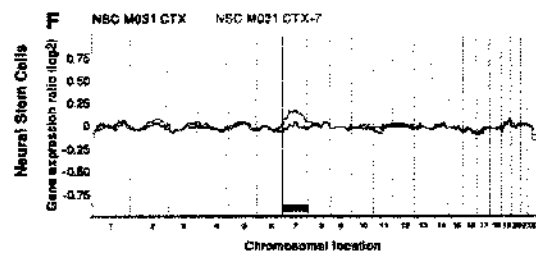
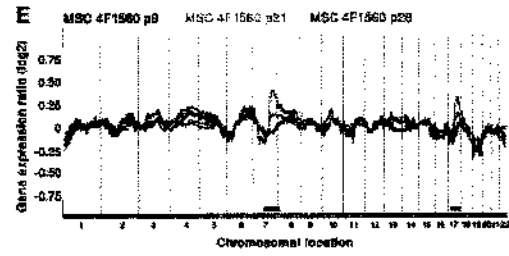
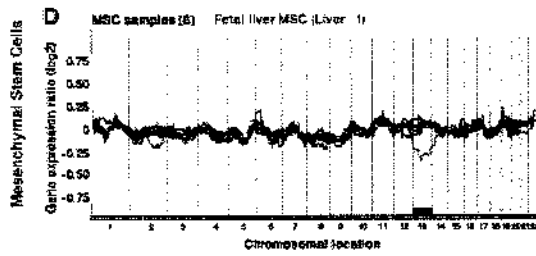
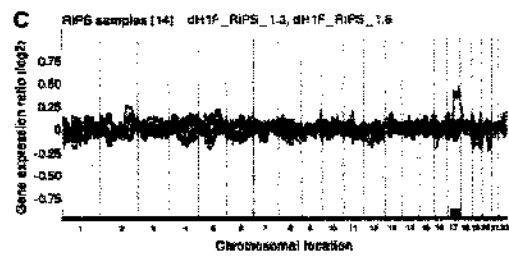
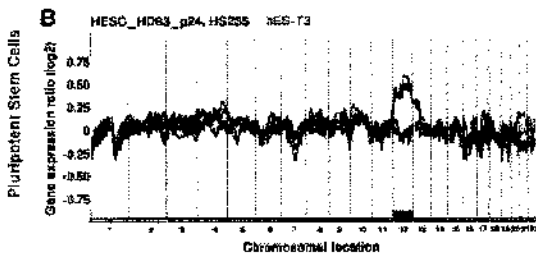
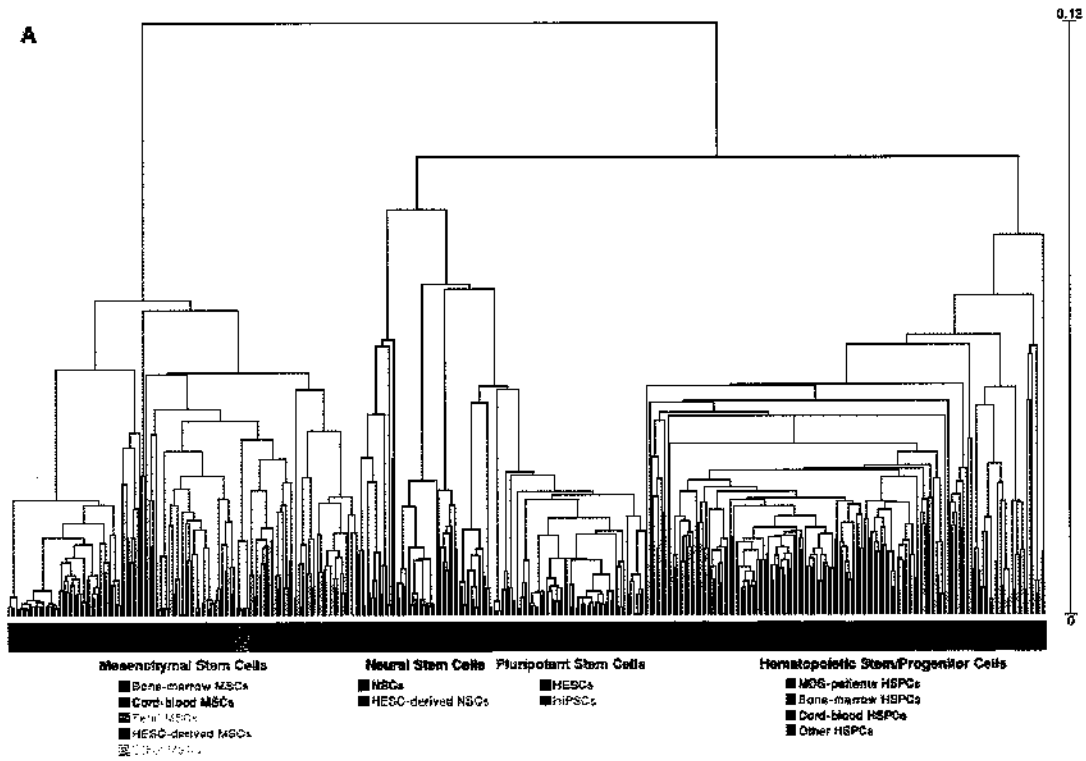


Figure 5: Gene expression patterns reveal chromosomal aberrations in human adult stem cells.

(A) Unsupervised hierarchical clustering of the MSCs (green branches), NSCs (blue branches), HSPCs (purple branches) and PSCs (red branches) analyzed in the current study (in Affymetrix HG-U133plus2 platform). The distinct groups of stem cells cluster apart from each other, and were thus analyzed separately. The various origins of the samples inside each stem cell group are color-coded. (B-G) Moving average plots of gene expression levels along the whole genome of: human pluripotent stem cells (B-C), mesenchymal stem cells (D-E) and neural stem cells (F-G). (B) Six samples of the HESC line, hES-T3, cultured in various conditions, demonstrate trisomy of chromosome 12 in this cell line (red lines, $p=5 \times 10^{-39}$). Two normal HESC lines from different studies (HESC_HD83_p24 and HESC HS235) are presented as controls (blue lines). (C) Synthetic mRNA-induced HiPSC lines, dH1F_RiPS_1.3 and dH1F_RiPS_1.6, demonstrate trisomy of 17q (red lines, $p=6 \times 10^{-26}$). 14 other samples from the same study are presented as controls (blue lines). (D) Fetal liver-derived MSC line, Liver_1, demonstrates monosomy 13 (red line, $p=3 \times 10^{-10}$). Six other MSC samples from the same study are presented as controls (blue lines). (E) Bone marrow-derived MSC line, #4F1560, acquired trisomies of 7q and 17q during its passaging in culture. The cells were normal at passage 9 (three replicates, blue lines), the trisomy is already evident at passage 21 (two replicates, orange lines), and it took over the culture by passage 28 (three replicates, red lines, $p=2 \times 10^{-7}$ and 1×10^{-7} for trisomies 7q and 17q, respectively). (F) Fetal cortex-derived NSC line, M031 CTX +7, demonstrates trisomy 7 (red line, $p=2 \times 10^{-8}$). This trisomy was previously identified in this cell line. The diploid NSC line from the same study, M031 CTX, is presented as control (blue line). (G) Fetal-derived NSC line acquired trisomy 18 during its passaging in culture. The cells were normal at passage 9 (two replicates, blue lines), but acquired the trisomy by passage 19 (red line, $p=1 \times 10^{-9}$).

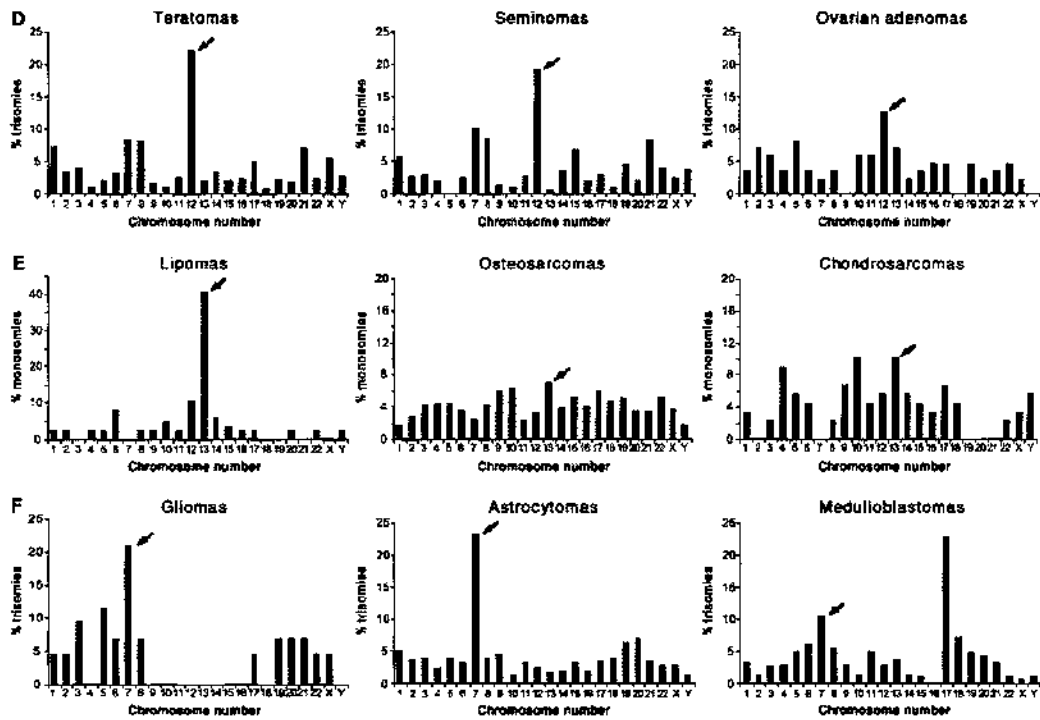
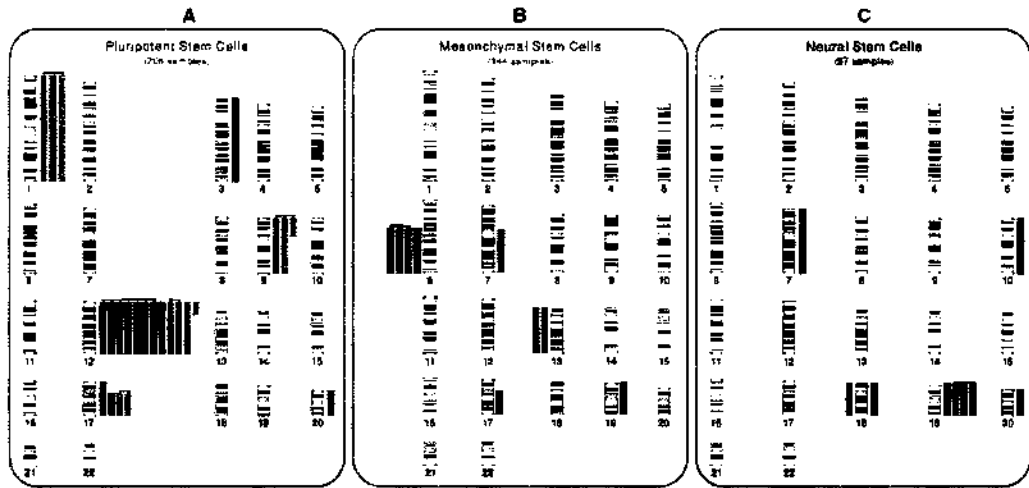


Figure 6: Different types of stem cells acquire distinct chromosomal aberrations.

(A-C) Ideograms representing the chromosomal aberrations identified in: (A) pluripotent, (B) mesenchymal, and (C) neural stem cells. Bars to the right of the chromosome represent gains, and bars to the left of the chromosome represent deletions. In the ideogram of PSCs (A), red and orange represent HESCs and HiPSCs, respectively. The PSC aberrations identified in the current study are shown together with the aberrations identified already in (Mayshar et al., 2010). Chromosomal aberrations in samples from similar cells from the same study are interconnected by a line, and were considered as a single aberration for the purpose of statistical analysis. (D-E) Some of the recurrent aberrations detected in stem cells, are the most common aberrations in tumors of the same tissue origin. The frequency of chromosomal aberrations in various types of tumors was calculated using the National Cancer Institute “Recurrent Chromosomal Aberrations in Cancer Database Searcher”. (D) The relative frequency of trisomies, gains and isochromosomes of each chromosome in three types of tumors of germ cell tissues. Trisomy 12 is the most common aberration in mature and immature teratomas (found in 182/827 cases), seminomas of the testis and the ovary (91/479) and adenomas of the ovary (11/87). (E) The relative frequency of monosomies of each chromosome in three types of tumors of mesenchymal tissues. Monosomy 13 is the most frequent aberration in lipomas (36/88), skeletal osteosarcomas NOS (49/694) and dedifferentiated chondrosarcomas (9/89). (F) The relative frequency of trisomies, gains and isochromosomes in three types of neural tumors. Trisomy 7 is the most common aberration in gliomas (9/43) and astrocytomas (256/1107), and is also frequently found in medulloblastomas (29/280). Trisomy 17, which was not detected in the analysis of neural stem cells, is the most common aberration in medulloblastomas. Arrows indicate chromosomes that were identified in stem cell cultures of a specific type and are also frequent in tumors of the same tissue, revealing a possible correlation between the two (the correlation is most evident for PSCs and germ cell tumors, less significant for MSCs and mesenchymal tumors, and marginal for NSCs and neural tumors).

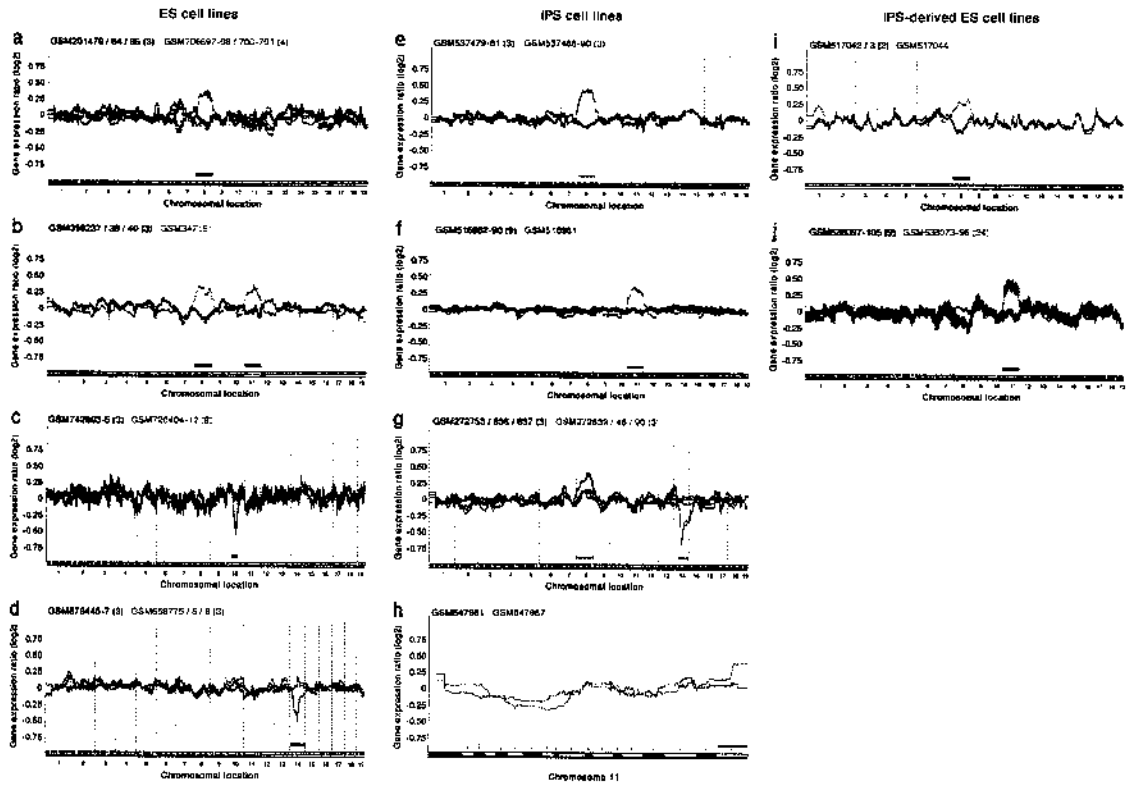


Figure 7: Gene expression patterns reveal hotspots of chromosomal aberrations in mouse pluripotent stem cells.

Moving average plots of gene expression levels along the autosomal genome of ESCs (a-d), iPSCs (e-h) and ESCs that were derived by nuclear transfer of iPSCs (i-j). (a) Four samples of 129sv ESCs demonstrate trisomy of chromosome 8 (red lines). Three normal ESC lines from another study are presented as controls (blue lines). (b) D3 ESC sample demonstrates trisomy of chromosomes 8 and 11 (red line). Three normal ESC lines from another study are presented as controls (blue lines). (c) Nine samples of D3 ESC demonstrate deletion of chromosome 10qB (red lines). Three normal ESC lines from another study are presented as controls (blue lines). (d) Three samples of E14Tg2A ESCs demonstrate monosomy of chromosome 14 (red lines). Three normal ESC lines from another study are presented as controls (blue lines). (e) Three samples of iPSCs demonstrate trisomy of chromosome 8 (red lines). These iPSCs were derived by retroviral transduction of mouse embryonic fibroblasts with Oct4, Sox2, Klf4 and c-Myc. Three samples of normal iPSCs, from another clone reported in the same study, are presented as controls (blue lines). (f) iPSC sample demonstrates trisomy of chromosome 11 (red line). This iPSC line was derived by drug-induction of peritoneal fibroblasts from transgenic mouse. Nine samples of normal iPSCs, from various clones reported in the same study, are presented as controls (blue lines). (g) Three samples of iPSCs demonstrate trisomy of chromosome 8 and deletion of chromosome 14qC-14qE (red lines). These iPSCs were derived by retroviral transduction of mouse neural stem cells with Oct4 and Klf4. Three samples of normal ESC lines from another study are presented as controls (blue lines). (h) iPSC sample demonstrates gain of chromosome 11qE2, acquired in culture within 12 passages. The cells were normal at passage 4 (blue line), and acquired the aberration by passage 16 (red line). This iPSC line was derived by drug-induction of T-cells from transgenic mouse. (i) iPSC-derived Rosa26 ESC sample demonstrates trisomy of chromosome 8 (red line). This ESC line was derived by nuclear transfer of iPSCs into enucleated oocytes. The iPSCs of origin were derived by adenoviral transduction of mouse tail-tip fibroblasts with the reprogramming factors. Two normal subclones of the same original cell line are presented as controls (blue lines), suggesting this aberration occurred at the ESC stage of the cells. (j) Twelve samples of iPSCs and twelve samples of iPSC-derived ESCs demonstrate trisomy of chromosome 11 (red lines). The iPSCs were derived by drug-induction of mouse embryonic fibroblasts from transgenic mouse, and the ESCs were derived by nuclear transfer of these iPSCs into enucleated oocytes. As the iPSCs and their ESC derivatives exhibit the exact same aberration, it has most likely occurred already at the iPSC stage of the cells. Nine normal ESC samples from the same study are presented as controls (blue lines).

Red, orange and purple bars indicate gains in ESCs, iPSCs and iPSC-derived ESCs, respectively; dark and light green bars indicate deletion in ESCs and iPSCs, respectively. Samples appear by their GSM numbers, as deposited in the GEO database. The numbers of normal (blue) and aberrant (red) samples in each panel appear within parentheses.

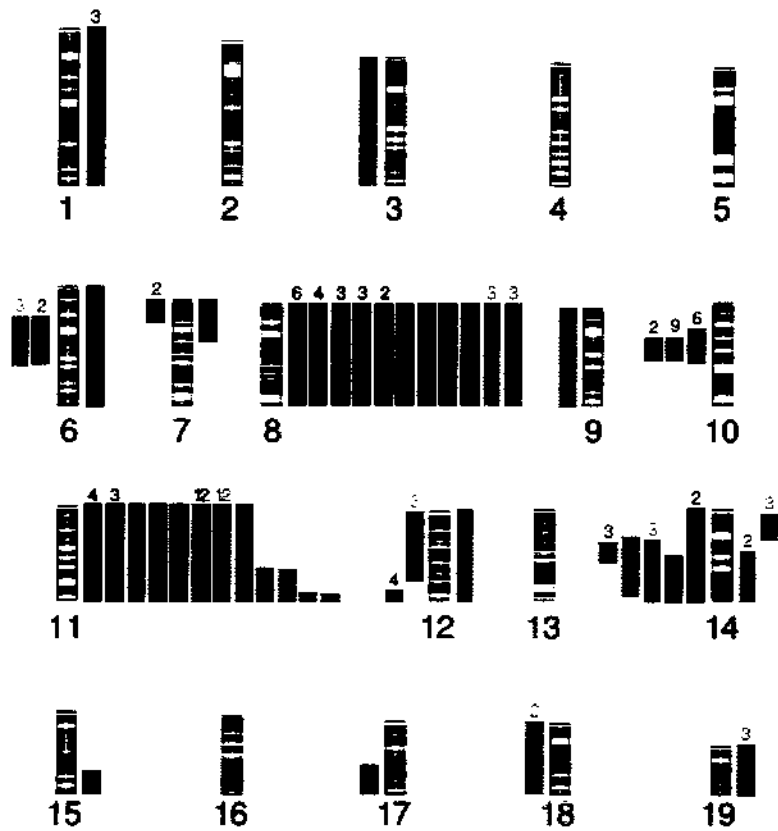


Figure 8: Ideogram of chromosomal aberrations in mouse pluripotent stem cells.

This ideogram represents the chromosomal aberrations identified in the autosomal genomes of mouse ESCs, iPSCs and iPSC-derived ESCs. Bars to the right of the chromosome represent gains, and bars to the left of the chromosome represent deletions. Red, orange and purple bars indicate gains in ESCs, iPSCs and iPSC-derived ESCs, respectively; dark and light green bars indicate deletions in ESCs and iPSCs, respectively. Chromosomal aberrations in samples from similar cells from the same study are shown and considered as a single aberration, for the purpose of all statistical analyses and graphic presentations. Whenever a bar represents more than one sample, the number of represented samples is indicated above it.

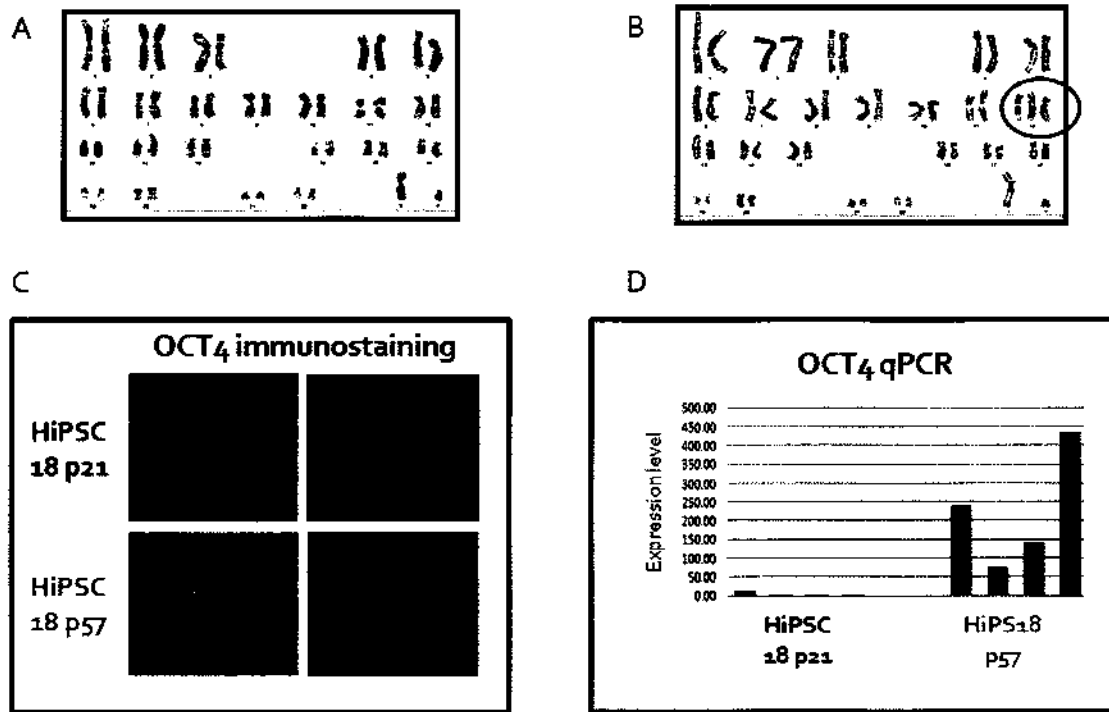


Figure 9: Trisomy 12 makes HiPSCs more tumorigenic.

(A) Karyotype analysis of the HiPSC line hiPS18 at passage 15, demonstrating this cell line was normal at that stage. (B) Karyotype analysis of the same cell line at passage 53, demonstrating the cells have acquired full trisomy of chromosome 12 by that stage. (C) Immunostaining for OCT4 of teratomas generated by diploid (upper panel) and aneuploid (lower panel) hiPS18 cells. Only teratomas generated by the aneuploid cells stained positively, showing these teratomas are less differentiated. (D) qPCR measurement of OCT4 levels in RNA derived from four teratomas generated by injection of diploid hiPS18 cells (blue bars), and from four teratomas generated by injection of aneuploid hiPS18 cells (red bars). OCT4 is highly expressed only in teratomas derived from aneuploid cells.

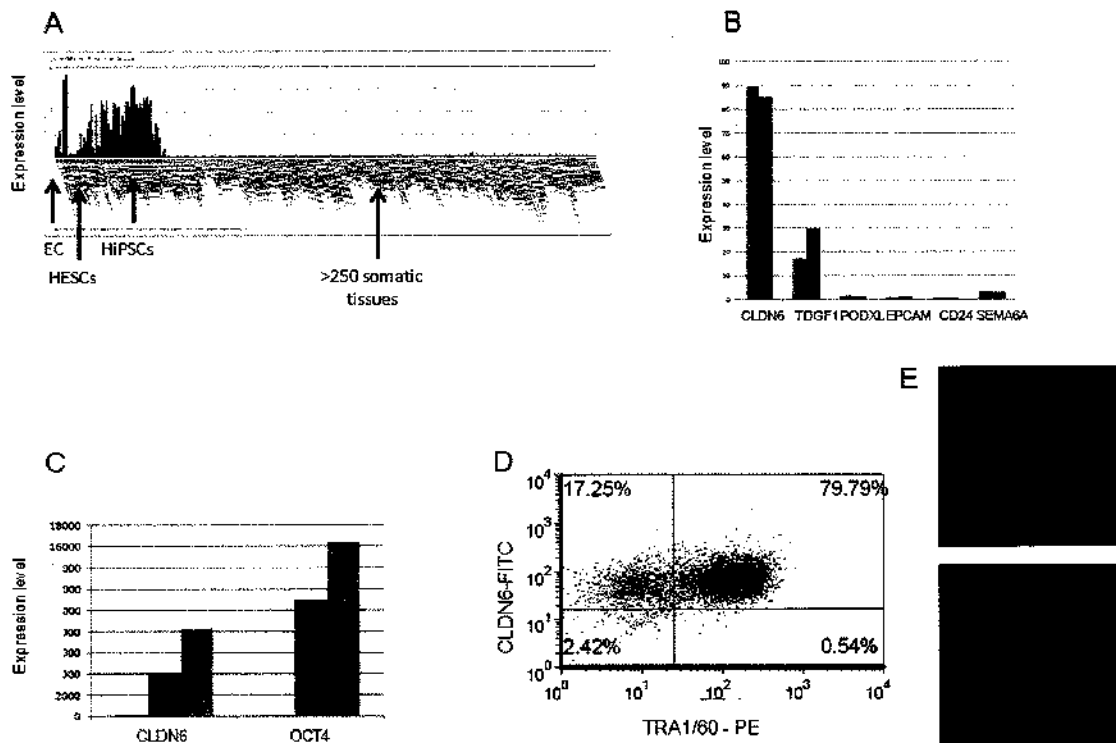


Figure 10: CLDN6 is a tight-junction gene expressed specifically in PSCs

(A) The RNA level of CLDN6, as measured by gene expression microarrays, in embryonic carcinoma (EC) cells, HESCs, HiPSCs and over 250 adult tissues (data was gathered from the Amazonia database). These results show that CLDN6 is highly expressed only in pluripotent stem cells, and is not expressed at all in other adult tissues (including adult stem cells). (B) A comparison of the expression level of CLDN6 and TDGF1 vs. known pluripotent-specific surface markers. Green: average expression in HESCs vs. the maximum expression among all differentiated tissues; Purple: Average expression in HiPSCs vs. the maximum expression among all differentiated tissues. (C) qPCR results of the relative expression level of CLDN6 and Oct4 in fibroblasts (blue), H9 HESCs (red) and CSES2-SO2 HESCs (green). Expression levels were normalized to GAPDH. (D) FACS analysis with antibodies against CLDN6 (Y axis) and TRA-1-60 (X axis). Practically all the cells that express TRA-1-60 also express CLDN6. (E) Confocal microscopy pictures after immunostaining of BJ28 HiPSCs with an antibody against CLDN6 (red). Nuclei were stained with Hoechst (blue). The results show that the pluripotent cells indeed express CLDN6 on their membranes.

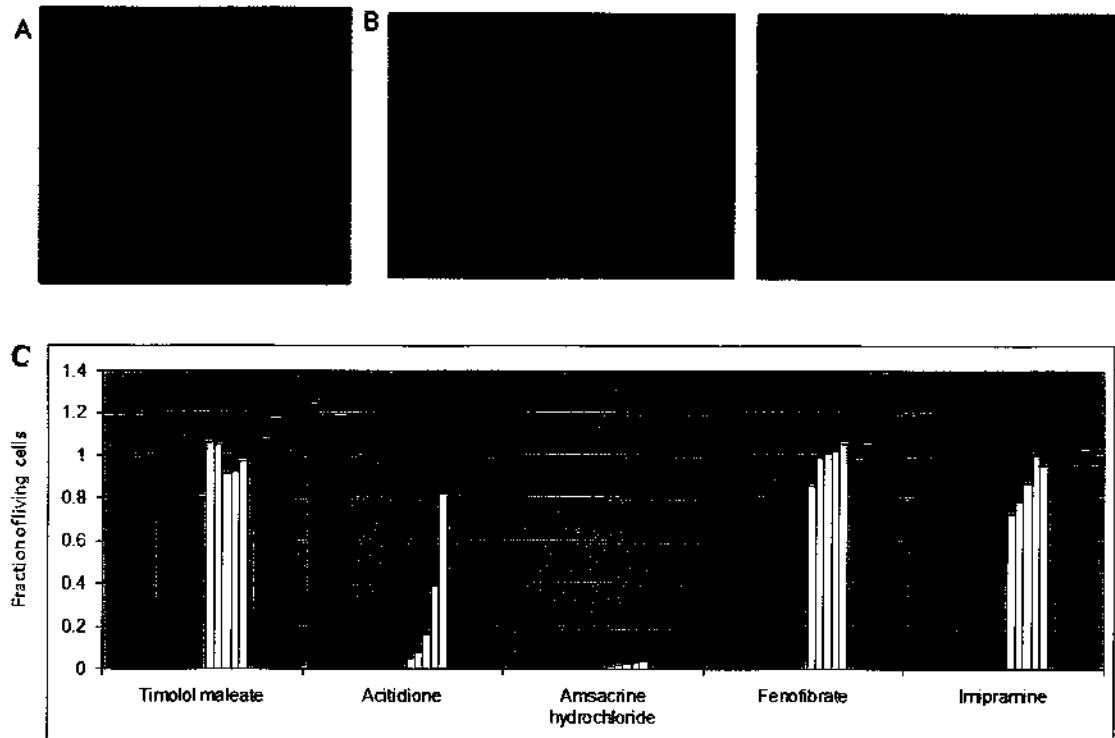


Figure 11: Assay development for screening HESCs and HiPSCs in a 384-well format

(A) CSES2 HESCs grown on matrigel with mTeSR defined medium (“feeder-free”) remain undifferentiated, as is evident by their morphology and by their positive staining to alkaline phosphatase. (B) Cells were then grown at these conditions in 384-well plates. The cells remained undifferentiated, as was validated by immunostaining to OCT4 (green, left), and by measurement of the red fluorescence of CSES2-SO2 cells (red, right). (C) A pilot screen of 50 known cytotoxic compounds was performed in order to select compounds that can use as positive controls (i.e. compounds that kill all cells in culture) and in order to determine the best time course for the primary screen. Five examples are shown in the graph, each compound was added to the cells for 6 (red), 24 (blue), 48 (yellow), and 72 (purple) hours. Compounds were added at 5 concentrations (from 30uM to 0.3uM, from left to right inside each group of bars). Amsacrine hydrochloride, a topoisomerase inhibitor that had a strong cytotoxic effect on the cells, was selected as positive control. The best exposure time was determined to be 24h.

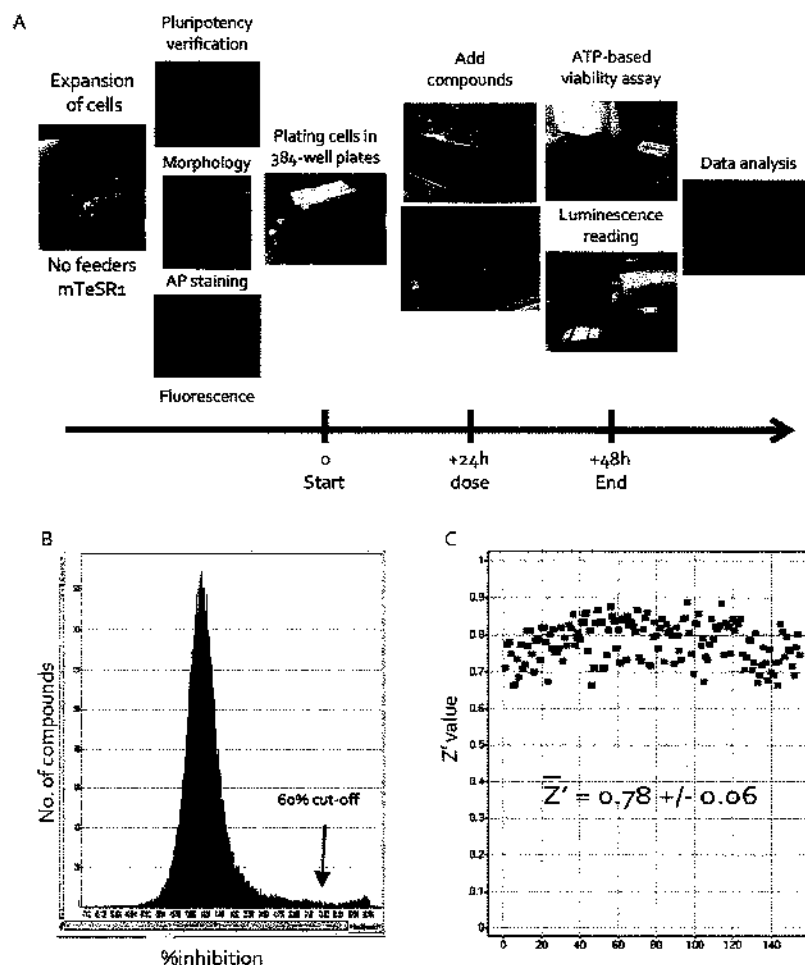


Figure 12: Primary screen

(A) A flow chart of the primary screen: CSES2 cells were grown on matrigel-coated plates with mTeSR defined medium. Before their transfer to 384-well plates, pluripotency was verified by morphology and alkaline phosphatase staining. Cells were harvested, counted, and dispensed in 384-well plates, using the Wellmate machine, so that 5,000 cells were plated in each well. Plates were incubated overnight (~24h) to allow the cells to settle down. The next day, compounds were added to the assay plates, together with positive and negative controls, using the Biomek FX liquid handling machine. 24 hours later, the number of living cells was quantified using the CellTiter-Glo viability assay. The raw data was normalized to the internal controls using the Genedata software. (B) Histogram of the response of CSES2 cells to 52,484 compounds. “Hits” were determined as compounds that caused over 60% inhibition (that is, above 60% reduction in the signal). Using this cut-off, 2,031 compounds were selected to be re-tested in confirmation screens. (C) Z' factor is an accepted index for screen quality, and it is calculated separately for each plate. The primary screen assay had very high robustness, with mean $Z' = 0.78 \pm 0.06$ (149 plates).

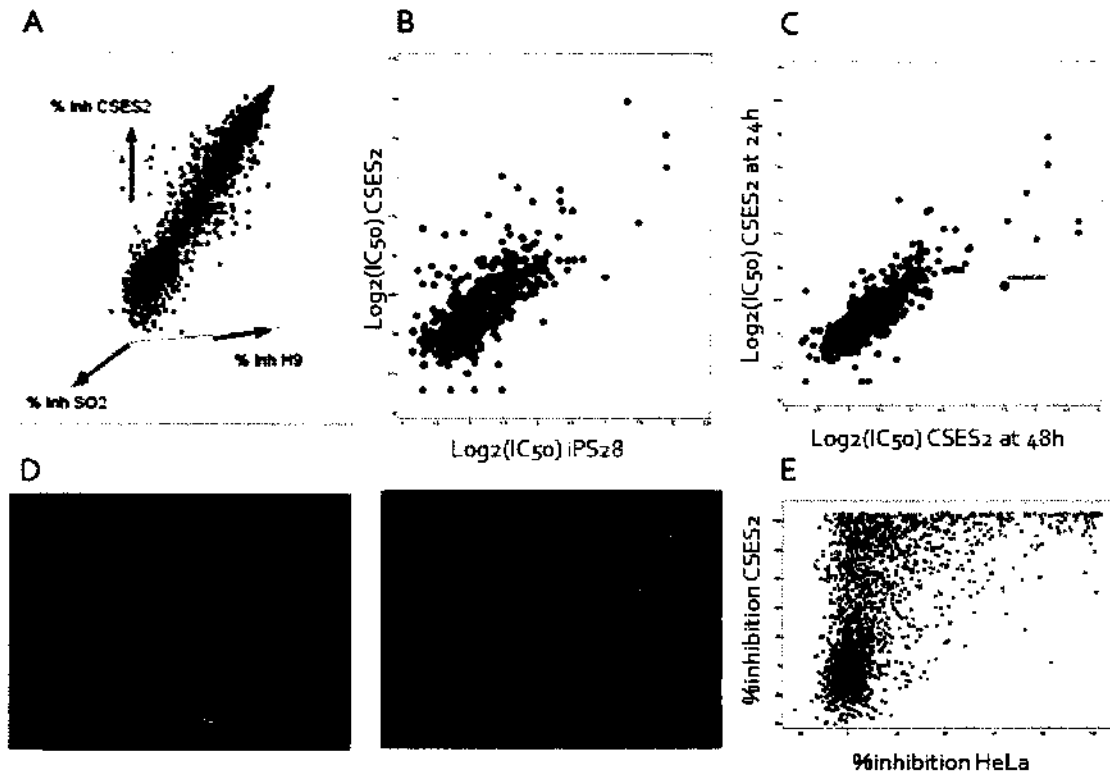


Figure 13: Confirmation and validation screens

(A) 3D dot-plot of the response of 3 HESC lines to the compounds. The confirmation screen was first conducted in triplicates at a single concentration. Data comparison between CSES2 and the other HESC lines revealed very high correlation, and ruled out line-specific effect. 696 compounds were confirmed to be genuine inhibitors of pluripotent cells. (B) These confirmed “hits” were screened again at multiple (6-8) concentrations in order to obtain their IC₅₀ values (the concentration in which each compound causes a 50% inhibition). Very high correlation values were observed within the HESC lines, as well as between the HESCs and HiPSCs. The dot-plot shows the correlation in the IC₅₀ values between CSES2 and HiPS18 cell lines. (C) Very high correlation was also observed between 24h and 48h exposure of CSES2 cells to the compounds. (D) The CSES2-SO2 cells were also subjected to a high-content screen (HCS), in which the level of red fluorescence was quantified in each well by microscopy imaging. Intensity of red fluorescence was automatically measured in each of the wells and compared to the negative control (left) and positive control (right). (E) HeLa cells were screened with the 696 “hits”, and the correlation of the inhibition values between HESCs and HeLa cells was low ($r^2 = 0.31$). ~200 compounds were found to have no significant effect on HeLa cells, although they dramatically reduce the viability of PSCs.

Materials and methods

Cell culture

HESCs and HiPSCs were cultured in standard ES cell culture media containing KnockOut DMEM (Gibco-Invitrogen) supplemented with 15% Knockout serum replacement (Gibco-Invitrogen), 2mM L-glutamine (Sigma-Aldrich), 1:100 dilution of non-essential amino acids (Gibco-Invitrogen), 0.1 mM β -mercaptoethanol (Sigma-Aldrich), 8 ng/ml basic fibroblast growth factor (bFGF) (PeproTech), penicillin (50 units/ml) and streptomycin (50 ug/ml) (Gibco-Invitrogen).

Alkaline phosphatase staining and Immunocytochemistry

Alkaline phosphatase staining was performed using the Leukocyte Alkaline Phosphatase Kit (Sigma-Aldrich) according to the manufacturer's instructions. For immunocytochemistry, cells were washed twice with PBS, crosslinked with 4% formalin solution for 10 min, washed twice with PBS and blocked for one hour at room temperature with PBS containing 2% bovine serum albumin (BSA, Sigma-Aldrich) and 0.1% Triton-X-100. Primary antibody staining was performed for 1 hour in room temperature with antibodies diluted in blocking solution. Antibodies were used diluted according to the recommendations of the manufacturer.

DNA microarray analysis

Total RNA was extracted according to the manufacturer's protocol (Affymetrix, Santa Clara, CA) from populations of BJ fibroblasts, BJ-HiPSCs and normal and HESCs. Hybridization to the GeneChip Human Gene 1.0 ST Arrays, washing, and scanning were performed according to the manufacturer's protocol (Affymetrix), and expression patterns were compared between samples.

Gene expression profiles database

Gene expression profiles from studies were obtained from the GEO database (<http://www.ncbi.nlm.nih.gov/geo>). Raw .CEL files for all samples were analyzed using MAS5 probeset condensation algorithm using Expression Console (Affymetrix, CA). Arrays were analyzed for quality control and outliers removed. Further outliers were removed following hierarchical clustering analysis. Probes absent in over 20% of the samples were discarded. In the case of multiple probesets for any given gene, multiple instances were discarded, so that each gene would be represented by one probeset only. Probesets without documented chromosomal location were also removed. Thus, separate datasets containing a single probeset for each expressed gene were generated. In order to reduce bias due to low expression levels, values under 50 were collectively raised to this level.

CGH-PCF over-expression analysis

For each sample, the expression value of each autosomal gene was divided by the median of the same gene across the entire dataset, in order to obtain a comparative value. In order to

reduce possible bias from any given experiment, groups of similar samples with highly similar gene expression profiles (as judged by hierarchical clustering) were averaged for the sake of calculating the grand population median. This median then served as the baseline for examining expression bias. The data was then processed using a freely available comparative genomic hybridization (CGH) analysis software program, CGH-Explorer (<http://www.ifi.uio.no/forskning/grupper/bioinf/Papers/CGH>). Gene expression regional bias was detected using the program's piecewise constant fit (PCF) algorithm, using a set of parameters as follows: Least allowed deviation = 0.25-0.30; Least allowed aberration size = 30-80; Winsorize at quantile = 0.001; Penalty = 12; Threshold = 0.01. Moving-average plots of lines and regions of interest were drawn using the moving-average fit tool.

Location enrichment analysis

For each sample in which an aberration was detected, a list of the autosomal genes that are over-expressed (>1.5 fold, for trisomies and gains) or under-expressed (<1.5 fold, for monosomies and deletions) relative to the median expression of that gene, was comprised. This list was then subjected to location enrichment analysis, using the Expander software (<http://acgt.cs.tau.ac.il/expander>). Significance was determined as Bonferroni corrected p-values lower than 1.0E-4, which is the default value of the Expander program.

RNA isolation and reverse transcription

Total RNA (DNase treated) was extracted using RNAeasy Mini Kit (Qiagen) according to the manufacturer's instructions. One microgram of total RNA was reverse transcribed with random hexamer primers using ImProm-II reverse transcriptase (Promega).

Real-time PCR

Real-time PCR was carried out in triplicates using SYBR Green ROX mix (Applied Biosystems) and 7300 Real-time PCR System (Applied Biosystems).

Teratoma formation assay

For teratoma formation, a confluent 10 cm plate was harvested using trypsin. Cells were centrifuged and cell pellets were resuspended in 50uL of ES cell medium. The cells were into the kidney capsule of 4-8 weeks old SCID-Beige male mice (The Jackson Laboratory). Tumors were dissected after 4-6 weeks and cryopreserved in O.C.T. (Sakura) Cryosections were stained with haematoxylin and eosin for histological analysis. All animal experiments were conducted under the supervision of the Hebrew University Faculty of Sciences and Animal Care and Use Committee.

FACS analysis

Cells were incubated with cell dissociating buffer (Gibco, 13151-014) for 20 min at 37°C and washed with 10% fetal calf serum in PBS with 0.05% sodium azide. Dissociated cells were suspended to a final concentration of 5×10^6 cells/ml. Incubation with primary antibody (45 minutes on ice) was followed with secondary antibody (45 minutes on ice). Cells were

washed and then analyzed with the FACSCalibur system (Becton Dickinson, <http://www.bd.com>). Analysis was performed on CELLQUEST software (Becton Dickinson, <http://www.bd.com>). Forward and side-scatter plots were used to exclude dead cells and debris.

High-throughput screen

HESCs and HiPSCs were grown on matrigel-coated plates with mTeSR defined medium (Stemcell Technologies Inc., <http://www.stemcell.com>). Cells were harvested using Accutase (Millipore, <http://www.millipore.com>), counted with Cedex Automated Cell Counter (Roche Innovatis AG, IN, <https://www.roche-applied-science.com>), and dispensed in 384-well plates using Matrix Wellmate cell dispenser (Thermo Scientific, <http://www.matrixtechcorp.com>), so that 5,000 cells were plated in each well. Plates were incubated overnight (~24h) to allow the cells to settle down. The next day, compounds were added to the assay plates, together with positive and negative controls, using the Biomek FX liquid handling machine (Beckman Coulter, <https://www.beckmancoulter.com>). 24 hours later, the number of living cells was quantified using the CellTiter-Glo viability assay (Promega, <http://www.promega.com>). Luminescence was quantified using the EnVision reader (Perkinelmer, <http://www.perkinelmer.com>). Data was analyzed using Genedata Screener software (Genedata, <http://www.genedata.com>).

List of Publications

Mayshar Y.*, **Ben-David U.***, Lavon N., Biancotti J.C., Yakir B., Clark A.T., Plath K., Lowry W.E. and Benvenisty N. Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. *Cell Stem Cell*, 2010 Oct, 7(4):521-31.

(* Equally contributing authors)

Ben-David U., Benvenisty N. and Mayshar Y. Genetic instability in human induced pluripotent stem cells: Classification of causes and possible safeguards. *Cell Cycle*, 2010 Dec, 9(23): 4603-4.

Ben-David U. and Benvenisty N. The tumorigenicity of human embryonic and induced pluripotent stem cells. *Nature Reviews Cancer*, 2011 Apr 11(4): 268-77.

Goldring C.E.P., Duffy P.A., Benvenisty N., Andrews P.W., **Ben-David U.**, Eakins R., French N., Hanley N.A., Kelly L., Kitteringham N.R., Kurth J., Ladenheim D., Lavery H., McBlane J., Narayanan G., Patel S., Reinhardt J, Rossi A., Sharpe M. and Park K. Assessing the safety of stem cell therapeutics. *Cell Stem Cell*, 2011 June, 8(6):618-28.

Ben-David U., Mayshar Y. and Benvenisty N. Large-scale analysis reveals acquisition of lineage-specific chromosomal aberrations in human adult stem cells. *Cell Stem Cell*, 2011 Aug, 9(2):97-102.