Stem Cell Reports

Report

TeratoScore: Assessing the Differentiation Potential of Human Pluripotent Stem Cells by Quantitative Expression Analysis of Teratomas

Yishai Avior,[1] Juan Carlos Biancotti,[2,3] and Nissim Benvenisty[1,*]

[1]The Azrieli Center for Stem Cells and Genetic Research, Department of Genetics, Silberman Institute of Life Sciences, The Hebrew University, Jerusalem 91904, Israel
[2]Zilkha Neurogenetic Institute, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA
[3]Department of Biomedical Sciences and Regenerative Medicine Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA
*Correspondence: nissimb@cc.huji.ac.il
http://dx.doi.org/10.1016/j.stemcr.2015.05.006

SUMMARY

Teratoma formation is the gold standard assay for testing the capacity of human pluripotent stem cells to differentiate into all embryonic germ layers. Although widely used, little effort has been made to transform this qualitative assay into a quantitative one. Using gene expression data from a wide variety of cells, we created a scorecard representing tissues from all germ layers and extraembryonic tissues. TeratoScore, an online, open-source platform based on this scorecard, distinguishes pluripotent stem cell-derived teratomas from malignant tumors, translating cell potency into a quantitative measure (http://benvenisty.huji.ac.il/teratoscore.php). The teratomas used for the algorithm also allowed us to examine gene expression differences between tumors with a diploid karyotype and those initiated by aneuploid cells. Chromosomally aberrant teratomas show a significantly different gene expression signature from that of teratomas originating from diploid cells, particularly in central nervous system-specific genes, congruent with human chromosomal syndromes.

INTRODUCTION

Pluripotent stem cells (PSCs) are defined by their ability for self-renewal and the capacity to differentiate into cells of the three embryonic germ layers: ectoderm, endoderm, and mesoderm (Thomson et al., 1998). These traits turn human PSCs (hPSCs) into a promising disease-modeling platform and a unique resource in regenerative medicine. PSC differentiation capacity, or cell potency, can be evaluated using several strategies. The gold standard assay to assess the pluripotency of mouse cells is injecting them into blastocysts, generating chimeric mice with germline transmission. As for hPSCs, several alternatives exist. One suggested approach (named PluriTest) is to analyze the expression profile of the undifferentiated cells and deduce their pluripotency based on similarities to known pluripotent cells (Müller et al., 2011). While this approach is promising, it does not validate a cell capacity to differentiate toward all lineages. Furthermore, this approach ignores the possibility that mutations in tissue master regulators (not analyzed in the assay) will affect the potency of the cells to differentiate. Numerous in vitro differentiation protocols have been published, enabling direct differentiation toward cells from all three germ layers. However, this strategy is seldom used, as it is highly expensive, time consuming, and complicated. Furthermore, there is no standard set of direct differentiation protocols sufficient to determine cell potency (Müller et al., 2012). In contrast, in vitro spontaneous differentiation into embryoid bodies (EBs) is a commonly used method to assess hPSC differentiation capability (Itskovitz-Eldor et al., 2000). Although

efforts have been made to standardize EB formation and analyze the differentiation capacity into the different germ layers, a commonly accepted method has yet to emerge (Müller et al., 2012).

Teratoma formation is currently the most abundant technique used to evaluate hPSC potency. In this assay, undifferentiated hPSCs are injected into an immune-deficient mouse, forming a heterogeneous tumor composed of terminally differentiated cells from all germ layers. The tumors are then sectioned and stained using H&E and evaluated by an experienced histologist providing qualitative examples for differentiation toward each lineage. Sections can also be stained with antibodies against tissue-specific markers, providing a more direct qualitative approach.

Almost no effort has been made to quantitatively measure teratoma cell composition (Müller et al., 2010), although it might be highly indicative of the initiating cell potency and variation. Gene expression data from teratomas can assess their composition and enable comparison of tumors originating from different PSC lines. However, this kind of analysis and quantitative evaluation had been so far halted by the lack of publicly published expression data.

Using teratoma gene expression data initiated from various hPSC lines, we have developed a method to assess teratoma tissue and lineage composition, hence evaluating the initiating cell's potency. The TeratoScore algorithm uses in vivo expression profiles to calculate and provide each cell line with a quantitative measure for pluripotency. Further comparison between the expression profiles of teratomas with a normal karyotype and ones with chromosomal

CrossMark

aberrations revealed an abnormal tissue-enriched gene expression, suggesting that they can aid in disease modeling.

RESULTS

In order to evaluate cell composition of hPSC-derived teratomas, we have analyzed teratomas from 12 different cell lines (Table S1). Five of the teratomas were generated from diploid hPSCs. Another five were from hPSCs with trisomy of chromosome 12, and four teratomas were initiated from hPSCs with trisomy of chromosome 21. Teratomas were subjected to serial sectioning and H&E staining in order to validate the representation of all three germ layers as previously shown (Ben-David et al., 2014; Biancotti et al., 2010; Chin et al., 2009; Lavon et al., 2008; Narwani et al., 2010). In many histological analyses, structures from all three embryonic germ layers were found in the same section, supporting the notion that different cell types exist in close proximity in the teratomas (Figure S1). RNA from different areas of the teratomas was extracted and analyzed by Affymetrix Human Genome U133 Plus 2.0.

Our aim was to quantitatively analyze cell and lineage differentiation of the teratomas derived from the hPSCs, as an indicator for the cells' potency. An algorithm assessing cell-specific differentiation of PSCs named CellNet was recently published by Cahan et al. (2014). While CellNet algorithm is extremely useful when dealing with a homogenous population of differentiated cells, when tested by us to examine the differentiation of teratomas, it has proven limited in assessing the complexity of hPSC-derived tumors (Figure S2). Although all three germ layer structures were histologically found in the teratomas (Figure S1), CellNet showed some similarity to undifferentiated ESCs by no clear enrichment of any specific tissues (Figure S2A). This was not the case when measuring homogenous tumors, such as hepatocellular carcinoma, where CellNet correctly identified the tumor composition. A more careful analysis of CellNet's results revealed that while some tissues, such as liver, were slightly enriched in the teratomas, the heterogeneous tumors still provided similar scores to the undifferentiated cells in most measured tissues (Figure S2B). These results encouraged us to develop a new platform to assess teratoma composition, reflecting the pluripotent capacity of human stem cells that initiated them.

Using previously published expression data sets of 26 different tissues and cell types (Table S2), we generated a list of 100 genes reflecting tissues from all three germ layers as well as placental genes, reflecting an extraembryonic tissue (Table S3). Genes were chosen according to tissue specificity, lack of expression in pluripotent cells, and sufficient expression in an initial database of

karyotypically normal teratomas (n = 5). An additional set of ten pluripotent markers was also generated, as a control for undifferentiated PSCs. Tissue expression specificity was validated using the Amazonia! online database, which represents expression profile of over 300 cell and tissue types (Le Carrour et al., 2010). The scorecard includes genes specific for the central and peripheral nervous system as well as skin-specific genes to represent the ectoderm; genes specific to the joint, heart, skeletal muscle, adipose tissue, kidney, and bone marrow to represent the mesoderm; and genes specific to the gut, liver, lung, and pancreas representing the endoderm. We also included a subset of placental genes, representing an extraembryonic tissue. Following the gene list assembly, we looked at their expression in teratomas compared with undifferentiated hPSCs and previously published expression data of tumors from seven different tissues (Figure 1). Teratomas show expression of all tissues from all germ layers as well as placental genes. Importantly, the expression of the different lineages in these tumors was similar. As expected, undifferentiated hPSCs show little to no expression of the tissue-specific genes and a high expression of pluripotency markers. Similarly, tumors originating from different tissues and composed mainly of a single cell type were successfully detected by the scorecard as expressing their specific tissue and lineage. Our scorecard showed specific ectoderm lineage expression for tumors such as medulloblastoma and basal cell carcinoma, specific mesoderm lineage expression for tumors such as liposarcoma and acute lymphoblastic leukemia, specific endoderm lineage expression for tumors such as hepatocellular carcinoma and colon cancer, and the extraembryonic tissue expression for choriocarcinoma (Figure 1). Congruent results were obtained when the equivalent normal tissues were measured for the same scorecard (Figure S3).

In order to quantitatively estimate the pluripotency of the tumor initiating cells, we calculated the mean gene expression of each lineage and extraembryonic tissue, multiplied them, and divided by 1,000, producing a single score. This analysis, termed TeratoScore, essentially estimates the differentiation potency of the tumor initiating cells—the very goal of teratoma formation. In other words, TeratoScore validates the potential of cells to differentiate toward all three germ layers, hence being qualified as pluripotent cells. Comparing the TeratoScore outcomes of 14 different teratomas with these of tissue-specific tumors reveals that it can successfully detect the hPSC-derived tumors (Figure 2), with average grades of 1,186 ± 334 (n = 14) compared with 7.3 ± 2.6 (n = 34) of the cell-specific tumors and hPSCs. Of 29 tissue-specific tumors, only one had a grade higher than 50 (colon cancer, TeratoScore value of 77). In contrast, only 1 teratoma in the 14 examined had a grade lower than 100, suggesting that an outcome higher
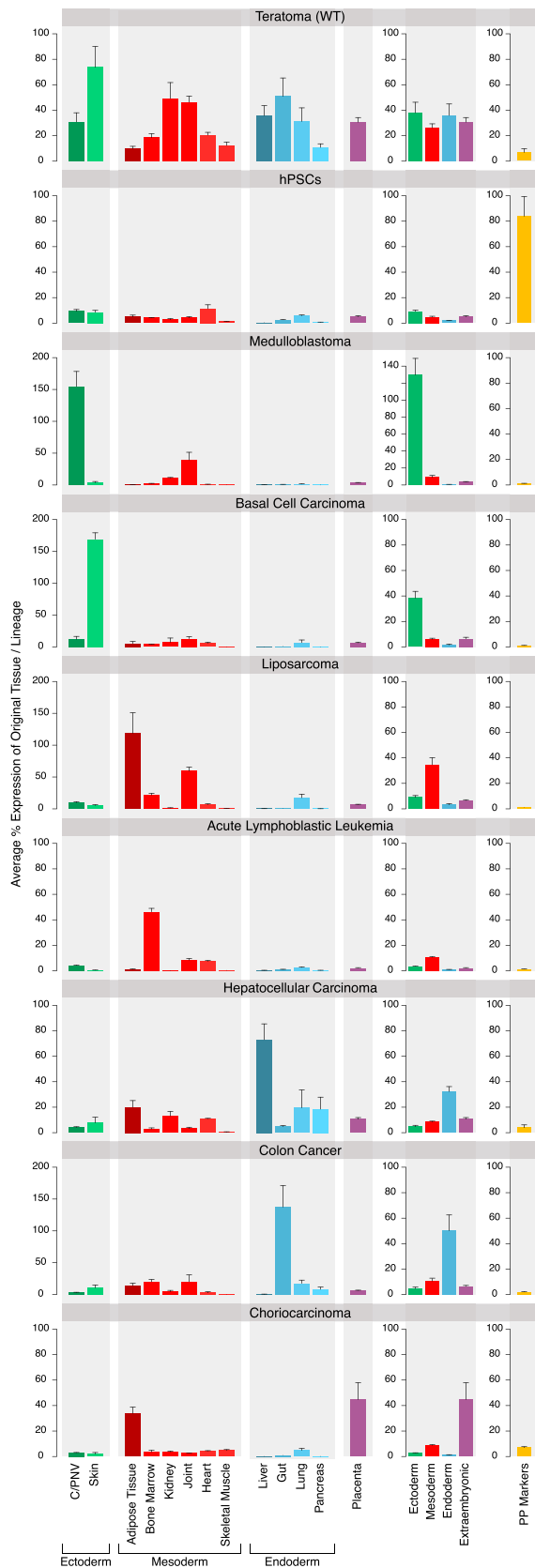
Figure 1. Average Tissue and Lineage Expression in Different Tumors and hPSCs

Average tissue gene expression was calculated in respect of its expression values in the original tissue. Lineage expression was calculated as the mean expression of all genes representing tissues in that lineage (n = 2 for basal cell carcinoma and choriocarcinoma; n = 5 for teratomas, medulloblastoma, liposarcoma, acute lymphoblastic leukemia, hepatocellular carcinoma, and colon cancer; n = 10 for hPSCs). Error bars represent SE. Teratomas show expression of all measured tissues and a similar expression of all lineages, while other tumors show a specific tissue and lineage expression. hPSCs show little to no expression of all the tissues. Pluripotent markers are highly expressed in hPSCs and show low to no expression in all tumors. C/PNV, central or peripheral nervous system.

than 100 sufficiently implies a pluripotent cell line-derived tumor. A TeratoScore value lower than 50 thus indicates a tissue-specific tumor, and a value between 50 and 100 was a borderline indicator of a tumor originated from pluripotent cells (Figure 2). Teratoma composition analysis and the unified outcome using TeratoScore are available using our online resource at http://benvenisty.huji.ac.il/teratoscore.php by uploading an Affymetrix U133 Plus 2.0 microarray expression CEL file.

As this is the first time a substantial amount of expression data regarding teratomas is combined, we were interested in identifying changes between teratomas with a diploid karyotype and teratomas initiated by aneuploid cells. We first examined the differences using the gene scorecard composing the TeratoScore (Figure 3). While some differences were statistically significant, the overall tissue and lineage expression between normal and aberrant cell lines was similar. Teratomas created by hPSCs with a trisomy of chromosome 12 show a significant decrease in adipose tissue and bone marrow-specific gene expression compared with normal cell teratomas (p < 0.02 and p < 0.05, respectively) and lower kidney-specific gene expression compared with normal and trisomy 21 teratomas (p < 0.05 and p < 0.03, respectively). TeratoScore values calculated using these results were 1,490 ± 580 (n = 5) for euploid hPSC-derived teratomas, 1,798 ± 825 (n = 4) for teratomas derived from trisomy 21 hPSC, and 392 ± 117 (n = 5) for trisomy 12-initiated teratomas. Curiously, the mean TeratoScore value of teratomas derived from trisomy 12 hPSC is significantly lower than that of teratomas derived from euploid cells (p = 0.05).

A live newborn with full trisomy 12 was never reported, suggesting that it is embryonic lethal in humans. However, individuals carrying mosaic tetrasomy of 12p were reported showing a variety of abnormalities (Segel et al., 2006). This condition, known as Pallister-Killian mosaic syndrome, is extremely rare and has a range of
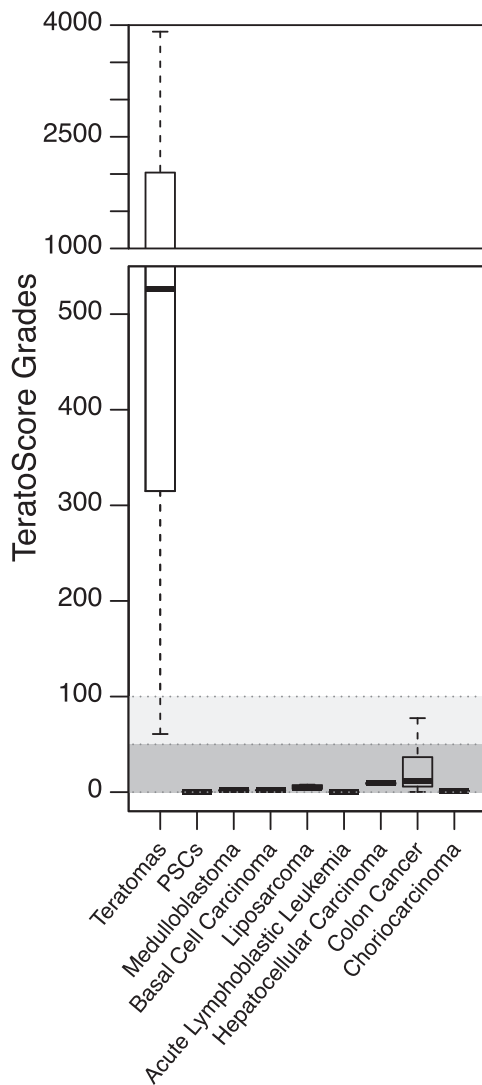
Figure 2. TeratoScore Grades of Different Tumors and PSCs
TeratoScore grades were calculated as the multiplication of the mean lineage gene expression of each tumor divided by 1,000. While teratomas show high and diverse TeratoScore grades (n = 14, mean = 1,186, SEM = 334), all other tumors and PSCs produce significantly lower ones (n = 34, mean = 7.3, SEM = 2.6, p < 10^{-6}).

developmental and behavioral outcomes (Kostanecka et al., 2012). Individuals carrying chromosome 21 trisomy also show a distinct set of abnormalities alongside intellectual disabilities, a condition known as Down syndrome (Patterson, 2009). With such an atypical in vivo development, we hypothesized that teratomas originating from cells carrying these mutations will show a more substantial differential gene expression. We suspected that the lack of significant changes originated from the small number of examined genes. While our 100-gene platform is sufficient for the TeratoScore analysis and defining a cell

as pluripotent, it is not capable of identifying minor expression changes and tissue distribution between euploid and aneuploid cells.

We therefore used another more sensitive bioinformatic strategy, establishing lists of the 200 most expressed genes in each tissue and looking for the ones most altered in chromosomally aberrant teratomas (Figure 4). In this analysis, trisomy 12-derived teratomas show a significant number of upregulated genes enriched for joint, spinal cord, brain, and fetal brain (p < 0.001) and a significant number of downregulated genes enriched for skeletal muscle, lung, small intestine, colon, fetal liver, and liver (p < 0.001) compared with normal karyotype teratomas. This wide set of gene expression alteration might suggest a pleiotropic effect of trisomy 12 on developmental patterns, contributing to its embryonic lethality. In contrast, teratomas initiated from cells with trisomy 21 show a more modest distortion of gene expression with a significant number of upregulated genes enriched for brain, fetal brain, and spinal cord (p < 0.001), as well as for joint, and a significant number of downregulated genes for skeletal muscle only compared with normal karyotype teratomas. The upregulation in brain genes in these teratomas is compatible with the neural phenotypes (such as mental impairment) observed in Down syndrome patients.

DISCUSSION

Teratoma formation is the most abundant method to demonstrate hPSC differentiation capacity and a critical step when defining its potency (Müller et al., 2010). TeratoScore transforms this qualitative obligatory milestone into a quantitative tool. Although alternatives have been suggested to replace teratoma formation, they all suffer from imperfections. The computational method PluriTest has shown promising results identifying PSCs and differentiating them from cancer stem cells and cells that have underwent differentiation (Müller et al., 2011). While this method appears simpler than in vivo or in vitro differentiations, it can only determine a cell resemblance to PSCs in their undifferentiated state. The uniqueness of PSCs, however, is in their ability to transform from the pluripotent state to terminally differentiated cell types. These differentiation processes require master regulators that are not expressed in PSCs. A mutation in one of these genes will not be exposed using PluriTest, claiming that a cell is pluripotent even though it cannot differentiate toward all germ layers. Differentiation per se can be evaluated using directed and spontaneous in vitro protocols. Directed differentiation protocols toward many cell types and tissues have been published, ranging from neurons to
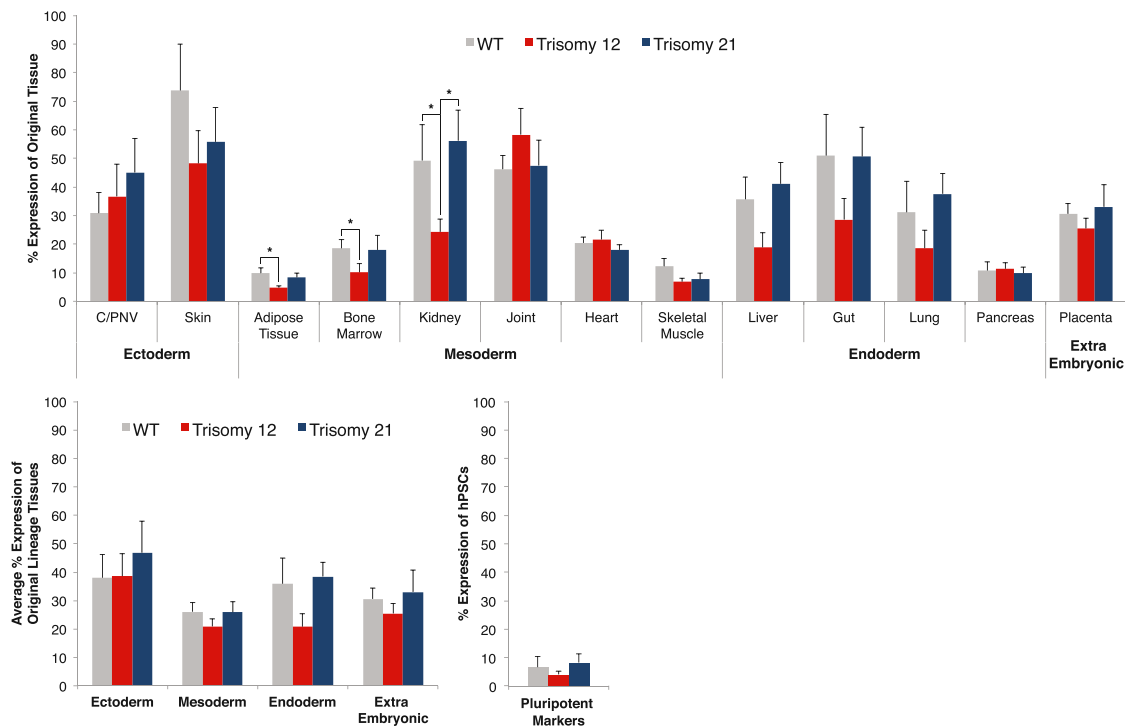
Figure 3. Average Tissue and Lineage Expression in WT and Chromosomally Aberrant Teratomas

Average gene expression was calculated in respect of its expression values in the original tissue. Lineage expression was calculated as the mean expression of all genes representing tissues in that lineage (*p < 0.05). WT (n = 5) and chromosomally aberrant teratomas (trisomy 12, n = 5; trisomy 21, n = 4) show a similar tissue and lineage expression, with exception for significantly lower expression of genes specific for kidney, adipose tissue, and bone marrow in teratomas with trisomy of chromosome 12 compared with WT. Error bars represent SE.

pancreatic progenitors (reviewed in Schuldiner and Benve- nisty, 2003; Tabar and Studer, 2014). Each protocol requires a set of specific factors and environment, presenting expen- sive and complex challenges to assess cell potency. Further- more, there is no agreed-upon assembly of protocols which successful utilization is sufficient in determining cells' plu- ripotency. In contrast, spontaneous differentiation is widely used and relatively simple to perform. EB formation has been suggested as a good replacement for teratoma formation. However, EBs do not always show complex differentiated structures and usually encapsulate a core of undifferentiated cells. When compared with teratomas, EBs show a lesser extant of mature tissue differentiation (Ozolek and Castro, 2011)—a difference that might pro- vide a limited estimation of the differentiation capabilities. Teratoma formation enables differentiation to a wide vari- ety of cells and tissues, presenting mature and complex structures. Although it is far from being a perfect assay, teratoma formation is still the best evaluator of pluripotent assessment.

Using expression data from teratomas, we were able to identify a large variety of differentiated tissues within

them. This allowed us to analyze a repertoire of differen- tiated cell types in each embryonic germ layer. Teratomas initiated from PSCs express characteristic genes of all analyzed tissues and a similar expression of all lineages. Other tumors are typically composed of uniform and ho- mogenous tissues, as seen by our 100-gene scorecard. We show that tumors originating from all lineages as well as from extraembryonic tissues express the respective genes in an exclusive manner. Weighing this expression creates a score that successfully separates PSC-derived mature tu- mors from tumors originating from other cell types. While this separation is evident, it is unclear why some teratomas provide a much higher grade in the TeratoScore analysis. We were not able to find a common ground for high or low scores, looking at karyotype integrity or ge- netic identity. In our study, we have analyzed two sets of different teratomas generated from the same cell types (CSES22 and CSES45; see Table S1). The TeratoScore of both pairs of teratomas was higher than 100, affirming the cell lines' pluripotency, but the numbers were some- what different (160 and 315 for CSES22-initiated tumors and 234 and 429 CSES45-initiated tumors). Examining a
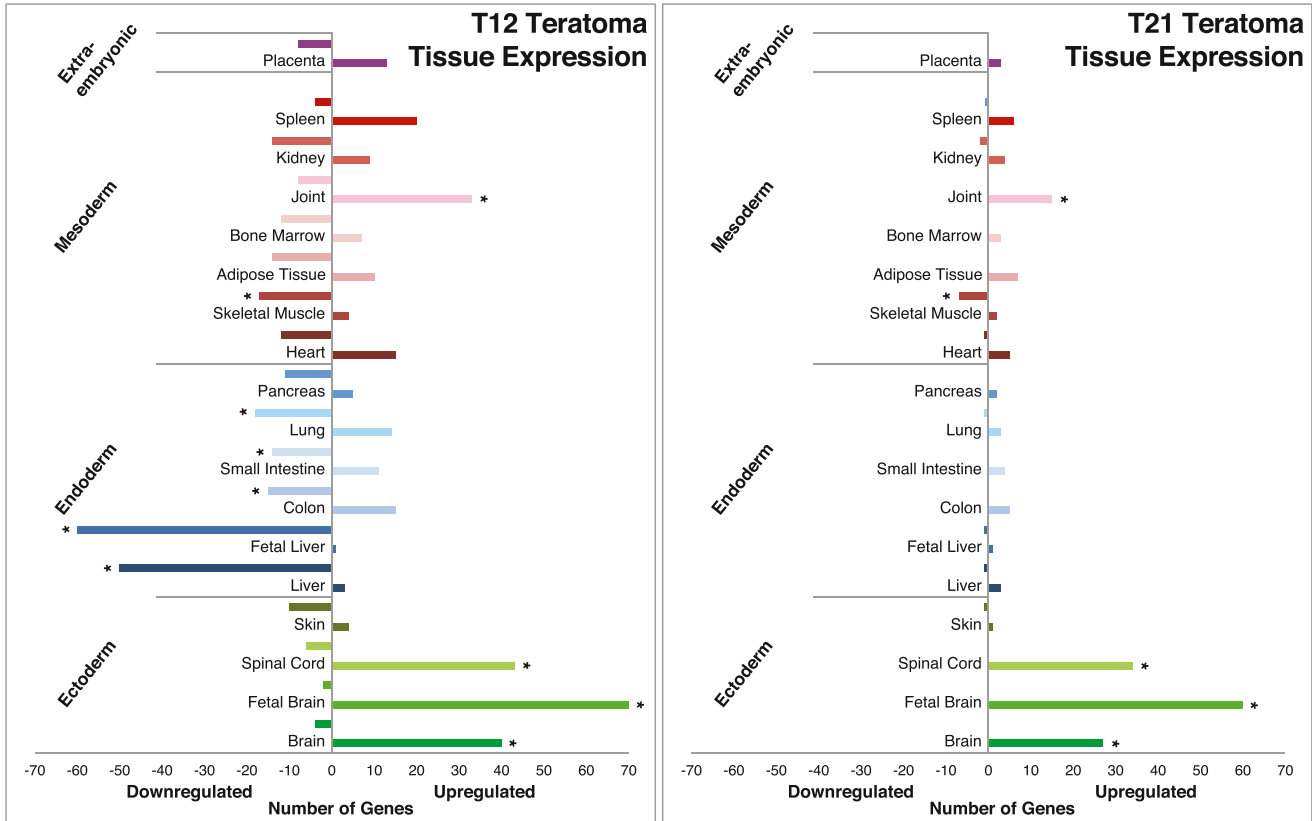
Figure 4. Tissue Classification of Genes Expressed Differently between WT and Aberrant Teratomas

Number of genes differentially expressed in chromosomally aberrant teratomas compared with teratomas initiated from diploid cells. Gene were considered differentially expressed when changed over 2-fold compared with control (*p < 0.001). Teratomas from cells carrying trisomy of chromosome 12 (T12, n = 5) showed a significant number of upregulated genes enriched for brain, fetal brain, spinal cord, and joint as well as downregulated genes enriched for liver and fetal liver, colon, small intestine, lung, and skeletal muscle compared with WT teratomas (n = 5). In contrast, teratomas from cells carrying trisomy of chromosome 21 (T21, n = 4) showed significant number of downregulated genes specific for skeletal muscle only, with a significant number of upregulated genes enriched for brain, fetal brain, spinal cord, and joint.

much larger teratoma sample from various sources and passages, including partially reprogrammed cells, might aid in identifying a biological origin for TeratoScore performance variation. It is possible that some of the variance originates from TeratoScore's limited sensitivity. Although we show that a 100-gene scorecard is enough to discover the existence or absence of tissues and lineage, it is insufficient in identifying possible lineage differentiation biases that might affect the TeratoScore outcome. This is also true for the measurement of pluripotency markers' expression. Pluripotent foci were previously shown to exist in teratomas initiated from trisomy 12 cells (Ben-David et al., 2014). While a small subset of pluripotent markers could not detect residual undifferentiated PSCs, analysis of the top 200 genes expressed in hPSCs revealed that 25 of them were upregulated T12 teratomas compared with WT teratomas (p < 0.06, following Bonfer-

roni correction for multiple testing) compared with only 5 in T21 teratomas (p < 0.8).

Several studies have analyzed the differentiation propensity of hPSCs in vitro, suggesting that different cell lines have differentiation bias, resulting in unequal ability to differentiate toward specific cell types (Bock et al., 2011; Boulting et al., 2011; Osafune et al., 2008). A method to evaluate these biases has been proposed (Bock et al., 2011), and their origins have been suggested to be both genetic and epigenetic (reviewed in Cahan and Daley, 2013). Furthermore, the International Stem Cell Initiative (ISCI) (Andrews et al., 2005) is currently conducting a large-scale study to compare different methods to analyze the pluripotency of human stem cells. Our aim in generating TeratoScore was to give a quantitative measure to differentiation of PSCs in vivo as a way to evaluate the pluripotency of the cells. Teratomas have long

been argued to be an intriguing developmental model; however, the extent of their resemblance to typical development is under debate (Ozolek and Castro, 2011). Using karyotypically abnormal cells, we were able to show significant changes in tissue-enriched gene expression. Although these changes were not very significant in the 100-gene scorecard made to identify pluripotent origins, they were highly significant when looking at the top 200 genes expressed in each tissue. Additional copies of chromosomes 12 or 21 caused a substantial increase in expression of genes related to the CNS. This upregulation may indicate these aneuploidies influence in vivo, for it is known that both conditions cause substantial intellectual disabilities (Chapman and Hesketh, 2000; Segel et al., 2006). Furthermore, case studies have reported joint and muscle malfunction in patients with Pallister-Killian mosaic syndrome (Speleman et al., 1991)—another significant trait in our analysis. It is important to note that we could not find a report of patients bearing a complete trisomy of chromosome 12. We can cautiously assume that this aberration is lethal—perhaps due to extant of developmental abnormalities implied in our analysis. In contrast, trisomy 21 seems to cause a significant downregulation in skeletal muscle-specific genes only—perhaps correlating with the successful birth of Down syndrome patients and their typical muscle hypotonia (Cisterna et al., 2013).

It is therefore still unclear how well a teratoma can reflect developmental processes in vivo. More data regarding these and other genetic conditions need to be collected before reaching final conclusions in this matter. However, we find these results encouraging in reflecting the implications chromosomal aberrations have on embryonic development.

Our scorecard and analysis are currently based on a specific microarray platform, but they might be applicable to other methods as well. Our 100-gene scorecard can be measured using a standard quantitative real-time PCR platform, generating a TaqMan low-density array (TLDA). As RNA sequencing is becoming more common and less expensive, it will also allow a reevaluation of our scorecard and TeratoScore analysis. This might enable a more elaborated analysis of teratomas, alongside additional biological material and data.

Our algorithm provides a fast and quantitative evaluation of cell potency. As teratomas are a reliable test for cell differentiation capacity, it is crucial to use this type of quantitative tools to evaluate and compare them. TeratoScore provides an estimation of tissue expression and lineage distribution, as well as a grade determining pluripotency. We show that teratomas originating from karyotypically aberrant cells present an aberrant gene expression, indicating that they might serve disease research and modeling.

EXPERIMENTAL PROCEDURES

Cell Culture

hPSCs were cultured and karyotyped using standard conditions as previously described (Ben-David et al., 2014). See also Supplemental Experimental Procedures.

Tumor Formation and Analysis

Teratomas derived from hPSCs were generated as previously described (Ben-David et al., 2014). Half of the tumor pieces were mixed together and taken for RNA extraction, while the rest were saved for H&E staining, carried out as previously described (Kopper et al., 2010). See also Supplemental Experimental Procedures. All experimental procedures in animals were in accordance with institutional guidelines.

Microarray Analysis

Microarray data of the various body tissues, hPSC, and previously published teratomas and tumors were downloaded from the GEO database (Table S2). An expression ratio was calculated between each tissue and the average of all other body tissues, indicating tissue specificity. Scorecard genes for each tissue were chosen by having a high tissue specificity (over 3-fold) and by being expressed in teratoma but not in hPSCs. Average expression for each lineage was calculated as the mean expression of all genes representing that lineage. TeratoScore grades were calculated as the multiplication of these means and dividing this product by 10^3. Top-200 tissue-enriched genes were chosen as the top unique genes in the tissue/body ratio mentioned above. Statistical calculations and figures were made in R programming language (http://www.r-project.org). See also Supplemental Experimental Procedures.

ACCESSION NUMBERS

The NCBI GEO accession numbers for the microarray data reported in this paper are GEO: GSM1576685 and GSE51455 (Table S1).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.stemcr.2015.05.006.

REFERENCES

Andrews, P.W., Benvenisty, N., McKay, R., Pera, M.F., Rossant, J., Semb, H., and Stacey, G.N.; Steering Committee of the International Stem Cell Initiative (2005). The International Stem Cell Initiative: toward benchmarks for human embryonic stem cell research. Nat. Biotechnol. *23*, 795–797.

Ben-David, U., Arad, G., Weissbein, U., Mandefro, B., Maimon, A., Golan-Lev, T., Narwani, K., Clark, A.T., Andrews, P.W., Benvenisty, N., and Carlos Biancotti, J. (2014). Aneuploidy induces profound changes in gene expression, proliferation and tumorigenicity of human pluripotent stem cells. Nat. Commun. *5*, 4825.

Biancotti, J.C., Narwani, K., Buehler, N., Mandefro, B., Golan-Lev, T., Yanuka, O., Clark, A., Hill, D., Benvenisty, N., and Lavon, N. (2010). Human embryonic stem cells as models for aneuploid chromosomal syndromes. Stem Cells *28*, 1530–1540.

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.

Boulting, G.L., Kiskinis, E., Croft, G.F., Amoroso, M.W., Oakley, D.H., Wainger, B.J., Williams, D.J., Kahler, D.J., Yamaki, M., Davidow, L., et al. (2011). A functionally characterized test set of human induced pluripotent stem cells. Nat. Biotechnol. *29*, 279–286.

Cahan, P., and Daley, G.Q. (2013). Origins and implications of pluripotent stem cell variability and heterogeneity. Nat. Rev. Mol. Cell Biol. *14*, 357–368.

Cahan, P., Li, H., Morris, S.A., Lummertz da Rocha, E., Daley, G.Q., and Collins, J.J. (2014). CellNet: network biology applied to stem cell engineering. Cell *158*, 903–915.

Chapman, R.S., and Hesketh, L.J. (2000). Behavioral phenotype of individuals with Down syndrome. Ment. Retard. Dev. Disabil. Res. Rev. *6*, 84–95.

Chin, M.H., Mason, M.J., Xie, W., Volinia, S., Singer, M., Peterson, C., Ambartsumyan, G., Aimiuwu, O., Richter, L., Zhang, J., et al. (2009). Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. Cell Stem Cell *5*, 111–123.

Cisterna, B., Costanzo, M., Scherini, E., Zancanaro, C., and Malatesta, M. (2013). Ultrastructural features of skeletal muscle in adult and aging Ts65Dn mice, a murine model of Down syndrome. Muscles Ligaments Tendons J. *3*, 287–294.

Itskovitz-Eldor, J., Schuldiner, M., Karsenti, D., Eden, A., Yanuka, O., Amit, M., Soreq, H., and Benvenisty, N. (2000). Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers. Mol. Med. *6*, 88–95.

Kopper, O., Giladi, O., Golan-Lev, T., and Benvenisty, N. (2010). Characterization of gastrulation-stage progenitor cells and their inhibitory crosstalk in human embryoid bodies. Stem Cells *28*, 75–83.

Kostanecka, A., Close, L.B., Izumi, K., Krantz, I.D., and Pipan, M. (2012). Developmental and behavioral characteristics of individuals with Pallister-Killian syndrome. Am. J. Med. Genet. A. *158A*, 3018–3025.

Lavon, N., Narwani, K., Golan-Lev, T., Buehler, N., Hill, D., and Benvenisty, N. (2008). Derivation of euploid human embryonic stem cells from aneuploid embryos. Stem Cells *26*, 1874–1882.

Le Carrour, T., Assou, S., Tondeur, S., Lhermitte, L., Lamb, N., Rème, T., Pantesco, V., Hamamah, S., Klein, B., and De Vos, J. (2010). Amazonia!: An Online Resource to Google and Visualize Public Human whole Genome Expression Data. Open Bioinformatics J. *4*, 5–10.

Müller, F.J., Goldmann, J., Löser, P., and Loring, J.F. (2010). A call to standardize teratoma assays used to define human pluripotent cell lines. Cell Stem Cell *6*, 412–414.

Müller, F.J., Schuldt, B.M., Williams, R., Mason, D., Altun, G., Papapetrou, E.P., Danner, S., Goldmann, J.E., Herbst, A., Schmidt, N.O., et al. (2011). A bioinformatic assay for pluripotency in human cells. Nat. Methods *8*, 315–317.

Müller, F.J., Brandl, B., and Loring, J.F. (2012). Assessment of human pluripotent stem cells with PluriTest. In StemBook, The Stem Cell Research Community, eds. (Cambridge, MA: Harvard Stem Cell Institute).

Narwani, K., Biancotti, J.C., Golan-Lev, T., Buehler, N., Hill, D., Shifman, S., Benvenisty, N., and Lavon, N. (2010). Human embryonic stem cells from aneuploid blastocysts identified by pre-implantation genetic screening. In Vitro Cell. Dev. Biol. Anim. *46*, 309–316.

Osafune, K., Caron, L., Borowiak, M., Martinez, R.J., Fitz-Gerald, C.S., Sato, Y., Cowan, C.A., Chien, K.R., and Melton, D.A. (2008). Marked differences in differentiation propensity among human embryonic stem cell lines. Nat. Biotechnol. *26*, 313–315.

Ozolek, J.A., and Castro, C.A. (2011). Teratomas derived from embryonic stem cells as models for embryonic development, disease, and tumorigenesis. In Embryonic Stem Cells-Basic Biology to Bioengineering, M. Kallos, ed. (InTech).

Patterson, D. (2009). Molecular genetic analysis of Down syndrome. Hum. Genet. *126*, 195–214.

Schuldiner, M., and Benvenisty, N. (2003). Factors controlling human embryonic stem cell differentiation. Methods Enzymol. *365*, 446–461.

Segel, R., Peter, I., Demmer, L.A., Cowan, J.M., Hoffman, J.D., and Bianchi, D.W. (2006). The natural history of trisomy 12p. Am. J. Med. Genet. A. *140*, 695–703.

Speleman, F., Leroy, J.G., Van Roy, N., De Paepe, A., Suijkerbuijk, R., Brunner, H., Looijenga, L., Verschraegen-Spae, M.R., and Orye, E. (1991). Pallister-Killian syndrome: characterization of the isochromosome 12p by fluorescent in situ hybridization. Am. J. Med. Genet. *41*, 381–387.

Tabar, V., and Studer, L. (2014). Pluripotent stem cells in regenerative medicine: challenges and recent progress. Nat. Rev. Genet. *15*, 82–92.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145–1147.

TeratoScore: Assessing the Differentiation

Potential of Human Pluripotent Stem Cells

by Quantitative Expression Analysis of Teratomas

Yishai Avior, Juan Carlos Biancotti, and Nissim Benvenisty

Supplementary Figure 1. Representative Histological Sections of Teratomas Showing Structures from All Three Germ Layers. The teratomas reported in this paper were heterogeneously comprised, with structures originating from endoderm, mesoderm and ectoderm lineages (marked with arrows). 4-μm sections of teratomas stained with hematoxylin and eosin are shown. The teratomas were initiated from CSES32 (top right), CSES21 (top left), CSES26 (bottom left) and CSES25 (bottom right) cells. Abbreviations: END: Endoderm, MES: Mesoderm, ECT: Ectoderm.
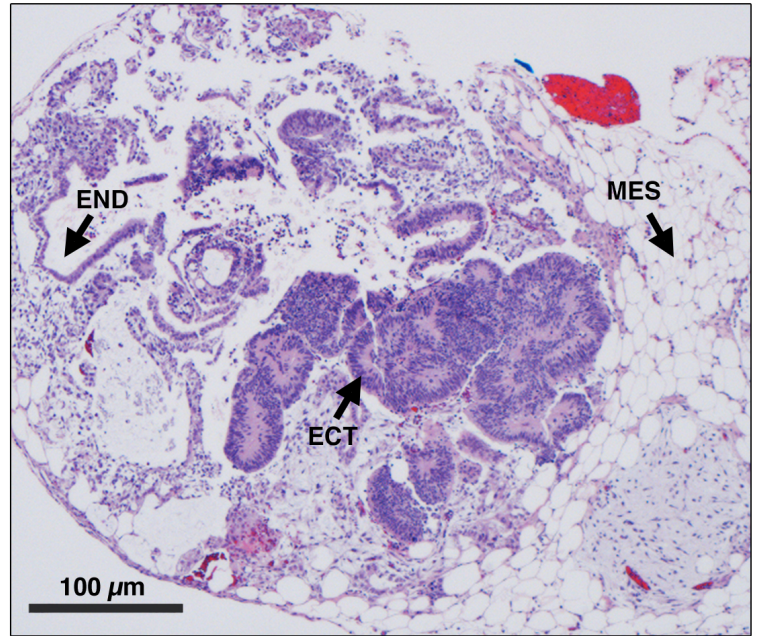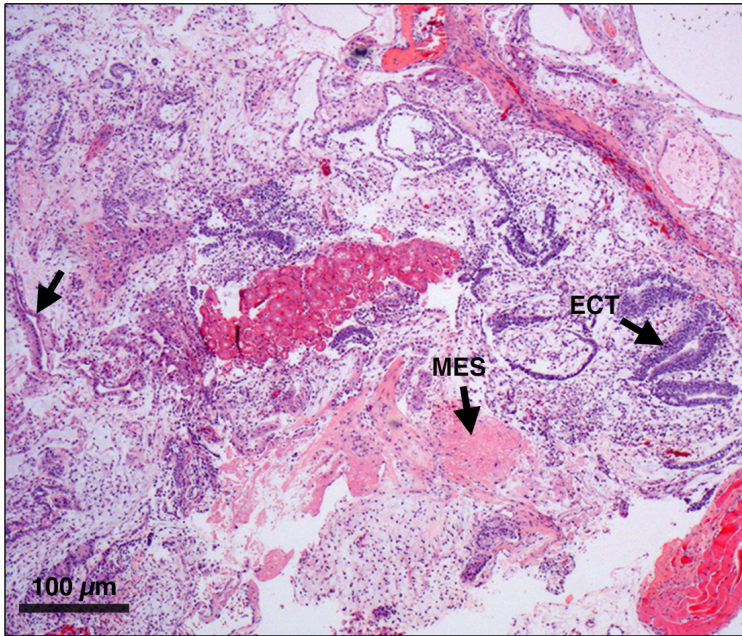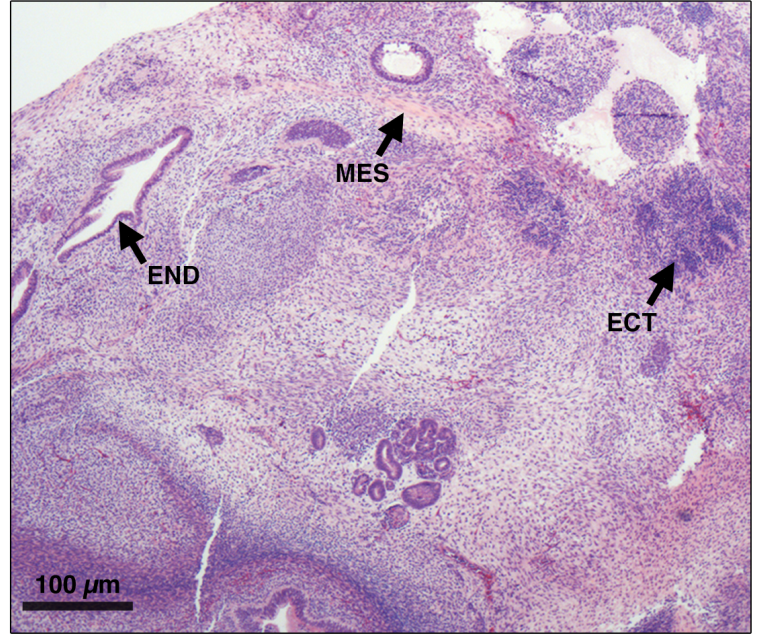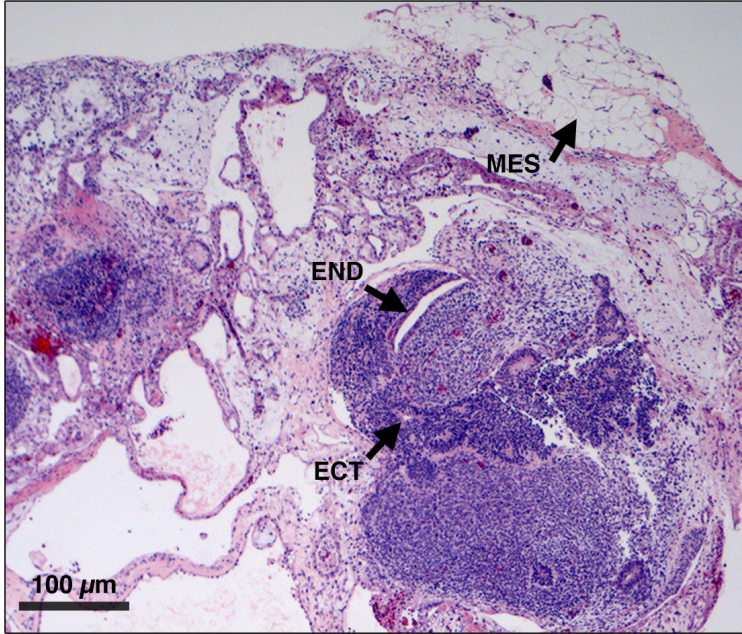
**Supplementary Figure 2. CellNet Online Resource Output for hESC, Teratoma and Tumor data.** Expression data sets from four human embryonic stem cell lines (ESCs), 14 teratomas and seven different somatic tumors were uploaded to the online CellNet resource (Cahan et al., 2014). While ESCs show a clear resemblance to the ESC categories, all teratomas do not show a distinct similarity to any of the cell types measured, as shown in both a heat map analysis of the individual data sets (A) and histograms of gene regulatory networks found in each teratoma-group representing different cell types (B). Somatic tumors do show a distinct classification according to their initiating tissue (A). Abbreviations: ESC: embryonic stem cells; BSC: basal cell carcinoma, ALL: acute lymphoblastic leukemia, HCC: hepatocellular carcinoma, CC: choriocarcinoma, HSC: Hematopoietic Stem Cells, S. Muscle: Skeletal Muscle.
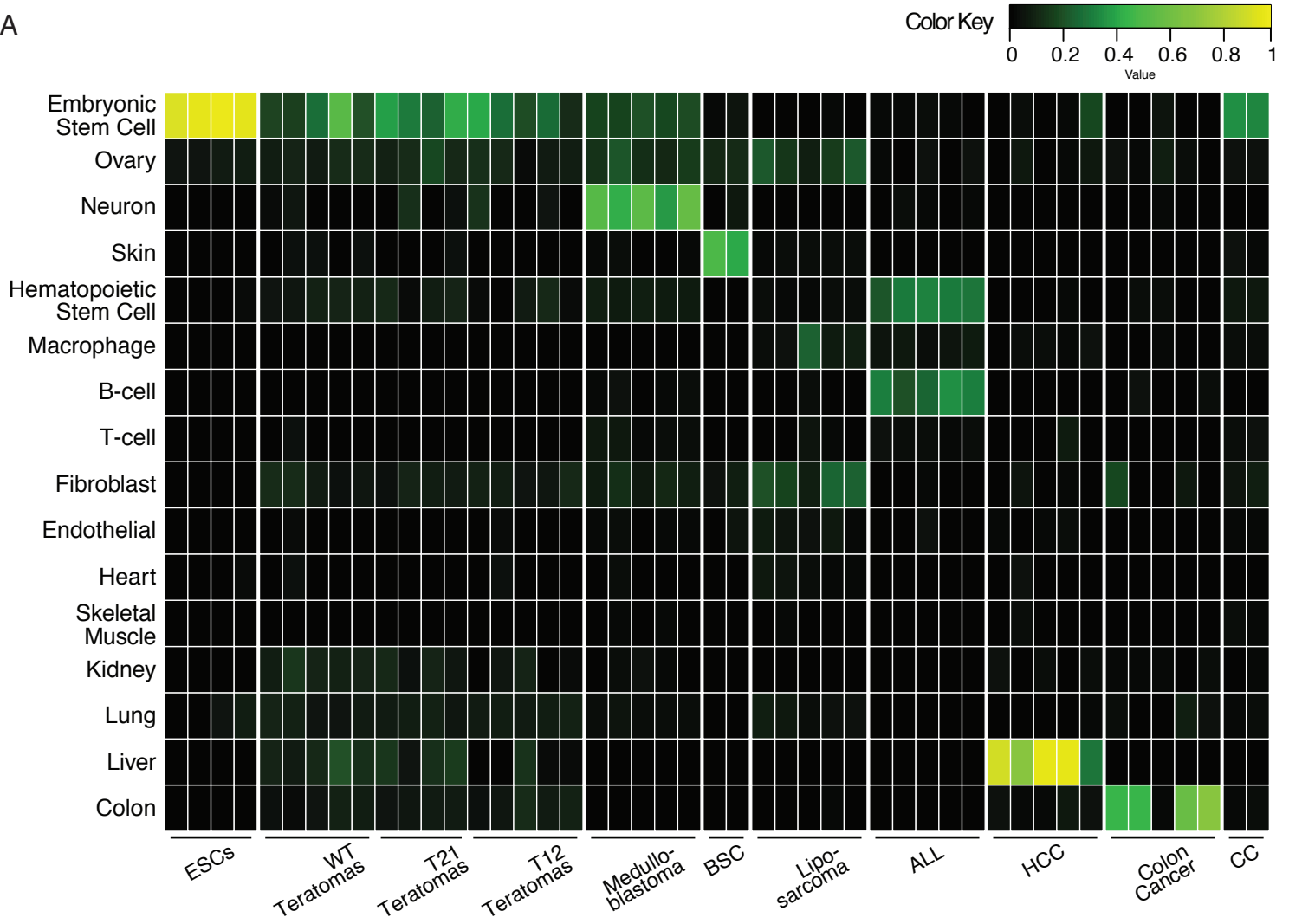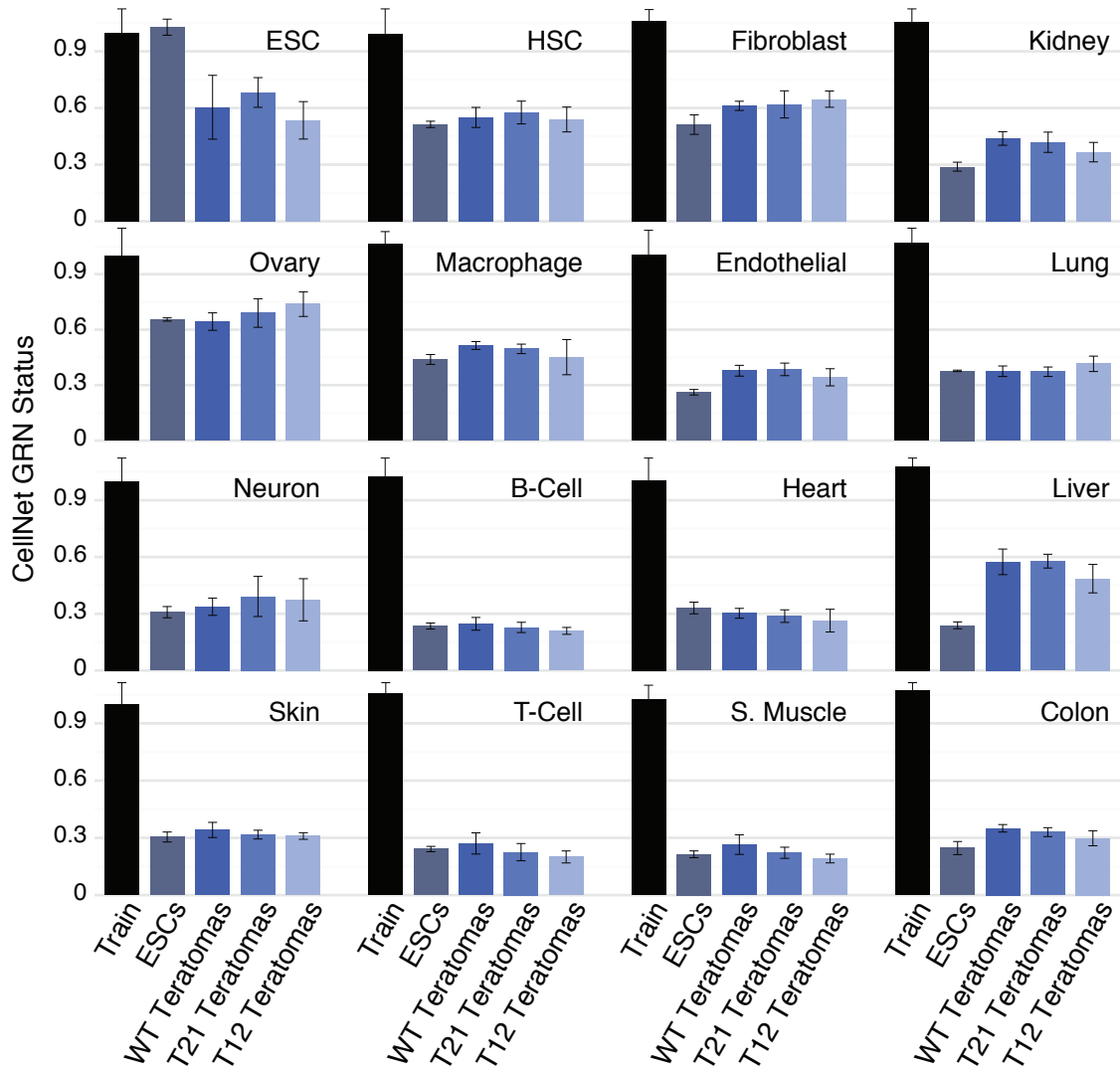
Supplementary Figure 3. TeratoScore Grades of Different Normal Tissues. Average tissue gene expression was calculated in respect of its expression values in the original tissue. Lineage expression was calculated as the mean expression of all genes representing tissues in that lineage. Error bars represent standard error. Each tissue shows a correct specific tissue and lineage expression. All tissues produced a low TeratoScore value, reflecting the lack of pluripotent origin.

Figure showing average % expression of original tissue/lineage across multiple tissue types (Central and Peripheral Nervous System, Skin, Adipose Tissue, Bone Marrow, Liver, Colon, Placenta) grouped by germ layer lineage (Ectoderm, Mesoderm, Endoderm).

Central and Peripheral Nervous System — TeratoScore Value: 0.8±0.2

Skin — TeratoScore Value: 6±0.4

Adipose Tissue — TeratoScore Value: 0.9±0.1

Bone Marrow — TeratoScore Value: 0.4±0.0

Liver — TeratoScore Value: 4.1±0.6

Colon — TeratoScore Value: 0.9±0.4

Placenta — TeratoScore Value: 6.5±0.5

Ectoderm: C/PNV, Skin
Mesoderm: Adipose Tissue, Bone Marrow, Kidney, Joint, Heart, Skeletal Muscle
Endoderm: Liver, Gut, Lung, Pancreas
Placenta

Lineages: Ectoderm, Mesoderm, Endoderm, Extraembryonic

Supplementary Table 1. Teratoma Expression Profiles' GEO Sample IDs. Cell lines used

to initiate the teratomas used in this study are listed, as well as their karyotype classification

and Gene Expression Omnibus (GEO) sample numbers.

Supplementary Table 1

Cell line	Karyotype	Geo Series ID	GEO Sample ID	Reference
CSES7	46, XX	GSE51455	GSM1246012	*
CSES12	46, XX	GSE64669	GSM1576685	*This Study*
CSES15	46, XX	GSE51455	GSM1246013	*
CSES26	46, XX	GSE51455	GSM1246015	*
CSES25	46, XX	GSE51455	GSM1246014	*
CSES22 (p16)	46, XX + 12	GSE64669	GSM1576686	*This Study*
CSES22 (p35)	46, XX + 12	GSE51455	GSM1246017	*
CSES45 (p20)	46, XX + 12	GSE51455	GSM1246018	*
CSES45 (p30)	46, XX + 12	GSE64669	GSM1576687	*This Study*
HiPS18	46, XX + 12	GSE51455	GSM1246019	*
CSES13	46, XX + 21	GSE64669	GSM1576688	*This Study*
CSES20	46, XX + 21	GSE64669	GSM1576689	*This Study*
CSES21	46, XX + 21	GSE64669	GSM1576690	*This Study*
CSES32	46, XX + 21	GSE64669	GSM1576691	*This Study*

* Ben-David et. al, 2014

Supplementary Table 2. Expression Profiles of Tissues and Tumors Used in this Study.

Expression profiles of tissues representing the three germ layers, the placenta and embryonic stem cells used to develop the TeratoScore are listed with their Gene Expression Omnibus (GEO) sample and series IDs. Tumors expression profiles used to compare TeratoScores are also listed with their Gene Expression Omnibus (GEO) sample and series IDs.

Supplementary Table 3. Genes Enriched for Tissues from Three Germ Layers and Placenta Used to Calculate TeratoScore. 100 genes were selected to represent tissues from all three germ layers and extraembryonic tissue in order to calculate tumors' TeratoScore. Genes enriched for each tissue were selected computationally and verified as being tissue-specific using Amazonia! online database (Le Carrour et al., 2010). Tissues with similar gene-expression were merged into a single, larger category. An additional set of ten genes representing pluripotent markers was selected. These genes are not taken into account when calculating TeratoScore. Listed are gene symbols, approved names and chromosomal locations.

Supplementary Table 4. Tissue-Specific Genes Up- and Downregulated in Teratomas Initiated from Aneuploid Pluripotent Cells. Tissue-specific genes differentially expressed in chromosomally aberrant teratomas compared to teratomas initiated from diploid cells. Out of the top 200 expressed genes in each tissue, a gene was considered differentially expressed when changed over 2-fold (either upregulated or downregulated) compared to control.

Supplemental Experimental Procedures

Cell Culture. Human embryonic stem cell lines CSES7, CSES 12, CSES13, CSES15, CSES20, CSES21, CSES22, CSES25, CSES26, CSES32 and CSES45 (Biancotti et al., 2010; Narwani et al., 2010) and iPSC line hiPS18 (Chin et al., 2009) were cultured using standard conditions as previously described (Mayshar et al., 2010). In brief, cells were grown on mouse embryonic fibroblast feeder (MEF) layer in KnockOut Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 15% KnockOut SR (Invitrogen), 1 mM glutamine, 0.1 mM b-mercaptoethanol (Sigma-Aldrich), 1% nonessential amino acids stock (Biological Industries), penicillin (50 U/ml), streptomycin (50 μg/ml) and FGF2 (4 ng/ml, Invitrogen). Cells were passaged using 0.25% trypsin solution (Biological Industries).

Karyotyping. A confluent, 10-cm plate of hESCs was used to analyze each cell line. Cells culture media were supplemented with KaryoMAX colcemid (100 ng/ml, Biological Industries) for 30 min at 37°C (with 5% CO2). Cultures were then trypsinized, treated with hypotonic solution and fixed using methanol:acetic acid solution (3:1). Metaphases were spread on glass slides and chromosomes were stained using standard G-banding technique. Chromosomal calssification was made using GenASIs software (Applied Spectral Imaging) according to the ISCN (International System for Human Cytogenetic Nomenclature).

Tumor Formation and Analysis. Undifferentiated hPSC colonies cultured on MEFs were trypsinized into single cells. ~3x106 cells were re-suspended in 50–100 μl of phosphate buffered saline (PBS) and injected under the kidney capsule of 6- to 8-week-old NOD-SCID mice. Three to four weeks post-injection, mice were sacrificed and resulting teratomas were extracted and dissected. Tumors were dissected into smaller pieces from different areas. Half of the tumor pieces were mixed together and taken for RNA extraction while the rest were saved for H&E staining, carried out as previously described (Kopper et al., 2010). The RNA portion was immediately homogenized in RLT buffer (RNeasy Mini kit, Qiagen) using a Polytron PT 1200 E system (Kinematica).

Microarray and Data Analysis. Total RNA was extracted from hPSC-derived teratomas using RNeasy mini kit (Qiagen), according to the manufacture instruction. RNA was subjected to Human Genome U133 Plus 2.0 microarray platform (Affymetrix); washing and scanning were performed according to the manufacturer's protocol. Original microarray data are accessible at the NCBI Gene Expression Omnibus database under the accession numbers GSM1576685 and GSE51455 (Table S1). Microarray data of the various body tissues, hPSC and previously published teratomas and tumors were downloaded from the Gene Expression Omnibus database (Table S2). Arrays were normalized using RMA algorithm in the Affymetrix Expression Console. Probe sets with expression values lower than 20 were raised to this level. The karyotype of the analyzed in-house cell lines was validated by virtual karyotyping (Ben-David et al., 2013).

Following averaging of samples from the same tissue, an expression ratio was calculated between each tissue and the average of all other body tissues, indicating tissue-specificity. Scorecard genes for each tissue were chosen by having a high tissue specificity (over 3 folds) and by being expressed in teratoma but not in hPSCs. Tissue-specific gene expression was further validated using Amazonia!, with a requirement for distinct tissue expression (an order of magnitude over all or most other tissues) (Le Carrour et al., 2010). Gene lists for central and peripheral nervous system, small intestine and colon, fetal brain and brain and fetal liver and liver were each merged together, as their specific-gene expression was considerably similar. Average expression for each lineage was calculated as the mean expression of all genes representing that lineage. TeratoScore grades were calculated as the multiplication of these means and dividing this product by 103. One-tail t-tests between tissue expressions were done using alpha 0.05. Top-200 tissue enriched genes were chosen as the top unique genes in the tissue/body ratio mentioned above. Genes were considered differentially expressed when changed over 2-fold compared to WT. To determine significance a chi-square independence test was performed followed by Bonferroni correction for multiple testing. Statistical calculations and figures were made in R.

Ben-David, U., Mayshar, Y., and Benvenisty, N. (2013). Virtual karyotyping of pluripotent stem cells on the basis of their global gene expression profiles. Nat Protoc 8, 989-997.

Ben-David, U., Arad, G., Weissbein, U., Mandefro, B., Maimon, A., Golan-Lev, T., Narwani, K., Clark, A.T., Andrews, P.W., Benvenisty, N., et al. (2014). Aneuploidy induces profound changes in gene expression, proliferation and tumorigenicity of human pluripotent stem cells. Nat Commun 5, 4825.

Biancotti, J.C., Narwani, K., Buehler, N., Mandefro, B., Golan-Lev, T., Yanuka, O., Clark, A., Hill, D., Benvenisty, N., and Lavon, N. (2010). Human embryonic stem cells as models for aneuploid chromosomal syndromes. Stem Cells 28, 1530-1540.

Cahan, P., Li, H., Morris, S.A., Lummertz da Rocha, E., Daley, G.Q., and Collins, J.J. (2014). CellNet: network biology applied to stem cell engineering. Cell 158, 903-915.

Chin, M.H., Mason, M.J., Xie, W., Volinia, S., Singer, M., Peterson, C., Ambartsumyan, G., Aimiuwu, O., Richter, L., Zhang, J., et al. (2009). Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. Cell Stem Cell 5, 111-123.

Kopper, O., Giladi, O., Golan-Lev, T., and Benvenisty, N. (2010). Characterization of gastrulation-stage progenitor cells and their inhibitory crosstalk in human embryoid bodies. Stem Cells 28, 75-83.

Le Carrour, T., Assou, S., Tondeur, S., Lhermitte, L., Lamb, N., Rème, T., Pantesco, V., Hamamah, S., Klein, B., and De Vos, J. (2010). Amazonia!: An Online Resource to Google and Visualize Public Human whole Genome Expression Data. Open Bioinformatics Journal 4.

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.

Narwani, K., Biancotti, J.C., Golan-Lev, T., Buehler, N., Hill, D., Shifman, S., Benvenisty, N., and Lavon, N. (2010). Human embryonic stem cells from aneuploid blastocysts identified by pre-implantation genetic screening. In Vitro Cell Dev Biol Anim 46, 309-316.

Schuldiner, M., and Benvenisty, N. (2003). Factors Controlling Human Embryonic Stem Cell Differentiation. Methods Enzymol 365, 446-461.