Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors

Jamie M. Sperger*†, Xin Chen†‡, Jonathan S. Draper§, Jessica E. Antosiewicz*, Chris H. Chon¶, Sunita B. Jones¶, James D. Brooks¶, Peter W. Andrews§, Patrick O. Brown‡‖**, and James A. Thomson*,**††

*Wisconsin National Primate Research Center and ††Department of Anatomy, School of Medicine, University of Wisconsin, Madison, WI 53715; ‡Howard Hughes Medical Institute, Departments of ¶Urology and ‖Biochemistry, Stanford University School of Medicine, Stanford, CA 94305; and §Department of Biomedical Science, University of Sheffield, Sheffield S10 27N, United Kingdom

Communicated by Neal L. First, University of Wisconsin, Madison, WI, September 8, 2003 (received for review June 11, 2003)

Remarkably little is known about the transcriptional profiles of human embryonic stem (ES) cells or the molecular mechanisms that underlie their pluripotency. To identify commonalties among the transcriptional profiles of different human pluripotent cells and to search for clues into the genesis of human germ cell tumors, we compared the expression profiles of human ES cell lines, human germ cell tumor cell lines and tumor samples, somatic cell lines, and testicular tissue samples by using cDNA microarray analysis. Hierarchical cluster analysis of gene expression profiles showed that the five independent human ES cell lines clustered tightly together, reflecting highly similar expression profiles. The gene expression patterns of human ES cell lines showed many similarities with the human embryonal carcinoma cell samples and more distantly with the seminoma samples. We identified 895 genes that were expressed at significantly greater levels in human ES and embryonal carcinoma cell lines than in control samples. These genes are candidates for involvement in the maintenance of a pluripotent, undifferentiated phenotype.

Embryonic stem (ES) cells are pluripotent cells that maintain the ability to differentiate to derivatives of all three embryonic germ layers, even after prolonged culture in the undifferentiated state (1, 2). A few genes involved in maintaining the ES cell state, such as Oct4 (3–5), fibroblast growth factor 4 (6–8), FoxD3 (9, 10), Sox2 (11), Nanog (12, 13), and genes involved in the leukemia inhibitory factor (LIF) signaling pathway (14, 15) have been functionally characterized in the mouse, but for human ES cells (16) almost nothing is known about the regulation of self-renewal and pluripotency. Microarray analysis allows broad comparisons of patterns of gene expression between different cell types. For example, microarray analysis has been used to define commonalties between mouse embryonic and adult stem cells (17, 18) and has recently been used to compare H1 human ES cells with differentiated cells (19). Here, we performed microarray analysis to identify commonalties among different pluripotent human cells and to look for clues to the genesis of human germ cell tumors (GCTs).

Mouse and human GCTs arise from primordial germ cells (PGCs), presumably from some dysfunction in their normal development shortly after they populate the genital ridge (20, 21). Teratocarcinomas are a particular type of GCT that contain undifferentiated stem cells, termed embryonal carcinoma (EC) cells, and differentiated derivatives that can include all three germ layers. In the laboratory mouse, testicular teratocarcinomas develop spontaneously in males of the 129 strain, and they can be induced in several other strains by transplantation of the genital ridge to ectopic sites between 11 and 13 days of embryonic development (21). When murine PGCs are cultured *in vitro* under appropriate conditions, they convert to cells that closely resemble EC and ES cells, and these are known as embryonic germ cells to distinguish their origins (22).

Human teratocarcinomas most commonly arise in the testes of young adult men (23). There is strong circumstantial evidence to support the general consensus that, as in the laboratory mouse, these tumors also arise from PGCs during early embryonic development, and consequently, they are classified as GCTs (24). The pathology of the human GCT is somewhat different from that of the mouse, the occurrence of a subset known as seminomas being the most striking difference (24). Seminomas are generally histologically uniform and seem to resemble a transformed state of the PGC. Nonseminomatous GCTs, on the other hand, typically include teratocarcinomas with EC components, as well as cell types corresponding to the extra embryonic tissues, the yolk sac as yolk sac carcinomas (YSCs), and the trophoblast as choriocarcinomas. Both teratocarcinomas and seminomas appear to arise, initially, as abnormal germ cells that form *carcinoma-in situ* (CIS) within the seminiferous tubules (25). It seems that human GCTs develop progressively, first by transformation of PGCs to form CIS that then progress into an invasive form. If during this process they retain the general features of CIS and PGCs, they constitute a seminoma. Alternatively, they may convert into a cell type resembling the embryonic inner cell mass (ICM). In this event, the cells further progress to an EC state, from which differentiation yields all of the various somatic and extraembryonic cell types that together constitute a teratocarcinoma. In this view, EC cells resemble ES cells derived directly from the ICM but are variants that have adapted to tumor growth; in the terminology of Pierce (26), they are caricatures of ES cells. Seminomas do not occur in the laboratory mouse in which the very earliest intratubular GCTs already contain EC cells. Moreover, whereas murine and human EC and ES cells show marked differences in the expression of markers and some aspects of cell biology, it is apparent that, like human EC cells, mouse EC cells also resemble the ICM cells of the peri-implantation blastocyst (27).

From this view of GCT development in humans, we might expect the gene expression patterns of seminoma, EC, and ES cells to be distinct but closely related, and the EC and ES expression patterns should more closely resemble one another than either resembles the seminoma cluster. Genes that are expressed in common between seminoma and EC cells, but not by ES cells, may reflect general aspects of their tumor origins. Further, differences between EC and ES cells may reflect adaptation of EC cells to tumor growth by suppression of differentiation and promotion of self-renewal and survival, with a reduced dependency on external cues for self-maintenance. Lastly, genes shared by ES and EC cell lines may reflect genes that are necessary for maintenance of a pluripotent or an undifferentiated phenotype.

Abbreviations: EC, embryonal carcinoma; ES, embryonic stem; GCT, germ cell tumor; ICM, inner cell mass; PGC, primordial germ cell; SAM, significance analysis of microarrays; YSC, yolk sac carcinoma.

†J.M.S. and X.C. contributed equally to this work.

**To whom correspondence may be addressed at: 1220 Capitol Court, Madison, WI 53715; e-mail: thomson@primate.wisc.edu (for J.A.T.); or Beckman B400, Stanford, CA 94305-5307; e-mail: patrick.brown@stanford.edu (for P.O.B.).

www.pnas.org/cgi/doi/10.1073/pnas.2235735100

Materials and Methods

Cell Lines, Cell Culture, and RNA Isolation. The human ES cell lines were derived from the ICM of blastocyst-stage embryos (16). The ES cell lines are all capable of both somatic and extraembryonic (trophoblast) differentiation. The human EC cell lines, NTERA2, 1777N, 2102Ep, 833KE, GCT27, and TERA1, have all been derived from testicular teratocarcinomas, whereas NCCIT was derived from an extragonadal GCT (28). NTERA2 (29), GCT27 (30, 31), 1777N (32), and NCCIT (33) have been reported to be capable of significant somatic and/or extraembryonic differentiation. In contrast, 2102Ep, 833KE, and TERA1 no longer show significant somatic or extraembryonic differentiation. Of these, 2102Ep and 833KE form xenograft tumors with a pure EC histology in *nu/nu* athymic mice (34–36). 1411H and GCT44 were both derived from human testicular GCTs but exhibit the features of YSC cells rather than human EC cells (30, 37). Both 1411H and GCT44 lack the expression of typical EC marker antigens and form endodermal sinus tumors when grown as xenografts in *nu/nu* athymic mice. Primary tissue samples were clinical samples from Stanford University Medical School. The 17 somatic cell lines are described in Table 4, which is published as supporting information on the PNAS web site.

The five human embryonic stem cell lines were grown as described (38). The human EC cell lines, NTERA2, NCITT, 2102EP, TERA1, 833KE, and 1777N (1777Nrpmet), were maintained by growth in DMEM (Invitrogen) supplemented with 10% FCS (GIBCO) and 2 mM L-glutamine (Invitrogen) (DMEM/FCS) in a 37°C humidified atmosphere of 10% CO_2 in air. NTERA2 were passaged by scraping with sterile, 3-mm glass beads (Philip Harris Scientific, Ashby-de-la-zouch, Leicestershire, U.K.) and were then seeded onto fresh tissue culture plastic (Nunc). All other cell lines were harvested by treatment with 1 ml of 0.25% trypsin (wt/vol)/1 mM EDTA (Invitrogen) per T75 flask for 5 min at 37°C, harvested with DMEM/FCS, centrifuged, and then resuspended in DMEM/FCS onto fresh tissue culture plastic (Nunc).

GCT44 and GCT27 were grown in 80% DMEM and 20% defined FBS (HyClone), with 1 mM L-glutamine, 0.1 mM 2-mercaptoethanol (Sigma), and 1% nonessential amino acids stock (Invitrogen) in a 37°C humidified incubator with 5% CO_2 in air. Cultures were passaged on mitotically inactivated mouse embryonic fibroblast (MEF) feeder layers as cells became confluent by incubation at 37°C by treatment with 3 ml of 0.25% trypsin (wt/vol)/1 mM EDTA (Invitrogen) per T75 flask for 5 min at 37°C, harvested with DMEM/FBS, centrifuged, and then resuspended in DMEM/FBS onto fresh tissue culture plastic (Falcon). After several passages, the GCT44 cell line was removed from feeder layers, and RNA was extracted several passages later.

To harvest cells for total RNA extraction, the EC cells were washed once with PBS, harvested by scraping with 3-mm glass beads and centrifuged to produce a pellet. ES cells were separated from the MEF feeder layer by adding 1 ml of 1 mg/ml collagenase IV (Invitrogen) to each well of a six-well plate and incubated ≈10–15 min at 37°C in a humidified chamber until colonies began to pull away from the feeder layers. At this time, 0.5 ml of 10 mg/ml dispase (Invitrogen) in media was added to each culture well and further incubated until colonies became detached and were recovered in the supernatant. Either TRI reagent (Sigma) or RNA-STAT 60 (Tel-Test, Friendswood, TX) was used to isolate the RNA according to the manufacturer's instructions. Subsequently, mRNA was extracted from total RNA with the FastTrack mRNA isolation kit (Invitrogen).

Testis tissue samples were flash-frozen on dry ice at the time of radical or simple orchiectomy and stored at −80°C. Histological characterization and purity of all samples were verified by frozen section before RNA extraction. Poly(A) mRNA extraction and purification were performed according to published protocols (ref. 39; The Brown Lab, http://brownlab.stanford.edu).

Microarray Procedure and Data Analysis. Microarray procedures were as described (39, 40). Briefly, for each hybridization, 2 μg of purified mRNA from each sample was reverse-transcribed, labeled with fluorescence-tagged nucleotides, and hybridized against 1.5 μg of a common reference pool of mRNA for 14–18 h at 65°C on cDNA microarrays containing >44,000 elements, representing ≈30,300 unique genes, including characterized genes and expressed sequence tags. After several washes, microarrays were scanned with a GenePix 4000 microarray scanner and analyzed with GENEPIX 3.0 software (Axon Instruments, Union City, CA). On visual inspection, spots of insufficient quality were excluded from further analysis. Data files containing fluorescence ratios were entered into the Stanford Microarray database (http://genome-www5.stanford.edu). Spots were required to have a signal over background of at least 1.5 for either the sample or the reference to be included in the analysis. A total of 45,087 spots met this criteria. We selected genes that had absolute value of $\log_2$-normalized red/green ≥1.59 from the mean expression values in at least four arrays across the sample set. Genes with technically inadequate measurements in >25% of spots were excluded from the analysis, resulting in 9,796 spots being used in the cluster analysis. The genes and the samples were organized by hierarchical clustering with the Pearson correlation metric and average linkage clustering (41). To demonstrate the statistical significance of differences in expression, significance analysis of microarrays (SAM) was performed on the genes selected for cluster analysis, as described (42).

Results

The mRNA expression pattern from 74 different cell lines or tissue samples was examined by using cDNA microarray analysis. We compared data from five different human ES cell lines (H1.1, H7, H9, H13, and H14) with 36 human GCT cell lines or tumor samples, 14 samples of normal testis and 17 somatic cell lines. The GCTs included 7 human EC cell lines (NTERA2, NCITT, 2102EP, TERA1, 833KE, GCT27, and 1777N), 2 YSC cell lines (GCT44 and 1141H), 2 primary yolk sac tumors, 2 primary ECs, and 23 primary seminoma tumors.

We used hierarchical cluster analysis to sort the 9,976 clones that were expressed at least 3-fold different from the mean expression values across all of the samples in at least four different arrays (Fig. 1). RT-PCR was performed to confirm the expression profiles of a number of the genes (Fig. 3, which is published as supporting information on the PNAS web site). Genes with similar expression patterns are clustered together. The different experiments were also clustered, and a dendrogram, illustrating the relationship between the expression patterns, is shown with arrays with the most similar gene expression profiles clustering together with the shortest branches (Fig. 1*A*). The five ES cell lines cluster together, indicating that their expression patterns are more similar to each other than to any of the other cell types in this analysis. The ES cells cluster as a branch of the EC cell lines, and the YSC cell lines cluster separately from the other EC cell lines, with significant differences in their expression patterns. The primary tissues clustered as a distinct branch separate from the cultured cells. Within the tissue samples, the ECs clustered with the seminomas and separated from YSCs and normal testis, which represent more differentiated tissues.

One goal of this analysis was to identify genes specifically expressed at a higher level in pluripotent cell types. To do this, we used SAM to determine the genes with statistically different expression in ES cells, EC cells, or seminomas relative to control-differentiated samples, including both somatic cell lines

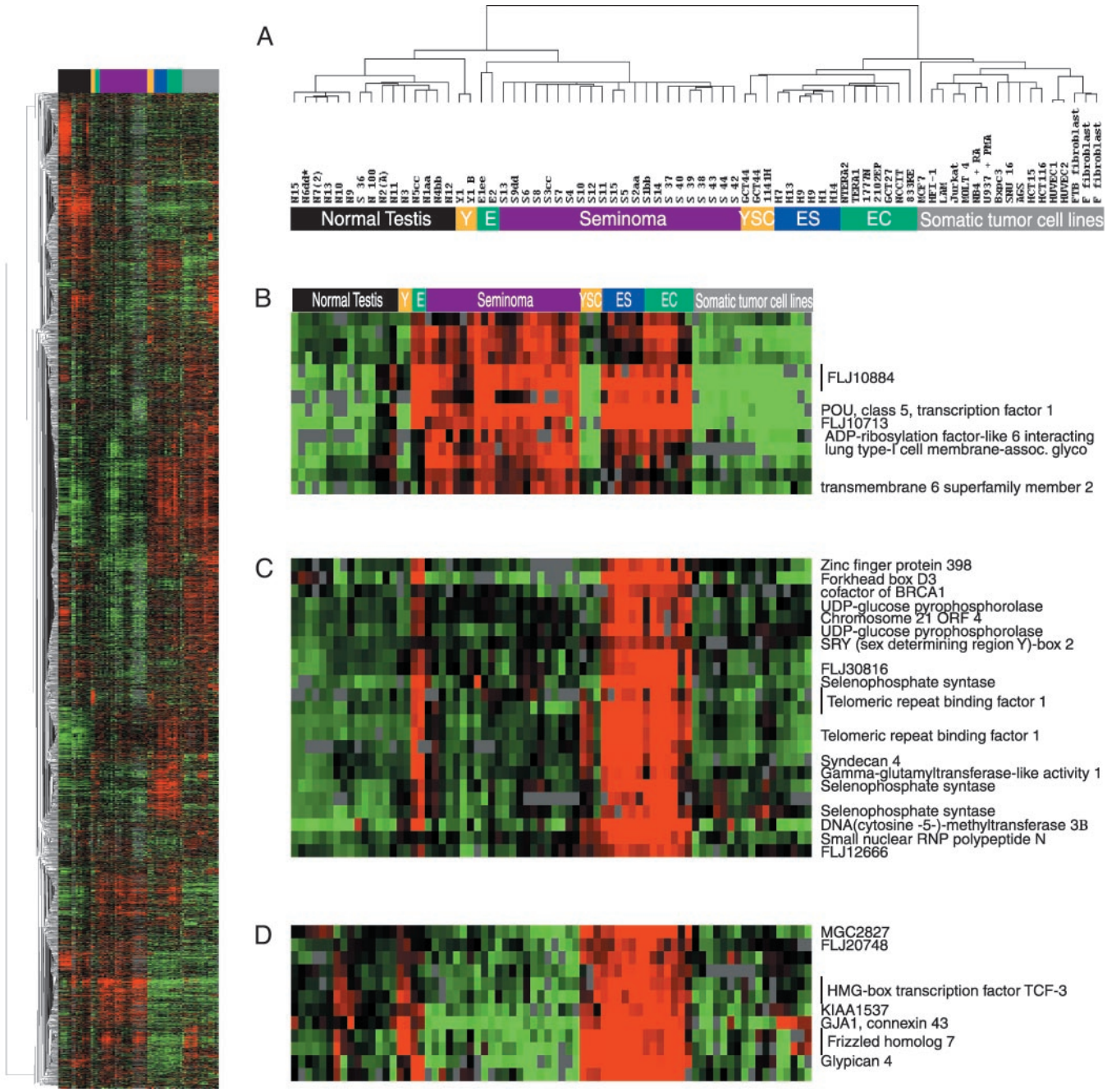

Fig. 1. Samples of the hierarchical cluster analysis of 9,976 cDNA clones that were expressed at least 3-fold different from the mean expression values across all of the samples in at least four different arrays. The colors indicates the relative expression levels of each gene, with red indicating positive expression above reference and green indicating negative expression below the reference. (*A*) The dendrogram represents the relationship of the expression patterns. The lengths of the arms are proportional to the similarities of the expression patterns, with shorter arms indicating closer relationships. (*B*) Cluster of genes surrounding *POU5F1*. These genes are highly expressed in ES and EC cells and seminomas but not YSCs and other differentiated tissue types. (*C*) Cluster of genes highly expressed in ES and EC cells but not seminomas, YSCs, or other differentiated tissues. (*D*) Cluster of genes including members of Wnt-β-catenin pathway expressed in ES and EC cells and YSCs. See Table 5, which is published as supporting information on the PNAS web site, or http://microarray-pubs.stanford.edu/es_cells_2/supplement.shtml for a complete list of the genes.

and normal testis. SAM identified 1,760, 1,299, and 1,518 cDNAs in ES and EC cells and seminomas, respectively, that were significantly more highly expressed than the control-differentiated cells (Tables 6–8, which are published as supporting information on the PNAS web site). Each of these samples has a median number of false significant genes of $<2\%$. The results from SAM are consistent with the hierarchical cluster analysis. ES and EC cells share 895 highly expressed genes, over half of the genes categorized as highly expressed genes in either ES or EC cells. This is a much higher number than either EC (481 genes) or ES (449 genes) cells share with seminomas. The genes shared by ES and EC cells represent a pool of genes that may be important in the maintenance of the pluripotent phenotype.

Among the 330 genes identified as highly expressed in ES and EC cells and seminomas is *POU5F1* (Oct4), a transcription factor that has been previously shown to be expressed only in the pluripotent

Table 1. The 25 most positively significant genes in human ES cells as defined by SAM

Clone ID	Gene name	Fold change
768508	*FLJ10713*, hypothetical protein FLJ10713	60.43
1607129	*POU5F1*, POU domain, class 5, transcription factor 1	51.34
276915	*DNMT3B*, DNA (cytosine-5-)-methyltransferase 3β	50.97
135773	*TERF1*, telomeric repeat binding factor 1	15.45
645485	*Homo sapiens* cDNA *FLJ36616*	16.18
809694	*CRABP1*, cellular retinoic acid binding protein 1	29.36
1551722	*AA922660*	15.79
504774	*GGTLA1*, γ-glutamyltransferase-like activity 1	15.15
358217	*GPC4*1, glypican 4	28.93
301878	*SCGB3A2*, secretoglobin, family 3A, member 2	22.64
1586535	*FOXD3*, forkhead box D3	37.48
50354	*OTX2*, orthodenticle homolog 2 (*Drosophila*)	40.76
52430	*SALL2*, sal-like 2 (*Drosophila*)	24.60
436070	*CA14*, carbonic anhydrase XIV	16.03
291448	*SILV*, silver homolog (mouse)	14.85
239256	*FZD7*, frizzled homolog 7 (*Drosophila*)	18.65
1946026	*FLJ10884*, hypothetical protein FLJ10884	24.11
341328	*TPM1*, tropomyosin 1 (α)	13.66
504763	*SDC4*, syndecan 4 (amphiglycan, ryudocan)	10.43
1499830	*H. sapiens* cDNA *FLJ25967*	21.22
743426	*KIAA1576*, KIAA1576 protein	12.38
204335	*CD24* or *FLJ37889*	16.96
788667	Multiple unigene clusters	14.66
32381	*GABRB3*, γ-aminobutyric acid (GABA) A receptor, β3	14.79
488651	*MGC16491*, hypothetical protein MGC16491	18.65

Table 2. The 25 most positively significant genes in human EC cells as defined by SAM

Gene ID	Gene name	Fold change
768508	*FLJ10713*, hypothetical protein FLJ10713	56.36
1607129	*POU5F1*, POU domain, class 5, transcription factor 1	66.73
276915	*DNMT3B*, DNA (cytosine-5-)-methyltransferase 3β	44.83
1586535	*FOXD3*, forkhead box D3	30.12
1946026	*FLJ10884*, hypothetical protein FLJ10884	24.52
1941870	ESTs, weakly similar to SWI/SNF related	15.76
432477	ESTs	17.16
136668	ESTs	14.30
358217	*GPC4*, glypican 4	14.62
1893670	ESTs	15.52
135773	*TERF1*, telomeric repeat binding factor	10.47
488651	*MGC16491*, hypothetical protein MGC16491	17.22
645485	*H. sapiens* cDNA *FLJ36616*	12.08
251404	EST	14.52
81409	*GABARAPL1*, GABA(A) receptor-associated protein like 1	16.99
208078	IMP-1, IGF-II mRNA-binding protein 1	11.98
1551722	*AA922660*	10.73
1628121	ESTs	18.58
1849084	*SEMA6A*, sema domain, (semaphorin) 6A	12.90
504774	*GGTLA1*, γ-glutamyltransferase-like activity 1	8.68
32381	*GABRB3*, GABA A receptor, β3	12.19
2237353	*GAL*, galanin, Hs. 278959, A1623173	14.10
824933	*AK3*, adenylate kinase 3	9.70
246808	*NALP2*, NALP2 protein	9.69
810873	*SCNN1A*, sodium channel, nonvoltage-gated 1 α	14.10

cells of the embryo and to promote differentiation when down-regulated (43, 44). *POU5F1* is one of the top two significantly differentially expressed genes in ES and EC cells and seminoma (Fig. 1*B*). The cluster analysis shows that *POU5F1* is highly expressed in these cell types, but not in YSCs, somatic tumor samples, or normal testis (Fig. 1*B*). Because *POU5F1* is a central regulator of pluripotency, genes having a similar expression pattern may also be central to maintaining pluripotent cells. The gene that clusters closest to *POU5F1* is *FLJ10713*, among the top five genes in ES and EC cells and seminomas. A homolog of *FLJ10713* was also shown to be highly expressed in mouse ES cells (17). ES and EC cells and seminoma share a total of 330 genes that SAM identified as positively significant genes. Forty percent of these genes are defined only as an accession number, EST, or hypothetical protein, and many of the other named genes have poorly defined functions, highlighting the fact that the genes with critical roles in early embryogenesis have been relatively unexplored.

A second class of 565 genes is highly expressed in ES and EC cell lines but not in seminomas. A number of these genes are among the most differentially expressed in ES and EC cells (Tables 1 and 2). Among the most highly expressed is DNA (cytosine-5-)-methyltransferase 3β (*DNMT3B*), a DNA methyltransferase that functions in early embryogenesis, specifically in the ICM, epiblast, and embryonic ectoderm (45). Also among these is *FOXD3*, a forkhead-family transcription factor that interacts with Oct4, which is essential for the maintenance of mouse primitive ectoderm (10). *FOXD3* is undetectable by PCR in mouse PGCs (46). These two genes cluster close together with other genes, including *SOX2*, which was previously shown to be important in pluripotent cell types of the early embryo (11). The expression of this particular set of genes in ES and EC cells, but not in seminomas, supports the hypothesis that seminomas are closely related to PGCs and that EC cells represent a reversion to a more ICM- or primitive ectoderm-like cell.

We specifically examined genes that might point to signaling pathways that may be important for maintaining or differentiating human ES cells (Table 3). Receptors were identified by using the gene ontology function of the Source database (47). Frizzled 7 (*Fzd7*) was the highest-ranked known receptor in human ES cells and is also highly expressed in EC cells. Frizzled 8 is also highly expressed in ES and EC cells and seminomas. Frizzled receptors are part of the Wnt-β-catenin pathway, and β-catenin signaling has been shown to both delay and modulate mouse ES cell differentiation (48). *Fzd7* clusters with *Tcf3*, a downstream element of β-catenin, with both being expressed in ES and EC cells and YSCs (Fig. 1*D*). SAM also shows that all four fibroblast growth factor receptor genes, *Fgfr1–Fgfr4*, are highly expressed by human ES cells compared with differentiated tissues (Table 3). We have previously shown that in serum-free conditions, fibroblast growth factor signaling promotes long term culture of human ES cells (38). The bone morphogenic protein (BMP) receptor, type 1A, is also highly expressed, and we have recently shown that BMPs induce human ES cells to differentiate to trophoblast (49). Two receptors that are notably absent from this list are the leukemia inhibitory factor (LIF) receptor and its coreceptor, *IL6ST* (*GP130*). LIF sustains undifferentiated proliferation of mouse ES cells in the absence of fibroblasts (14) but does not sustain human ES cells (16). Relative to the control RNA sample, both the LIF receptor and *IL6ST* (*GP130*) are expressed at lower levels in human ES cells, with *IL6ST* significantly less expressed than in differentiated cell types by SAM (Fig. 4*A* and Table 6, which are published as supporting information on the PNAS web site). These results were confirmed by RT-PCR (Fig. 4*B*).

Genes differentially expressed in EC cells compared with ES cells may reflect adaptation of tumor growth by suppression of differentiation. We identified genes with different levels of expression in

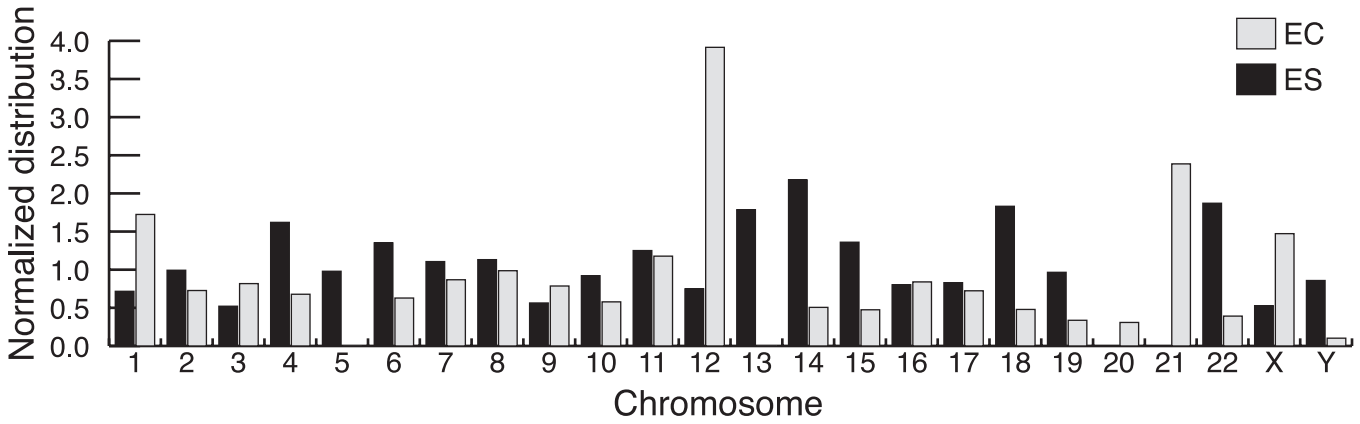

Fig. 2. Chromosomal distribution of genes expressed significantly differently in ES cell lines than in EC cells. Genes expressed significantly higher in ES cell lines (black) relative to EC cell lines and in EC cell lines (gray) relative to ES cell lines were determined by direct comparison using SAM. The chromosomal location of each unique gene was retrieved from April 2003 freeze Golden Path (http://genome.ucsc.edu). The distribution of genes on the chromosomes is expressed as the ratio of the number genes per chromosome in the subset to the number of genes per chromosome in the starting microarray set illustrated in Fig. 1. Chromosome 12 in EC cells is significantly different from expected [hypergeometric test ($P < 0.00001$)].

ES and EC cell lines by direct comparison of the two data sets using SAM (see Table 9, which is published as supporting information on the PNAS web site). The most striking observation from this analysis is that genes from chromosome 12 are greatly overrepresented ($P < 0.00001$ by hypergeometric test) among those significantly more highly expressed in EC cell lines when directly compared with ES cell lines (Fig. 2). From studies of human GCTs themselves, the frequent presence of one or more isochromosomal 12p, or amplification of at least part of chromosome 12p, has previously focused attention on this region of the genome (50). Studies of the early stages of GCT development have led to the suggestion that this amplification has more to do with tumor progression than tumor initiation, and it is generally common to both seminoma and nonseminomatous forms of GCTs (51). A number of transcription factors on chromosome 12 are more highly expressed in EC cells than ES cells, including *FOXM1*, *CREBL2*, and *TEAD4*. Also found among these genes is *WNT5B*, a ligand that binds to frizzled receptors; *DUSP16* (*MPK-7*), a MAPK phosphatase that may act as a negative regulator of a MAPK pathway; and *FKBP4*, which binds to steroid receptor complexes and may play a role in intracellular signaling. A common minimal amplicon (12p11.2–12p1) involved in tumor progression has been suggested (52). We have identified a different region of chromosome 12 (12p12.2–12p13.33) containing 25 of the 36 genes we have identified as differentially expressed on chromosome 12. Interestingly, the gene for Nanog, a protein recently described as being required for maintenance of pluripotency in mouse ES cells (12, 13), is also located in this region of chromosome 12 (12p13.31). Further studies will be necessary to determine whether this region of the chromosome or any of these genes are important for tumor progression.

Table 3. The 20 most positively significant genes in ES cells defined as receptors in the SOURCE database

Clone ID	Gene name	Fold change
239256	*FZD7*, frizzled homolog 7 (*Drosophila*)	18.65
32381	*GABRB3*, GABA A receptor, β3	14.80
309929	*GPR*, putative G protein-coupled receptor	5.75
365147	*ERBB2*, v-erb-b2 erythroblastic leukemia viral oncogene homolog 2	6.45
150361	*IGF1R*, insulin-like growth factor 1 receptor	7.63
154472	*FGFR1*, fibroblast growth factor receptor 1	6.45
305606	*EPHA1*, EphA1	4.98
1323361	*NR2F6*, nuclear receptor subfamily 2, group F, member 6	4.03
137794	*ACVR2B*, activin A receptor, type IIB	3.83
768597	*MUC1*, mucin 1, transmembrane	4.49
1606300	*BMPRIA*, bone morphogenetic protein receptor, type 1A	3.76
785148	*PTPRZ1*, protein tyrosine phosphatase, receptor-type, Z polypeptide 1	3.18
2016775	*GPRC5B*, G protein-coupled receptor, family C, group 5, member B	3.69
1160531	*ERBB3*, v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)	2.42
45231	*GPR19*, G protein-coupled receptor 19	3.02
610883	*PCDHA12*, protocadherin α12	3.32
395955	*GPR27*, G protein-coupled receptor 27	3.74
810459	*FZD8*, frizzled homolog 8 (*Drosophila*)	2.92
752802	*PSK-1*, type I transmembrane receptor (seizure-related protein)	2.91
1678361	*PTPRF*, protein tyrosine phosphatase, receptor type, F	2.75

Taken together, this cDNA microarray analysis is consistent with the hypothesis that human seminomas most closely resemble transformed PGCs, and EC cells most closely resemble transformed ICM or primitive ectoderm cells. The key difference that distinguishes a malignant human teratocarcinoma from a benign teratoma is the persistence of the undifferentiated EC cell population in teratocarcinomas. Human ES cells injected into immunocompromised mice form benign teratomas in which the undifferentiated ES cells do not persist. Because we can now genetically manipulate human ES cells either by overexpressing genes (53, 54) or subtracting function by homologous recombination (55), it will be possible to test the roles of the candidate genes identified by this microarray analysis in converting human ES cells to their malignant EC counterparts.

We thank Martin Pera for the gift of GCT44 and GCT27 cell lines, Christine Pigott and Ping He for technical support, Thomas Zwaka and Lynn Schmidt for critical reading of the manuscript, and Karthik Ramanchandran and members from Stanford Microarray Database and Stanford Functional Genomic Facility for support. This work was supported, in part, by grants to P.O.B. from the National Cancer Institute, to P.W.A. from the Wellcome Trust and Yorkshire Cancer Research, to J.S.D. and P.W.A. from the Biotechnology and Biological Sciences Research Council, and to J.D.B. from the Lance Armstrong Foundation. X.C. was a Howard Hughes Medical Institute Postdoctoral Fellow of the Life Sciences Research Foundation. P.O.B. is an investigator of the Howard Hughes Medical Institute.

1. Evans, M. J. & Kaufman, M. H. (1981) *Nature* **292,** 154–156.
2. Martin, G. R. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 7634–7638.
3. Scholer, H. R., Balling, R., Hatzopoulos, A. K., Suzuki, N. & Gruss, P. (1989) *EMBO J.* **8,** 2551–2557.
4. Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H. & Smith, A. (1998) *Cell* **95,** 379–391.
5. Boiani, M., Eckardt, S., Scholer, H. R. & McLaughlin, K. J. (2002) *Genes Dev.* **16,** 1209–1219.
6. Wilder, P. J., Kelly, D., Brigman, K., Peterson, C. L., Nowling, T., Gao, Q. S., McComb, R. D., Capecchi, M. R. & Rizzino, A. (1997) *Dev. Biol.* **192,** 614–629.
7. Abud, H. E., Skinner, J. A., McDonald, F. J., Bedford, M. T., Lonai, P. & Heath, J. K. (1996) *Dev. Genet.* **19,** 51–65.
8. Yuan, H., Corbi, N., Basilico, C. & Dailey, L. (1995) *Genes Dev.* **9,** 2635–2645.
9. Sutton, J., Costa, R., Klug, M., Field, L., Xu, D., Largaespada, D. A., Fletcher, C. F., Jenkins, N. A., Copeland, N. G., Klemsz, M. & Hromas, R. (1996) *J. Biol. Chem.* **271,** 23126–23133.
10. Hanna, L. A., Foreman, R. K., Tarasenko, I. A., Kessler, D. S. & Labosky, P. A. (2002) *Genes Dev.* **16,** 2650–2661.
11. Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N. & Lovell-Badge, R. (2003) *Genes Dev.* **17,** 126–140.
12. Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M. & Yamanaka, S. (2003) *Cell* **113,** 631–642.
13. Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S. & Smith, A. (2003) *Cell* **113,** 643–655.
14. Smith, A. G., Nichols, J., Robertson, M. & Rathjen, P. D. (1992) *Dev. Biol.* **151,** 339–351.
15. Niwa, H., Burdon, T., Chambers, I. & Smith, A. (1998) *Genes Dev.* **12,** 2048–2060.
16. Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S. & Jones, J. M. (1998) *Science* **282,** 1145–1147.
17. Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R. C. & Melton, D. A. (2002) *Science* **298,** 597–600.
18. Ivanova, N. B., Dimos, J. T., Schaniel, C., Hackney, J. A., Moore, K. A. & Lemischka, I. R. (2002) *Science* **298,** 601–604.
19. Sato, N., Sanjuan, I. M., Heke, M., Uchida, M., Naef, F. & Brivanlou, A. H. (2003) *Dev. Biol.* **260,** 404–413.
20. Stevens, L. C. (1967) *J. Natl. Cancer Inst.* **38,** 549–552.
21. Stevens, L. C. (1967) *Adv. Morphog.* **6,** 1–31.
22. Matsui, Y., Zsebo, K. & Hogan, B. L. (1992) *Cell* **70,** 841–847.
23. Andrews, P. W. (1988) *Biochim. Biophys. Acta* **948,** 17–36.
24. Damjanov, I. (1993) *Eur. Urol.* **23,** 5.
25. Skakkebaek, N. E. (1972) *Lancet* **2,** 516–517.
26. Pierce, G. B. (1974) *Am. J. Pathol.* **77,** 103–118.
27. Henderson, J. K., Draper, J. S., Baillie, H. S., Fishel, S., Thomson, J. A., Moore, H. & Andrews, P. W. (2002) *Stem Cells* **20,** 329–337.
28. Andrews, P. W. & Damjanov, I. (1994) in *Atlas of Human Tumor Cell Lines*, eds. Hay, R. J., Park, J.-G. & Gazdar, A. (Academic, London), pp. 443–476.
29. Andrews, P. W., Damjanov, I., Simon, D., Banting, G., Carlin, C., Dracopoli, N. C. & Fogh, J. (1984) *Lab. Invest.* **50,** 147–162.
30. Pera, M. F., Blasco Lafita, M. J. & Mills, J. (1987) *Int. J. Cancer* **40,** 334–343.
31. Pera, M. F., Cooper, S., Mills, J. & Parrington, J. M. (1989) *Differentiation* **42,** 10–23.
32. Bronson, D. L., Andrews, P. W., Vessella, R. L. & Fralley, E. E. (1983) in *Cold Spring Harbor Conferences on Cell Proliferation*, eds. Silver, L. M., Martin, G. R. & Strickland, S. (Cold Spring Harbor Lab. Press, Plainview, NY), Vol. 10, pp. 597–605.
33. Teshima, S., Shimosato, Y., Hirohashi, S., Tome, Y., Hayashi, I., Kanazawa, H. & Kakizoe, T. (1988) *Lab. Invest.* **59,** 328–336.
34. Andrews, P. W., Bronson, D. L., Benham, F., Strickland, S. & Knowles, B. B. (1980) *Int. J. Cancer* **26,** 269–280.
35. Andrews, P. W., Goodfellow, P. N., Shevinsky, L., Bronson, D. L. & Knowles, B. B. (1982) *Int. J. Cancer* **29,** 523–531.
36. Bronson, D. L., Andrews, P. W., Solter, D., Cervenka, J., Lange, P. H. & Fraley, E. E. (1980) *Cancer Res.* **40,** 2500–2506.
37. Vogelzang, N. J., Bronson, D. L., Savino, D., Vessella, R. L. & Fraley, E. E. (1985) *Cancer* **55,** 2584–2593.
38. Amit, M., Carpenter, M. K., Inokuma, M. S., Chiu, C. P., Harris, C. P., Waknitz, M. A., Itskovitz-Eldor, J. & Thomson, J. A. (2000) *Dev. Biol.* **227,** 271–278.
39. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., *et al.* (2000) *Nature* **406,** 747–752.
40. Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., *et al.* (2000) *Nature* **403,** 503–511.
41. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 14863–14868.
42. Tusher, V. G., Tibshirani, R. & Chu, G. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 5116–5121.
43. Pesce, M., Gross, M. K. & Scholer, H. R. (1998) *BioEssays* **20,** 722–732.
44. Niwa, H., Miyazaki, J. & Smith, A. G. (2000) *Nat. Genet.* **24,** 372–376.
45. Watanabe, D., Suetake, I., Tada, T. & Tajima, S. (2002) *Mech. Dev.* **118,** 187–190.
46. Saitou, M., Barton, S. C. & Surani, M. A. (2002) *Nature* **418,** 293–300.
47. Diehn, M., Sherlock, G., Binkley, G., Jin, H., Matese, J. C., Hernandez-Boussard, T., Rees, C. A., Cherry, J. M., Botstein, D., Brown, P. O. & Alizadeh, A. A. (2003) *Nucleic Acids Res.* **31,** 219–223.
48. Kielman, M. F., Rindapaa, M., Gaspar, C., van Poppel, N., Breukel, C., van Leeuwen, S., Taketo, M. M., Roberts, S., Smits, R. & Fodde, R. (2002) *Nat. Genet.* **32,** 594–605.
49. Xu, R. H., Chen, X., Li, D. S., Li, R., Addicks, G. C., Glennon, C., Zwaka, T. P. & Thomson, J. A. (2002) *Nat. Biotechnol.* **20,** 1261–1264.
50. Atkin, N. B. & Baker, M. C. (1982) *Lancet* **2,** 1349.
51. Rosenberg, C., Van Gurp, R. J., Geelen, E., Oosterhuis, J. W. & Looijenga, L. H. (2000) *Oncogene* **19,** 5858–5862.
52. Rodriguez, S., Jafer, O., Goker, H., Summersgill, B. M., Zafarana, G., Gillis, A. J., van Gurp, R. J., Oosterhuis, J. W., Lu, Y. J., Huddart, R., *et al.* (2003) *Oncogene* **22,** 1880–1891.
53. Eiges, R., Schuldiner, M., Drukker, M., Yanuka, O., Itskovitz-Eldor, J. & Benvenisty, N. (2001) *Curr. Biol.* **11,** 514–518.
54. Ma, Y., Ramezani, A., Lewis, R., Hawley, R. G. & Thomson, J. A. (2003) *Stem Cells* **21,** 111–117.
55. Zwaka, T. P. & Thomson, J. A. (2003) *Nat. Biotechnol.* **21,** 319–321.