Issues in Development

Karyotypic Stability, Genotyping, Differentiation, Feeder-Free Maintenance, and Gene Expression Sampling in Three Human Embryonic Stem Cell Lines Derived Prior to August 9, 2001

SANDII N. BRIMBLE,^{1,5} XIANMIN ZENG,^{2,5} DEBORAH A. WEILER,¹ YONGQUAN LUO,³
YING LIU,³ IAN G. LYONS,^{1,4} WILLIAM J. FREED,² ALLAN J. ROBINS,¹
MAHENDRA S. RAO,³ and THOMAS C. SCHULZ¹

ABSTRACT

The number of human embryonic stem cell (hESC) lines available to federally funded U.S. researchers is currently limited. Thus, determining their basic characteristics and disseminating these lines is important. In this report, we recovered and expanded the earliest available cryopreserved stocks of the BG01, BG02, and BG03 hESC lines. These cultures exhibited multiple definitive characteristics of undifferentiated cells, including long-term self-renewal, expression of markers of pluripotency, maintenance of a normal karyotype, and differentiation to mesoderm, endoderm, and ectoderm. Each cell line exhibited a unique genotype and human leukocyte antigen (HLA) isotype, confirming that they were isolated independently. BG01, BG02, and BG03 maintained in feeder-free conditions demonstrated self-renewal, maintenance of normal karyotype, and gene expression indicative of undifferentiated pluripotent stem cells. A survey of gene expression in BG02 cells using massively parallel signature sequencing generated a digital read-out of transcript abundance and showed that this line was similar to other hESC lines. BG01, BG02, and BG03 hESCs are therefore independent, undifferentiated, and pluripotent lines that can be maintained without accumulation of karyotypic abnormalities.

INTRODUCTION

HUMAN DEVELOPMENT from the fertilized oocyte follows a defined sequence of events, and the stages

prior to implantation can be successfully replicated in culture. The ability to culture fertilized eggs until blastocyst hatching has allowed in vitro fertilization (IVF) centers to manipulate the developing egg to assist re-

¹BresaGen, Inc., Athens, GA 30605.

²Cellular Neurobiology Branch, National Institute on Drug Abuse, Department of Health and Human Services, Baltimore, MD 21224.

³Laboratory of Neuroscience, National Institute of Aging, Department of Health and Human Services, Baltimore, MD 21224.

⁴Present address: Gibco/Invitrogen, Grand Island, NY 14072.

⁵These authors contributed equally.

production where conventional strategies have failed. More recently, blastocysts collected for IVF purposes, but unneeded because of successful pregnancy (1,2) or discarded due to poor morphological grading (3), have been used to derive human embryonic stem cell (hESC) lines. Seventy-eight cell lines were reported isolated prior to the August 9, 2001, deadline established for lines eligible for federal research funding; however, at least an additional hundred lines have subsequently been derived. The actual number of hESC lines available for distribution to federally funded researchers is currently only 22 (<http://stemcells.nih.gov/index.asp>), because other isolates failed to recover from cryopreservation or are not available for other reasons. If the current guidelines are not changed, it is imperative that the lines that qualify for federal funding are characterized thoroughly and made available for dissemination to the research community.

Current reports analyzing the basic cell properties of this limited number of lines have suggested that hESC are similar in their expression of cell-surface antigens and markers characteristic of the hESC state. While not all markers have been tested in all lines, expression of OCT4 (POU5F1), NANOG, SOX-2, ABCG2, CRIPTO, UTF1, connexin, gap junction proteins, and the cell-surface antigens recognized by the SSEA-3, SSEA-4, TRA-1-60, and TRA-1-81 antibodies is common, as is the ability to self-renew and differentiate into ectoderm, endoderm, and mesoderm in vitro and in vivo (reviewed in ref. 4). Microarray analyses of multiple cell lines have highlighted a wider set of transcripts that may mark or function in the pluripotent state (5).

Recent reports have suggested, however, that differences between lines may exist based on the method of passaging, timing of isolation, and allelic differences. For example, Bongso and colleagues have noted differences in the expression of *REX1* and several other genes by SAGE analysis (6) and expression of nestin has been detected in some lines but not others by reverse transcriptase (RT)-PCR (5,7,8). Others have reported differences in karyotype stability (9,10). Although the mechanism of the origin of hESCs containing triploidies of chromosomes 17 and/or 12 has not been determined, it is possible that the disaggregation of hESC colonies to single cells during passaging leads to selective advantage for aneuploid cells. It will be important to compare the characteristics of the available lines under standardized culture and differentiation conditions. However, information describing the basic characteristics of many of these lines is incomplete.

The BG01, BG02, and BG03 hESC lines were derived from discarded IVF-derived embryos that had not been given a morphological grading that warranted cryopreservation or uterine transfer (3). Isolated inner cell

masses (ICM) were expanded on layers of mouse embryonic fibroblasts (MEF) using microdissection passaging, and stocks were cryopreserved between passages 4–7. Initial characterization of these lines demonstrated expression of the markers alkaline phosphatase, SSEA-3, SSEA-4, TRA-1-60, TRA-1-81, and OCT4. The BG01 and BG02 lines were successfully thawed and cultured for over 80 and 30 passages, respectively, and retained a normal 46, XY karyotype after 6 months in culture. Differentiation experiments demonstrated the capacity to generate human chorionic gonadotropin (hCG)-producing trophoblastic cells and cardiomyocytes (3), representatives of multiple differentiated lineages (11), and directed differentiation to neuronal lineages (12–14). A detailed microarray analysis of the BG01 and BG02 cell lines demonstrated that their gene expression profile was similar to other hESC lines, and contributed to the identification of ~92 transcripts that mark the pluripotent state (5,11).

The present report details further characterization of these lines, including the first examination of the properties of BG03 cells following recovery and expansion to generate a viable cell line. The properties of these three lines were similar to each other as well as to other reported hESC lines.

MATERIALS AND METHODS

hESC culture

The BG01, BG02, and BG03 hESC lines (3) used in this work are registered with the National Institutes of Health (NIH) (<http://stemcells.nih.gov/index.asp>). hESCs were primarily cultured on a layer of low- or high-density MEF, in 20% knockout serum replacer (KSR) hESC medium that had been conditioned on MEFs for 24 h as described (12,15). Cultures were expanded by passaging clumps of ~10–100 cells as described (14). hESCs were also maintained on Matrigel (BD Biosciences, Bedford, MA, <http://www.bdbiosciences.com>) or fibronectin (Sigma, St. Louis, MO, <http://www.sigma-aldrich.com>)-coated plates in MEF conditioned medium (CM) essentially as described (15), except that EDTA-free trypsin was used to passage the cells (14). The passage 24 BG03 culture used herein was cultured for six passages on Matrigel, followed by passaging on fibronectin. The BG01, BG02, and BG03 hESC lines were negative for hepatitis B and C, human immunodeficiency virus-1 (HIV-1), and -2, human T cell leukemia virus (HTLV I/II), (HSV-1) and -2, cytomegalovirus (CMV), and mycoplasma. A detailed manual of the hESC protocols described here is available online at the NIH

stem cell web site (<http://stemcells.nih.gov/research/protocols.asp>). *RT-PCR*

Immunofluorescence, alkaline phosphatase staining, and differentiation of hESCs as embryoid bodies

hESCs and differentiated cultures were fixed in 4% paraformaldehyde and 4% sucrose in 1× phosphate-buffered saline (PBS) and immunostained as described previously (14). Alkaline phosphatase activity was detected using the Vector Red substrate kit (Vector Laboratories, Burlingame, CA, <http://www.vectorlabs.com>). hESC were differentiated as embryoid bodies. Whole hESC colonies were lifted from the MEF layer and placed in suspension culture in Dulbecco's modified Eagle medium (DMEM)/F12 with 15% fetal cell serum (FCS), 5% KSR, 20 mM L-glutamine, 0.5 U/ml penicillin, 0.5 U/ml streptomycin, 0.1 mM β -mercaptoethanol, and 1× nonessential amino acids. Suspension aggregates were differentiated for 8–9 days, and were then plated to adherent culture and differentiated for a further 8–10 days in the same medium.

DNase I-treated RNA and first-strand cDNA was prepared as described previously (14) and amplified using Ready Mix Taq DNA polymerase (Sigma). Primers were as used previously (11) except for (forward and reverse, respectively): GAPDH 5'-TGAAGGTCGGAGTCAACGGATTTGGT-3' and 5'-CATGTGGGCCATGAGGTCCACCAC-3'; NANOG, 5'-GACTGAGCTGGTTGCCTCAT-3' and 5'-TTTCTTCAGGCCACAAATC-3'; FOXD3, 5'-CGTCGCTCATCAAGTCCGA-3' and 5'-AGAAACTGTCCGAGAGTGG-3'; TERT, 5'-AGCTATGCCCGGACCTCCAT-3' and 5'-GCCTGCAGCAGGAGGATCTT-3'; SOX-1, 5'-TACAGCCCC ATCTCCAATC-3' and 5'-GCTCCGACTTCACCAGAGAG-3'; AFP 5'-CCGAACCTTCCAAGCCATAA-3' and 5'-TGGCATTCAAGAGGGTTTTC-3'; MSX1 5'-CCTTCCCTTTAACCCTCACAC-3' and 5'-CCGATTCTCTGCGCTTTTC-3'; HAND1 5'-GCCTAGCCACCA-GCTACATC-3' and 5'-CGGCTCACTGGTTTAACTCC-3'; NEUROFILAMENT HEAVY CHAIN 5'-TGAACACAGACGCTATGCGCTCAG-3' and 5'-CACCTTTATGTGAGTGGACACAGAG-3'; β III TUBULIN,

TABLE 1. SUMMARY OF CYTOGENETIC ANALYSES OF hESC CULTURES

Method	Line	Passage	Karyotype	Count
Trypsin/EDTA	BG01	p32 ^a	46,XY	1
			47,XY,+17	12
			48,XY,+12,+17	3
			49,XY,+1,+12,+17	4 (20 total)
		p42 ^a	46,XY	4
			47,XY,+17	26 (30 total)
		p84 ^a	50,XXY,+12,+14,+17	40
			51,XXY,+7,+12,+14,+17	2
			51,XXY,+8,+12,+14,+17	2 (44 total)
Microdissection	BG01	p35	46,XY	20/20
		p52	46,XY	17/17
	BG02	p19	46,XY	10/10
		p19	46,XY	15/15
		p28	46,XY	15/15
		p32	46,XY	14/14
	BG03	p17	46,XX; (47,XX,+11 ^b)	29; (1 ^b)/30
		p30	46,XX	20/20
		p31	46,XX	20/20
Matrigel and MEF-CM	BG01	p32 M5 ^c	46,XY	20/20
	BG02	p32 M4 ^c	46,XY	20/20
	BG03	p23 M5 ^c	46,XX	20/20
		p34 M16 ^{c,d}	46,XX	20/20

^aTotal passages using collagenase and trypsin/EDTA.

^bRare triploidy that had not accumulated in the culture.

^cTotal number of manual or EDTA-free trypsin passages; M, passages on Matrigel or fibronectin in conditioned medium.

^dSix initial passages on Matrigel, and 10 subsequent passages on fibronectin.

5'-GCCATCTTCCCGCGCAACCG-3' and 5'-CTTGGGGCCCTGGGACTCCG-3'; GATA4, 5'-CTCCTTCA-GGCAGTGAGAGC-3' and 5'-GAGATGCAGTGTGC-TCGTGC-3'. End products were identified with ethidium bromide staining of agarose gels.

Karyotyping, genotyping, and HLA typing

Karyotyping was performed using colcemid arrest and G banding. Genotyping was performed (Identigene, Houston, <http://www.identigene.com>) using 10 short tandem repeat (STR) loci (vWA, FGA, D8S1179, D2S11, D18S51, D16S539, TH01, D2S1338, D19S433). High-resolution typing of HLA alleles was performed using local procedures at the Immunogenetics Laboratory of John's Hopkins University, which were adapted from the 13th International Histocompatibility Workshop (<http://www.ihwg.org/>) (16). A combination of sequence specific priming (SSP) and sequence based typing (SBT) was used. Details of the reagents are available from this journal upon request.

Focused cDNA arrays

Focused arrays for stem cell markers (Hs601), common cytokines (Hs3), and TGF/BMP signaling (Hs023) were assayed with probes generated from 4 μ g of BG03 total RNA as described previously (11,14,17).

Massively parallel signature sequencing

A digital read out of global gene expression in BG02 hESCs was generated using massively parallel signature sequencing (MPSS) (Lynx Therapeutics Inc. Hayward,

CA) as described previously (18–20). Briefly, poly(A)⁺ RNA was used to generate a library of microbeads, each bearing a unique cDNA fragment. The microbead library was loaded onto a flow cell and sequenced. Signatures were derived in two overlapping frames to avoid potential loss of signatures due to palindromes. Seventeen-nucleotide data were collected for each sequencing frame and pooled to generate the 17-nucleotide signature data set. An additional round of sequencing of these reactions, and pooling, generated the 20-nucleotide data set. Multiple signatures representing a single Unigene cluster were pooled and 17- and 20-nucleotide data sets were organized under transcripts per million (tpm) per individual Unigene cluster (Unigene build Hs 169).

RESULTS

Passaging methods and accumulation of aneuploidies in hESC cultures

The derivation of the four hESC lines examined here has been reported previously (3). BG01 and BG02 were recovered from cryopreservation and expanded for in vitro characterization and retained a normal karyotype after 6 months of continual culture using microdissection passaging (3). To enable larger-scale culture of these lines, BG01 cells were passaged with an EDTA-based cell dispersal buffer (CDB) (11) or with collagenase and trypsin/EDTA (14). However, as reported by others (9), accumulation of karyotypic abnormalities was observed

TABLE 2. STR GENOTYPING AND HLA CLASS I AND CLASS II TYPING OF BG01, BG02, AND BG03 LINES

	<i>Genotyping^a</i>									
	<i>D3S1358</i>	<i>vWA</i>	<i>FGA</i>	<i>D8S1179</i>	<i>D2S11</i>	<i>D18S51</i>	<i>D16S539</i>	<i>TH01</i>	<i>D2S1338</i>	<i>D19S433</i>
BG01	15,17	16,17	22	10,12	28,29	19	9,11	7,9,3	17,24	13,15
BG02	15,17	18,19	20,25	13,14	27,30	14,16	12	8,9	17,24	13,14.2
BG03	14,15	16	20,22	12,13	29	17,19	11,13	6,7	21,23	13,14
	<i>HLA isotyping</i>									
	<i>A</i>	<i>Cw</i>	<i>B</i>	<i>Bw</i>	<i>DRB1</i>	<i>DRB3</i>	<i>DRB4</i>	<i>DRB5</i>	<i>DQB1</i>	
BG01	0201	0401	3503	6	301	xx			0201	
	0301	0701	4901	4	405		xx		0302	
BG02	2402	0304	3503	6	1502			xx	0601	
	6801	0401	3511	6	1602				0301	
BG03	0201	05AC	4402	4	0401		xx		0301	
	3201								0302	

^aNumber of repeats of STR for each loci.

FIG. 1. Alkaline phosphatase activity and expression of markers of pluripotency in hESC cultures. **(A)** Early passage cultures of BG01, BG02, and BG03 expressed alkaline phosphatase activity and the markers of pluripotency OCT4, SSEA-3, SSEA-4, TRA-1-60, TRA-1-81, but were negative for SSEA-1 as expected. Nuclei were stained with DAPI. **(B)** RT-PCR analysis of early passage cultures of BG01, BG02, and BG03 demonstrated expression of OCT4, NANOG, REX1, SOX2, UTF1, ABCG2, CRIPTO, FOXD3, TERT, and DPPA5. (+) Reverse transcriptase⁺ experimental cDNA template; (–) reverse transcriptase minus mock cDNA, negative control template.

with long-term culture, when colonies were disaggregated to essentially single cells at passaging (Table 1, representative data). Trisomies of chromosomes 12 and 17 as well as the X chromosome were primarily observed, with occasional trisomy of chromosomes 1, 7, 8, and 14. The timing of the first appearance of these trisomies was not determined, but a significant proportion of cells were aneuploid within ~32 passages of switching to collagenase and trypsin/EDTA passaging, and a karyotype of 49, XXY, +12, +17 was predominant in a culture after 45

passages using CDB (X.Z., submitted). There may be some variability in the kinetics of this accumulation, because other collagenase and trypsin/EDTA-passaged cultures showed only ~50% of cells with an abnormal chromosome count after 33 passages (14). Despite the accumulation of trisomies, obvious differences in the differentiation capacity of normal and karyotypically abnormal cells were not observed (14) (and X.Z., submitted), implying that these variant cells may still be a useful experimental resource.

TABLE 3. SUMMARY OF EXPRESSION OF MARKERS ASSOCIATED WITH PLURIPOTENCY IN BG01, BG02, AND BG03 hESCs

Marker	BG01 RT-PCR	BG01 staining	BG02 RT-PCR	BG02 staining	BG03 RT-PCR	BG03 staining
AP		+		+		+
SSEA-1		—		—		—
SSEA-3		+		+		+
SSEA-4		+		+		+
TRA-1-60		+		+		+
TRA-1-81		+		+		+
TERT	+		+		+	
OCT4	+	+	+	+	+	+
NANOG	+		+		+	
SOX-2	+		+		+	
REX-1	+		+		+	
UTF1	+		+		+	
DPPA5	+		+		+	
FOXD3	+		+		+	
CRIPTO	+		+		+	
CX43	+		+		+	
CX45	+		+		+	
ABCG2	+		+		+	

Long-term maintenance of karyotypically normal hESC cultures

To prepare karyotypically normal cultures of these hESC lines that could be used for additional characterization, the earliest available stocks of BG01, BG02, and BG03 cells were recovered at passages 7, 6, and 4, respectively. These cultures were expanded and cryopreserved again between passages ~9–15 to ensure long-term availability of these cell lines. The BG04 cell line was not successfully recovered from cryopreservation and has been removed from the NIH registry. Expansion from these secondary stocks was used to derive cultures for the experiments described here, as well as to establish stocks from passages ~15–30, which are available to the research community. Cultures were expanded by passaging clumps of cells, primarily by microdissection

passaging, but also by using EDTA-free trypsin. These cultures retained a normal karyotype of 46,XY for BG01 and BG02 and 46,XX for BG03 for at least 6 months continuous culture, or more than ~30–50 passages (Table 1). This data confirmed that these hESC lines retained a stable karyotype when disaggregation to single cells during passaging was avoided.

Genotyping and HLA isotyping of BG01, BG02, and BG03

Information demonstrating genetic uniqueness and also enabling accurate identification is not available for most hESC lines. Genotyping was therefore performed on BG01, BG02, and BG03 cells using STR markers (Table 2) and demonstrated that these lines were independent and provides a reference for their identification.

FIG. 2. Differentiation of hESCs to endoderm, mesoderm, and ectoderm in vitro. (A) Immunofluorescence of BG01, BG02, and BG03 cells differentiated as embryoid bodies demonstrated the generation of representatives of all three germ layers. Cells expressing markers of endoderm (α -fetoprotein, AFP), mesoderm (smooth muscle actin, SMA), or ectoderm (β III Tubulin) were detected. Nuclei were stained with DAPI. (B) RT-PCR analysis of the same populations. Expression of the markers OCT4 and NANOG (pluripotency) was detected in undifferentiated cells, and GATA4, AFP (endoderm), MSX1, and HAND1 (mesoderm) were all expressed at a higher level in the embryoid bodies. Neurofilament heavy chain (NF heavy) and β III tubulin (ectodermal markers) were expressed in both populations, but expression of SOX1 confirmed the generation of ectodermal lineages in embryoid bodies. (*) BG01 hESC sample in the SOX1 RT-PCR.

FIG. 3. Pluripotent marker expression in BG03 cells maintained in MEF-CM on matrigel and fibronectin. BG03 hESCs were maintained in MEF-CM on matrigel for 6 passages and subsequently on fibronectin for a total of more than 24 passages without a MEF feeder layer. Cells were passaged using EDTA-free trypsin, and immunostained at passage 7 for OCT4, SSEA-4, TRA-1-60, and TRA-1-81. Nuclei were stained with DAPI. Images were captured with a 5 $\times$ objective.

A

B

FIG. 2.

FIG. 3.

Furthermore, these lines also exhibited unique HLA isotypes (Table 2). This provides an additional reference for identifying these cells and a guide for HLA-matching criteria in any potential transplantation setting for these hESC lines or their differentiated derivatives.

Gene expression in undifferentiated hESC cultures and during differentiation in vitro

Immunofluorescence was performed to analyze the expression of markers of pluripotency in undifferentiated, early-passage BG01, BG02, and BG03 cell lines. All three cell lines expressed alkaline phosphatase activity, and were positive for OCT4, SSEA-3, SSEA-4, TRA-1-60, and TRA-1-81 and negative for SSEA-1 (Fig. 1A). RT-PCR analysis also demonstrated expression of OCT4, NANOG, REX1, SOX2, UTF1, ABCG2, CRIPTO, FOXD3, TERT, and DPPA5 (Fig. 1B). This analysis demonstrated the expression of cell-surface antigens and transcriptional markers of pluripotency expected for undifferentiated hESCs. A summary of the markers expressed by hESCs in these analyses is shown in Table 3.

To investigate the differentiation potential of these cells, early-passage hESC cultures were differentiated as embryoid bodies. Cultures were immunostained for either α -fetoprotein (endoderm), smooth muscle actin (mesoderm), or β III tubulin (ectoderm), and positive cells of each of these lineages were detected (Fig. 2A). Gene expression in these populations was also examined using RT-PCR (Fig. 2B). This demonstrated that expression of the pluripotent markers OCT4 and NANOG was down-regulated as hESCs differentiated in embryoid bodies. At the same time, the expression of markers of endoderm (GATA4 and AFP) and mesoderm (MSX1 and HAND1) was up-regulated in embryoid bodies, demonstrating the generation of differentiated cells from these lineages. Neurofilament heavy-chain and β III tubulin, markers of ectoderm, were expressed in both the hESCs and the embryoid bodies. Because the low level of OCT4 and NANOG expression in the embryoid bodies indicated a low proportion of undifferentiated cells in these populations, the expression of neurofilament and β III tubulin suggested the presence of ectodermal derivatives in embryoid bodies. Expression of SOX1, a marker of neural ectoderm, was detected in the embryoid body populations but not in a control BG01 hESC culture, confirming ectodermal differentiation. These analyses indicated that BG01, BG02, and BG03 cells could differentiate to representatives of all three germ layers.

Maintenance of BG01, BG02, and BG03 in feeder-free conditions

The maintenance of hESC cultures in the absence of feeder layers offers some clear advantages for certain experiments. Long-term culture of hESCs on the Matrigel

substrate in MEF-CM has been described previously (15,21,22). BG01, BG02, and BG03 hESCs were maintained on Matrigel or fibronectin using MEF-CM and passaging with EDTA-free trypsin. These cultures exhibited normal karyotypes (Table 1) and the expected morphology of monolayer growth and prominent nucleoli of hESCs that are cultured under these conditions (not shown). In particular, BG03 hESCs were maintained for more than 24 passages under these conditions, retained expression of OCT4, SSEA4, TRA-1-60, and TRA-1-81 after seven passages (Fig. 3) and were karyotypically normal after 16 passages (Table 1). This demonstrated that these hESC lines could be adapted to culture in the absence of a MEF feeder layer.

Focused array analysis of gene expression in the BG03 hESC line

To begin assembling a comprehensive gene expression data set for the BG03 cell line, focused cDNA arrays were used to examine expression of 454 stem cell markers, cytokines, and members of the TGF/BMP signaling pathway (Fig. 4). This data was generated from BG03 cells that had been maintained with MEF-CM on Matrigel or fibronectin for 15 passages with EDTA-free trypsin. This provides an important validation of the quality of cultures grown under these conditions by enabling comparison to previous focused array analyses of hESC lines maintained on MEFs (11,13,14). Overall, the profile of gene expression in this culture (Fig. 4) was highly similar to that observed previously, including, for example, the expression of the pluripotent markers POU5F1 (OCT4), CD9, SOX2, ZFP42, and EGR2. However, the expression of SOX1 was also detected, which may indicate some sporadic differentiation to early neural lineages under these conditions. A total of 34 cytokines and transforming growth factor (TGF)/bone morphogenic protein (BMP) pathway members were detected including FGF2, FGF11, IL10, TGFB1, CER1, GDF3, ID2,3 and 4. FGF2 and TGF- β have been demonstrated to maintain hESCs in feeder-free conditions (23), and it is also possible that autocrine activity of some of these cytokines may contribute to the maintenance of pluripotency. These focused array experiments demonstrated that the BG03 hESC line, which was maintained without contact with feeder layers for more than 2 months, maintained a gene expression profile highly similar to other hESCs.

Global survey of gene expression in BG02 hESCs using MPSS

A detailed expression analyses of the BG02 hESC line was performed using MPSS (19,20). MPSS generates a digital read out of gene expression, with abundance indicated as tpm, and, unlike array hybridization ap-

Marker expression in BG03 hESC

Stem cell array (Hs601): ACTA2, ACTG2, BMPR1A, BMPR2, C3orf4, CTNNA1, CTNNAL1, CTNNB1, CCNE1, CD24, CD9, CDH1, CDH2, CDH3, CDK4, CER1, CXCR4, DNMT3B, EGR2, FGF17, FGF8, FGF9, FOXM1, FZD3, FZD7, GDF3, GJA7, HSPA9B, IGF1R, IGF2R, INS, ITGA5, ITGA6, ITGAV, ITGB1, ITGB5, KDR, KRT14, KRT8, MAP1B, MDM2, MYH11, NES, NODAL, LOC145957, NUMB, PDGFA, POU5F1, PROM1, PTEN, PUM1, PUM2, SIAT8A, SLC2A1, SNAI1, SOX1, SOX2, SOX3, SOX4, TERF1, TGFB1, TGFB1, THY1, TP53, VEGF, VIM, ZFP42.

Common Cytokine array (Hs003): BMP2, CSF1, FGF11, FGF17, FGF19, FGF2, IL10, IL11, DKFZp451J0118, IL15, IL19, PDGFA, PTN, TGFB1, TNFSF14, VEGF.

TGF/BMP signaling array (Hs023): BMPR1A, CDC25A, CER1, COL1A2, EBAF, FST, GDF3, ID2, ID3, ID4, INHA, ITGB5, JUNB, LEFTB, MADH6, NODAL, SOX4, TDGF1.

FIG. 4. Focused arrays of BG03 gene expression. The gene expression of BG03 cells that had been maintained in MEF-CM on matrigel and fibronectin was examined by focused array analysis of stem cell markers, common cytokines, and the TGF/BMP signaling pathways. BG03 gene expression was highly similar to previous analyses of hESCs using this focused stem cell cDNA array, and expression of 34 cytokines and members of the TGF/BMP signaling pathways were detected. Transcripts detected by each array are listed.

proaches, MPSS is not limited by a predetermined probe set. This analysis therefore complements the microarray analyses of BG01 and BG02 performed previously (5,11). Gene expression in BG02 cells was assayed by collecting 17-, and with a subsequent sequencing round, 20-nucleotide signatures in two sequencing frames. The 20-nucleotide signature set enables more accurate identification of UniGene clusters, but some signatures are lost with the additional sequencing cycle. Signatures representing a single UniGene cluster were pooled and the expression level shown in tpm. A summary of transcriptional abundance in these surveys is shown in Table 4 and a list of the most highly expressed transcripts in Table 5. MPSS detected expression of several of the markers of pluripotency used in the above expression analyses (Table 5), and showed a good correlation of expression

with a larger panel of pluripotency markers identified by a microarray comparison of multiple hESC lines (5). A total of 78 of these 92 (84.7%) hESC markers were detected in BG02 cells by MPSS (Supplementary Table 1),* indicating that this subset of transcripts can also be used to define the pluripotent state of these cells. Furthermore, numerous markers of differentiated lineages, such as α -fetoprotein, Brachyury, FGF5, MSX1, SOX1, and KERATIN 15 were not detected (Table 5), confirming the undifferentiated state of these cells. The data sets for the 17-nucleotide and 20-nucleotide MPSS surveys of BG02 gene expression are available in Supplementary Tables 2 and 3,* respectively.

*Available upon request from the Publisher at stemcells@liebertpub.com

TABLE 4. SUMMARY OF MPSS SURVEYS OF BG02 GENE EXPRESSION

<i>tpm</i> ^a	20 nt ^b	%	17 nt ^b	%
>10,000	6	0.09	2	0.02
>5,000	17	0.26	15	0.18
>1,000	131	1.99	119	1.42
>500	186	2.84	152	1.80
>100	1447	22.08	1384	16.51
>50	1224	18.67	1346	16.06
>10	2699	41.20	3534	42.17
Total genes ^c	6,553		8,381	
Total signatures	1,063,090		961,601	

^atpm, Transcripts per million. Organized into clusters of expression level.

^b20 nt and 17 nt, 20- and 17-nucleotide signature MPSS reads.

^cIdentified as Unigene clusters. Some Unigene clusters are represented by multiple types of signatures.

DISCUSSION

The limitation on the number of pluripotent hESC lines that can be used for federally funded U.S. research means that the available lines must be characterized carefully and disseminated. Early-passage stocks of the BG01, BG02, and BG03 lines (3) were successfully recovered, expanded, and characterized as described here. This represents the first expansion of BG03 cells, and these three lines exhibited similar self-renewal, gene expression, and pluripotency and were also similar to other characterized hESC lines. Failure to generate viable lines has occurred with a large proportion of qualifying hESC derivations, including BG04. Only 22 of the predicted 78 hESC lines are currently available for distribution; 17 of the qualifying lines have since been irrevocably lost (<http://stemcells.nih.gov/index.asp>). It is clear that individual hESC lines can exhibit differences in their growth rates (24) and aspects of their gene expression (5), characteristics that may also be predictive of different differentiation capacities. The relatively low success rate in establishing viable hESC lines from the initial presidential-qualified derivations may critically limit the diversity of the hESC lines available.

STR genotyping and HLA isotyping defined unique molecular signatures for BG01, BG02, and BG03 cells, which will be useful for accurate tracking of cultures when multiple lines are maintained simultaneously. Furthermore, one promise of hESC research is the possibility of cell therapy treatments for degenerative diseases such as Parkinson's disease or diabetes (25–28). Should the NIH-qualified hESC lines eventually reach clinical applications, many potential transplantation therapies will be governed by typical HLA matching criteria: Potential recipients may have to be matched for suitable HLA loci with the donor hESC line. Currently, only the HLA isotype of BG01, BG02, and BG03 and four other

hESC lines has been reported (22). All these seven lines exhibited a different HLA isotype. Because of the limitation on the hESC lines available to federal funding, and how this may relate to the cost of eventual therapies and the population demographic that may be able to afford such therapies, it is critical to determine the range of different HLA types represented by the currently available lines, and how effectively this represents the potential patient population. This information is highly relevant in determining the effectiveness of policies that govern hESC research in the long term and the actual application of potential therapies.

The BG01, BG02, and BG03 hESC lines expressed the expected profile of cell-surface antigens, alkaline phosphatase activity, and transcriptional markers, indicating that they were undifferentiated and pluripotent. A more detailed MPSS analysis of the BG02 line revealed the expression of 8381 transcripts (Table 4, 17-nucleotide survey), and defined their expression level in a digital readout. This MPSS data should be directly comparable to other hESC lines or their differentiated derivatives assayed with the same approach, providing a powerful reference for future analyses. The BG02 line expressed a large majority (84.7%) of an expanded panel of 92 markers of pluripotency, which have previously been suggested to define this cell state (5,11), confirming its similarity to other undifferentiated hESC lines. Similarly, BG03 cells maintained in feeder-free conditions expressed a similar profile of transcripts using focused arrays, confirming that these cultures consisted primarily of undifferentiated cells. The generation of these gene expression data sets provides reference points for future analysis of cellular differentiation as well as highlighting cytokines and cell signaling pathways that may be involved in regulating pluripotency or differentiation.

The pluripotency of BG01, BG02, and BG03 cells was demonstrated by the differentiation of embryoid bodies

CHARACTERIZATION OF THE BG01, BG02, AND BG03 HUMAN ES CELL LINES

TABLE 5. MPSS QUANTITATION OF MARKERS OF PLURIPOTENCY AND DIFFERENTIATED LINEAGES IN BG02 hESCs

<i>Unigene number</i>	<i>20-nt signature tpm^a</i>	<i>17-nt signature tpm^a</i>	<i>Description</i>
Markers of pluripotency			
Hs.250769	235	187	Alkaline phosphatase, liver/bone/kidney
Hs.439911	28	19	Telomerase reverse transcriptase
Hs.249184	2096	1978	POU domain, class 5, transcription factor 1
Hs.329296	11	11	Nanog homeobox
Hs.816	269	220	SRY (sex determining region Y)-box 2
Hs.458406	0	123	Undifferentiated embryonic cell transcription factor 1
Hs.385870	0	175	Teratocarcinoma-derived growth factor 1
Hs.74471	2622	2012	Gap junction protein, alpha 1, 43 kDa (connexin 43)
Hs.532492	0	19	Gap junction protein, alpha 7, 45 kDa (connexin 45)
Markers of differentiated lineages			
Hs.155421	0	0	AFP
Hs.389457	0	0	Brachyury
Hs.37055	0	0	Fibroblast growth factor 5
Hs.424414	0	0	Msx1
Hs.202526	0	0	Sox1
Hs.80342	0	0	Keratin 15
Highly expressed transcripts			
Hs.439552	17862	19227	Eukaryotic translation elongation factor 1 alpha 1
Hs.410817	15550	11532	Ribosomal protein L13
Hs.426930	10497	9773	Actin, beta
Hs.449070	12958	9562	Ribosomal protein L13a
Hs.443796	11303	8361	Ribosomal protein, large, P0
Hs.169476	10350	7825	Glyceraldehyde-3-phosphate dehydrogenase
Hs.178551	7166	7530	Ribosomal protein L8
Hs.498569	6352	7192	Ribosomal protein S2
Hs.181307	8882	6717	H3 histone, family 3A
Hs.388664	8774	6441	Ribosomal protein L11
Hs.14376	4169	6024	Actin, gamma 1
Hs.356721	6155	5970	Heterogeneous nuclear ribonucleoprotein A1
Hs.374588	7556	5892	Ribosomal protein L17
Hs.311640	7412	5543	Ribosomal protein S27a
Hs.8102	7434	5466	Ribosomal protein S20
Hs.74497	7286	5338	Nuclease sensitive element binding protein 1
Hs.469653	6321	5076	Ribosomal protein L5
Hs.446579	5707	4670	Heat shock 90kDa protein 1, alpha
Hs.356331	5556	4625	Peptidylprolyl isomerase A (cyclophilin A)
Hs.433701	5526	4467	Ribosomal protein L37a
Hs.401929	5301	4326	Ribosomal protein L10
Hs.186350	5235	4029	Ribosomal protein L4
Hs.275865	5057	4149	UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 4
Hs.406682	5032	3680	Ribosomal protein L26

^atpm, Transcripts per million.

in vitro to representatives of the three primary germ layers. The emergence of cells exhibiting differentiated phenotypes and expressing markers of mesoderm, endoderm, and ectoderm, was concomitant with a reduction in OCT4 and NANOG expression. The directed differentiation of

BG01 and BG03 cells to neuroepithelia and multiple differentiated neuronal lineages, including cells expressing multiple markers of the midbrain dopaminergic lineage, has previously been demonstrated (13,14). These differentiation experiments were responsive to exogenous cues

found during embryogenesis, such as BMP4, and highlight that these hESC lines have pluripotent potential and form a powerful model of human embryogenesis.

Although some laboratories have reported that hESCs can be maintained with a normal karyotype for extended periods (10,15), others have described recurrent accumulation of trisomies of chromosomes 12 and 17 (9). Triploidies of chromosomes 12, 17, and X were observed in BG01 cultures that had been routinely passaged by disaggregation to single cells. One interpretation of this data is that disaggregation of hESCs to single cells during passaging exerts selection pressure on viable cells, such that sporadic and random segregation events for these chromosomes imparts a significant selective advantage. Consistent with this, when these three hESC lines were passaged as clumps of cells, only a single aneuploidy (of chromosome 11) was detected in 13 karyotype assays with a combined total of 241 metaphase spreads. Furthermore, the BG03 line could be maintained in feeder-free conditions for more than 24 passages, and exhibited a normal karyotype at passage 16 and gene expression generally consistent with the undifferentiated state. These results provide strong confirmation that these hESC lines can be maintained in an undifferentiated state over long-term culture in multiple conditions, without accumulation of karyotypic abnormalities.

The properties of the BG01, BG02, and BG03 hESC lines described here provide a range of essential information about the molecular and cellular characteristics of these cells. However, a more coordinated approach would involve examination of the growth, differentiation potential in vitro and in vivo, epigenetic status, and marker expression analyses of a large number of hESC lines, which would be maintained and differentiated in standardized conditions. Such a large-scale collaborative effort has been proposed as an International Stem Cell Initiative (Peter Andrews, personal communication), and is a necessary step to addressing fundamental questions such as defining unique qualities of individual lines and exposing subtle differences between these and other hESC lines.

ACKNOWLEDGMENTS

We would like to acknowledge Michelle Sturgeon and Erin Fotter for technical assistance. This work, and the distribution of these hESC lines, is partly supported by NIH grant R24DK063689 awarded to BresaGen Inc.

REFERENCES

1. Reubinoff BE, MF Pera, CY Fong, A Trounson and A Bongso. (2000). Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. *Nature Biotechnol* 18:399–404.
2. Thomson JA, J Itskovitz-Eldor, SS Shapiro, MA Waknitz, JJ Swiergiel, VS Marshall and JM Jones. (1998). Embryonic stem cell lines derived from human blastocysts. *Science* 282:1145–1147.
3. Mitalipova MM, JD Calhoun, S Shin, W Wininger, TC Schulz, SA Noggle, V Venable, IG Lyons, A Robins and SL Stice. (2003). Human embryonic stem cell lines derived from discarded embryos. *Stem Cells* 21:521–526.
4. Carpenter MK, E Rosler and MS Rao. (2003). Characterization and differentiation of human embryonic stem cells. *Cloning Stem Cells* 5:79–88.
5. Bhattacharya B, T Miura, R Brandenberger, J Mejido, Y Luo, AX Yang, BH Joshi, I Ginis, RS Thies, M Amit, I Lyons, BG Condie, J Itskovitz-Eldor, MS Rao and RK Puri. (2004). Gene expression in human embryonic stem cell lines: unique molecular signature. *Blood* 103:2956–2964.
6. Richards M, SP Tan, JH Tan, WK Chan and A Bongso. (2004). The transcriptome profile of human embryonic stem cells as defined by SAGE. *Stem Cells* 22:51–64.
7. Ginis I, Y Luo, T Miura, S Thies, R Brandenberger, S Gerecht-Nir, M Amit, A Hoke, MK Carpenter, J Itskovitz-Eldor and MS Rao. (2004). Differences between human and mouse embryonic stem cells. *Dev Biol* 269:360–380.
8. Reubinoff BE, P Itsykson, T Turetsky, MF Pera, E Reinhartz, A Itzik and T Ben-Hur. (2001). Neural progenitors from human embryonic stem cells. *Nature Biotechnol* 19:1134–1140.
9. Draper JS, K Smith, P Gokhale, HD Moore, E Maltby, J Johnson, L Meisner, TP Zwaka, JA Thomson and PW Andrews. (2004). Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. *Nature Biotechnol* 22:53–54.
10. Buzzard JJ, NM Gough, JM Crook and A Colman. (2004). Karyotype of human ES cells during extended culture. *Nature Biotechnol* 22:381–382.
11. Zeng X, T Miura, Y Luo, B Bhattacharya, B Condie, J Chen, I Ginis, I Lyons, J Mejido, RK Puri, MS Rao and WJ Freed. (2004). Properties of pluripotent human embryonic stem cells BG01 and BG02. *Stem Cells* 22:292–312.
12. Schulz TC, GM Palmarini, SA Noggle, DA Weiler, MM Mitalipova and BG Condie. (2003). Directed neuronal differentiation of human embryonic stem cells. *BMC Neurosci* 4:27.
13. Zeng X, J Cai, J Chen, Y Luo, Z-B You, E Fotter, Y Wang, B Harvey, T Miura, C Backman, G-J Chen, MS Rao and WJ Freed. (2004). Dopaminergic differentiation of human embryonic stem cells. *Stem Cells* (in press).
14. Schulz TC, SA Noggle, GM Palmarini, DA Weiler, IG Lyons, KP Pensa, ACB Meedeniya, BP Davidson, NA Lambert and BG Condie. (2004). Differentiation of human embryonic stem cells to dopaminergic neurons in serum free suspension culture. *Stem Cells* (in press).
15. Rosler ES, GJ Fisk, X Ares, J Irving, T Miura, MS Rao and MK Carpenter. (2004). Long-term culture of human embryonic stem cells in feeder-free conditions. *Dev Dynam* 229:259–274.
16. Hansen JA and B Dupont, ed. (2004). *HLA 2004: Immunobiology of the Human MHC. Proceedings of the 13th International Histocompatibility Workshop and Congress*. IHWG Press, Seattle, WA.

17. Luo Y, J Cai, I Ginis, Y Sun, S Lee, SX Yu, A Hoke and MS Rao. (2003). Designing, testing, and validating a focused stem cell microarray for characterization of neural stem cells and progenitor cells. *Stem Cells* 21:575–587.
18. Jongeneel CV, C Iseli, BJ Stevenson, GJ Riggins, A Lal, A Mackay, RA Harris, MJ O'Hare, AM Neville, AJ Simpson and RL Strausberg. (2003). Comprehensive sampling of gene expression in human cell lines with massively parallel signature sequencing. *Proc Natl Acad Sci USA* 100: 4702–4705.
19. Brenner S, M Johnson, J Bridgham, G Golda, DH Lloyd, D Johnson, S Luo, S McCurdy, M Foy, M Ewan, R Roth, D George, S Eletr, G Albrecht, E Vermaas, SR Williams, K Moon, T Burcham, M Pallas, RB DuBridge, J Kirchner, K Fearon, J Mao and K Corcoran. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnol* 18:630–634.
20. Brenner S, SR Williams, EH Vermaas, T Storck, K Moon, C McCollum, JI Mao, S Luo, JJ Kirchner, S Eletr, RB DuBridge, T Burcham and G Albrecht. (2000). In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. *Proc Natl Acad Sci USA* 97:1665–1670.
21. Xu C, MS Inokuma, J Denham, K Golds, P Kundu, JD Gold and MK Carpenter. (2001). Feeder-free growth of undifferentiated human embryonic stem cells. *Nature Biotechnol* 19:971–974.
22. Carpenter MK, ES Rosler, GJ Fisk, R Brandenberger, X Ares, T Miura, M Lucero and MS Rao. (2004). Properties of four human embryonic stem cell lines maintained in a feeder-free culture system. *Dev Dynam* 229:243–258.
23. Amit M, C Shariki, V Margulets and J Itskovitz-Eldor. (2004). Feeder layer- and serum-free culture of human embryonic stem cells. *Biol Reprod* 70:837–845.
24. Cowan CA, I Klimanskaya, J McMahon, J Atienza, J Witmyer, JP Zucker, S Wang, CC Morton, AP McMahon, D Powers and DA Melton. (2004). Derivation of embryonic stem-cell lines from human blastocysts. *N Engl J Med* 350: 1353–1356.
25. Colman A. (2004). Making new beta cells from stem cells. *Semin Cell Dev Biol* 15:337–345.
26. Bjorklund A, SB Dunnett, P Brundin, AJ Stoessl, CR Freed, RE Breeze, M Levivier, M Peschanski, L Studer and R Barker. (2003). Neural transplantation for the treatment of Parkinson's disease. *Lancet Neurol* 2:437–445.
27. Freed WJ. (2004). A perspective on transplantation therapy and stem cells for Parkinson's disease. *Cell Transplant* 13:319–327.
28. Ginis I and MS Rao. (2003). Toward cell replacement therapy: promises and caveats. *Exp Neurol* 184:61–77.

Address reprint requests to:
Dr. Thomas C. Schulz
BresaGen, Inc.
111 Riverbend Road
Athens, GA 30605

E-mail: tschulz@novocell.com

Received July 31, 2004; accepted September 6, 2004.