

# GeneChip analysis of human embryonic stem cell differentiation into hemangioblasts: an *in silico* dissection of mixed phenotypes

Shi-Jiang Lu<sup>✉\*</sup>, Jennifer A Hipp<sup>✉†</sup>, Qiang Feng<sup>\*</sup>, Jason D Hipp<sup>†</sup>, Robert Lanza<sup>\*†</sup> and Anthony Atala<sup>†</sup>

Addresses: \*Advanced Cell Technology, Worcester, MA 01605, USA. †Institute of Regenerative Medicine, Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA.

✉ These authors contributed equally to this work.

Correspondence: Anthony Atala. Email: aatala@wfubmc.edu

Published: 13 November 2007

Genome **Biology** 2007, **8**:R240 (doi:10.1186/gb-2007-8-11-r240)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2007/8/11/R240>

Received: 11 June 2007

Revised: 10 July 2007

Accepted: 13 November 2007

© 2007 Lu et al.; licensee BioMed Central Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

**Background:** Microarrays are being used to understand human embryonic stem cell (hESC) differentiation. Most differentiation protocols use a multi-stage approach that induces commitment along a particular lineage. Therefore, each stage represents a more mature and less heterogeneous phenotype. Thus, characterizing the heterogeneous progenitor populations upon differentiation are of increasing importance. Here we describe a novel method of data analysis using a recently developed differentiation protocol involving the formation of functional hemangioblasts from hESCs. Blast cells are multipotent and can differentiate into multiple lineages of hematopoietic cells (erythroid, granulocyte and macrophage), endothelial and smooth muscle cells.

**Results:** Large-scale transcriptional analysis was performed at distinct time points of hESC differentiation (undifferentiated hESCs, embryoid bodies, and blast cells, the last of which generates both hematopoietic and endothelial progenies). Identifying genes enriched in blast cells relative to hESCs revealed a genetic signature indicative of erythroblasts, suggesting that erythroblasts are the predominant cell type in the blast cell population. Because of the heterogeneity of blast cells, numerous comparisons were made to publicly available data sets *in silico*, some of which blast cells are capable of differentiating into, to assess and characterize the blast cell population. Biologically relevant comparisons masked particular genetic signatures within the heterogeneous population and identified genetic signatures indicating the presence of endothelia, cardiomyocytes, and hematopoietic lineages in the blast cell population.

**Conclusion:** The significance of this microarray study is in its ability to assess and identify cellular populations within a heterogeneous population through biologically relevant *in silico* comparisons of publicly available data sets. In conclusion, multiple *in silico* comparisons were necessary to characterize tissue-specific genetic signatures within a heterogeneous hemangioblast population.

## Background

The establishment of human embryonic stem cells (hESCs) raised the possibility of being able to treat/cure many human diseases that are nowadays untreatable. This therapeutic potential, however, largely relies on the efficient and controlled differentiation of hESCs towards a specific cell type and the generation of homogeneous cell populations. Many differentiation protocols utilize the formation of progenitors through a stepwise approach. Thus, characterizing and understanding mixed populations of progenitor stages will be of increasing importance in stem cell research.

hESCs have been shown to be able to differentiate into a variety of cell types, including hematopoietic precursors and endothelial cells, *in vitro* under various culture conditions [1-9]. Hemangioblasts are the precursors of both hematopoietic and endothelial cells [10]. The existence of hemangioblasts was first demonstrated using an *in vitro* differentiation system of mouse ESCs. Replating of embryonic bodies (EBs) of mouse ESCs resulted in the formation of blast colony forming cells (BL-CFCs), which possessed hemangioblastic characteristics: BL-CFCs generated both hematopoietic and endothelial cells upon transfer to appropriate conditions [11,12]. Cells with hemangioblastic characteristics have been reported in both mouse and human adult tissues [13-18]. In an hESC system, Wang *et al.* [3] found that a fraction of a percent (0.18%) of CD45<sup>neg</sup>FVP cells with hemangioblast-like properties in hESCs derived from EBs. Zambidis *et al.* [8] demonstrated the formation of multi-potential colonies from hEBs, although it is unclear whether these colonies can be expanded and/or whether they have any functional activity *in vivo*. Umeda *et al.* [19] also identified the presence of CD34+/KDR+ bipotential cells in non-human (*Cynomolgus*) ESCs. Kennedy *et al.* [20] recently reported the generation of BL-CFCs from hESCs. However, the rarity of the cells with hemangioblast properties both from adult tissues and from ESC systems precluded comprehensive analysis of gene expression and comparison with other populations.

We have recently developed a two-step strategy that can efficiently and reproducibly generate blast colonies (BCs), the human counterparts of BL-CFCs, from hESCs [21]. These BC cells expressed gene signatures characteristic of hemangioblasts, and could be differentiated into multiple hematopoietic cell lineages as well as endothelial cells. When the BC cells were injected into animals with spontaneous type II diabetes or ischemia/reperfusion injury of the retina, they homed to the site of injury and showed robust reparative function of the damaged vasculature. The cells also showed a similar regenerative capacity in NOD/SCID  $\beta 2^{-/-}$  mouse models of both myocardial infarction (50% reduction in mortality rate) and hind limb ischemia, with restoration of blood flow in the latter model to near normal levels, demonstrating the functional properties of hemangioblasts *in vivo* [21]. In contrast to previous studies, these cells could be readily obtained in large scale, which allowed us to perform comprehensive analysis of

gene expression in these cells and compare this with other cell populations from which the BC cells originated.

Microarrays assess the total amount of RNA in a population and can be influenced by a predominating cell type. Variation in the homogeneity of the population can influence the number of genes identified as differentially expressed. Here, we show how comparisons to publicly available tissues *in silico* can identify differentially expressed genes representative of the various cell types within a heterogeneous population.

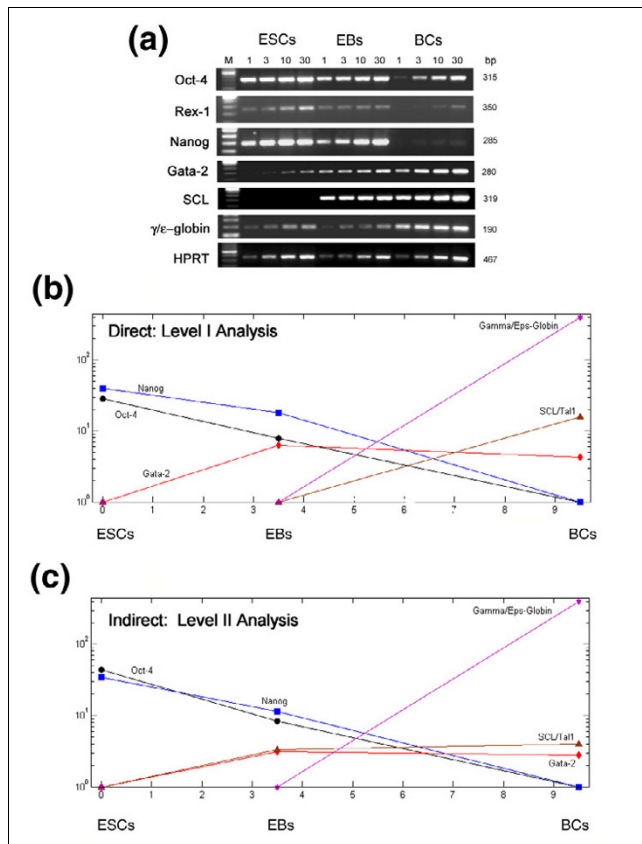
In the current study, we analyzed the global gene expression profiles with robust multi-chip average (RMA) normalization to provide a relative value of gene expression between two samples. The first analysis consisted of direct comparisons with ESCs and their derivatives (EBs and BCs). Genes enriched in BCs relative to hESCs revealed a genetic signature indicative of erythroblasts, suggesting that erythroblasts are the predominant cell type in the BC population. The next analysis consisted of multiple but biologically meaningful *in silico* comparisons to publicly available data sets that identified other progenitor cell types within the BC population. The significance of this microarray study is in its ability to assess and identify heterogeneous cellular populations through biologically relevant *in silico* comparisons.

## Results

### Strategy

Microarrays assess the total amount of RNA in a population and can, therefore, be influenced by a predominating cell type. Variations in the homogeneity of the population can influence the number of genes identified as differentially expressed, especially if both populations are relatively homogeneous. Here, we show how comparisons to publicly available tissues *in silico* can identify differentially expressed genes representative of the various cell types within the heterogeneous population of BCs.

We describe our method of assessing heterogeneous samples in three levels of analysis. The first level consists of making direct comparisons within the ESCs and their differentiated derivatives (EBs and BCs). The advantage of this technique is that it provides a kinetic-like relationship of changes in gene expression upon differentiation. The second level of analysis consists of indirect comparisons to a baseline, or reference tissue. Breast epithelium was chosen as a reference tissue because it represents a genetically distinct cell type that BCs are not capable of differentiating into. ESCs, EBs, and BCs were compared to breast epithelial tissue and differentially expressed genes were compared and contrasted to each other. Because genes that are up-regulated in BCs when compared to breast epithelia could represent those that are under-expressed in breast tissue, we removed those that were also up-regulated in a genotypically similar but different cell type (hESCs), when compared to breast epithelia.



**Figure 1**  
Validation of differentially expressed genes by RT-PCR in human ESCs, EBs and BCs. **(a)** Total RNA from human ESCs, EBs and BCs was used to construct cDNA pools, and the expression of genes was examined by semi-quantitative PCR. The number at the top of each lane indicates the amount (microliters) of cDNA used in the 50  $\mu$ L PCR reaction. M = 100 bp DNA ladder. **(b)** Direct and **(c)** indirect analysis of differentially expressed genes matched the expression patterns obtained by RT-PCR. The fold change data are presented on the y-axis using logarithm-base-10.

The third level of analysis consists of comparing BCs to tissues they are capable of differentiating into as a way to mask that cell type's 'genetic signature' and reveal signatures of the more minor cell types. Samples were chosen based on type of GeneChip and their public availability - leukocytes, and endothelial and stromal cells. These biologically relevant comparisons identified tissue specific genetic signatures that would have otherwise been missed in the level I and II analyses.

The reliability of the microarray data generated from our multi-comparison analysis is demonstrated by the consistent identification of a set of genes among multiple comparisons, of which a subset of genes were confirmed by immunocytochemistry (Table 1) and RT-PCR (Figure 1). To summarize, comparing BCs to leukocytes identified genes involved in vasculogenesis, to endothelial cells identified genes involved in hematopoiesis, and to stromal cells identified genes involved in heart development.

**Level I analysis**

*Genes down-regulated upon differentiation of ESCs into EBs and BCs*  
We began our data analysis by verifying the expression of 'stemness' genes that are down-regulated in ESCs upon differentiation into EBs and BCs. We identified 87 genes that were down-regulated upon differentiation into EBs. Genes with the highest fold change include *SOX2*, *LEFTY1*, *GAL*, *NODAL*, *OCT4*, and *THY1*, which play a critical role in maintaining the undifferentiated status of ESCs [22-24]. To uncover enriched processes, data sets were analyzed by DAVID, a web-based tool that identifies over-represented biological themes in a data set based on their Gene Ontology (GO) terms. GO provides consistent descriptions of genes in terms of biological processes and molecular function. When these genes were clustered with DAVID based on their GO terms, processes involved in development, cell differentiation, and proliferation were identified. The genes identified in the development ontology were *DNA methyltransferase 3B*, *FGF2*, *THY1*, *SFRP2*, *LEFTY1*, *GREM1*, and *NODAL* (Additional data file 3).

We also identified 267 genes that were down-regulated upon differentiation of ESCs to BCs. These genes include *GAL*, *TDGF*, *NANOG*, *LEFTY1*, and *OCT4*, most of which are stemness genes [22-24]. When genes were clustered with DAVID using their GO terms, the processes included development, cell differentiation, and morphogenesis (Additional data file 4). These data demonstrate that *OCT4*, *NODAL*, *GAL*, and *THY1* are initially down-regulated in stage 1 (ESCs→EBs) and are further down-regulated in stage 2 (EBs→BCs).

*Genes up-regulated upon differentiation of ESCs into EBs*

While the focus of this paper is to evaluate the pathways involved in hemangioblast differentiation, we begin by identifying those genes that were up-regulated in the early stage of differentiation into EBs (day 3.5) from which BCs were derived [21]. We identified 128 genes that were up-regulated upon differentiation of ESCs into EBs (Additional data file 5). These genes include *HAND1*, *WNT5*, *HEY1*, *LMO2*, *BMP4*, *TBX3*, and *MYL4*. Clustering these genes with EASE identified processes involved in development, transcription, organ development and system development, some of which are related to hemangioblastic differentiation. These genes include *SOX9*, *HOXB2*, *HOXB3*, *Neuregulin 1*, *LMO2* [25] and *GATA2* [26]. This data set also included numerous genes encoding transcription factors, such as *MESP1*, *HAND1*, *TBX3*, *GATA2*, *SOX7*, *SOX9*, *HOXB2*, and *HOXB3*.

*Genes down-regulated upon differentiation of EBs into BCs*

When EBs were compared to BCs, 185 genes were identified as down-regulated upon differentiation. This data set contained processes that were similar to those that were down-regulated upon ESC differentiation into EBs, such as tissue and organ development. The most significantly down-regulated genes included *NANOG*, *WNT5*, *OCT4*, *GAL*, *TDGF1*, *BMP4*, *endothelin receptor B*, and *VEFG* (Additional data file

**Table 1**

**Characterization of hESCs and BCs by immunocytochemistry and Affymetrix arrays**

	IC	Level I	Level II	Level III		
		BC direct	Epithelium	Leukocytes	Endothelium	Stromal
GATA1	+	+	+	+	+	+
GATA2	+	+	+	--	--	--
CD71	+	+	+	+	--	+
EPO-R	+	--	+	+	+	+
TPO-R	+	--	--	--	--	+
LMO2	+	+	+	+	--	+
β-Catenin	+	--	--	--	--	--
EGR1	+	--	--	--	--	--
Integrin α4	+/-	--	--	--	--	--
Integrin β1	+/-	--	--	--	--	--
VE-cadherin	--	--	--	+	--	--
E-cadherin	--	--	--	--	--	--
KDR	--	--	--	--	--	--
PECAM (CD31)	--	--	--	+	--	--
CD34	--	--	--	+	--	--
CD41	--	--	CD41b	--	CD41b	CD41b
CD43	--	--	--	+	+	+

The reliability of the microarray data generated from our multi-comparison analysis is demonstrated by the consistent identification of a set of genes among multiple comparisons, of which a subset of genes were confirmed by immunocytochemistry. For immunochemistry (IC): +, moderate to strong staining; -, negative staining; +/-, very weak staining. For Affymetrix arrays: +, detected as up-regulated in BCs; --, not detected as up-regulated in BCs.

6). These data demonstrate that *BMP4*, *WNT5*, and *HEY1* are initially up-regulated upon differentiation into EBs but then down-regulated upon further differentiation into BCs.

*Genes up-regulated upon differentiation of EBs into BCs*

In contrast, 82 genes were up-regulated upon differentiation of EBs into BCs. The genes with the greatest fold change were hemoglobin genes and erythropoietic genes, such as *hemoglobins γ, ζ, α, and ε, Alas2, AFP, TUBB1, GYPA, and RHAG* (fold change, (FC) 31x-886x). Genes with moderate increases in expression (FC 6.2x-7.2x) were *KLF1, TAL1/SCL, GATA1* and *CD71*. When genes were clustered with DAVID using their GO terms, processes characteristic of erythropoiesis

(heme and porphyrin biosynthesis and oxygen transport) were identified (Additional data file 7).

*Genes up-regulated upon differentiation of ESCs into BCs*

There were 107 genes up-regulated upon differentiation of ESCs into BCs. Similar to the data set above (EBs→BCs), the genes with the greatest fold change (FC 29x-810x) were involved in hemoglobin synthesis (hemoglobins *γ, ε, and α, ALA2, GYPA, and TAL1/SCL*), similar to the comparison of EBs to BCs. In addition, this data set contained many key transcription factors involved in hemangioblastic differentiation, such as *GATA2* [26], *LMO2* [25] and *TAL1/SCL* [27,28] (Additional data file 8). *GATA2* and *MYL4* were

**Table 2**

**Gene ontologies for up-regulated processes in BCs versus ESCs**

Gene category	List hits	EASE score
Oxygen transport (erythrocytic)	6	1.69E-09
Gas transport (erythrocytic)	6	1.69E-09
Transcription from Pol II promoter	7	1.64E-02
Regulation of transcription from Pol II promoter	5	2.38E-02
Development (developmental)	15	2.56E-02

EASE analysis of up-regulated genes in BCs identified biologically relevant themes, such as oxygen and gas transport, and development.

up-regulated upon differentiation into EBs and remained at a constant level upon differentiation into BCs. EASE analysis of up-regulated genes in BCs identified biologically relevant themes, such as oxygen and gas transport, and development (TAL1/SCL, KLF1, LMO2, GATA1; Table 2).

### Level II analysis

#### Genes enriched in ESCs

Since genes that are up-regulated in ESCs when compared to breast epithelia could represent those genes that are under-expressed in breast epithelium, we filtered out those that were also up-regulated in BCs when compared to breast epithelium. This analysis identified 2,108 genes, which comprised GO processes involved in cell cycle, DNA and RNA metabolism, and DNA replication, as expected (Additional data file 9). Genes with the highest fold change include *TDGF1*, *GAL*, *LEFTY1/2*, *OCT4*, and *NANOG* (FC 130.0x, 69.6x, 68.7x, 47.6x, 43.8x, and 31.5x). When this data set was clustered based on their GO terms, processes involved in development, cell differentiation and nervous system development were identified (data not shown). This data set was then analyzed with GenMapp, and then used for pathway analysis. Each genetic signature was assigned a color: ESCs, green; EBs, orange; and BCs, red. The ESC pathway confirms that most of the embryonic genes were not removed when compared to breast epithelial cells (Additional data file 1).

#### Genes enriched in EBs

Since genes up-regulated in EBs when compared to breast epithelia could similarly represent those that are under-expressed in breast epithelium, we also filtered out those that were also up-regulated in ESCs when compared to breast epithelium. We identified 939 genes as up-regulated in EBs relative to breast epithelium and filtered out those that are enriched in ESCs relative to breast epithelium (Additional data file 10). When these genes were clustered with DAVID, processes involved in development, transcription, wnt and frizzles signaling, cell cycle and blood vessel morphogenesis (*KDR*, *VEGF*, *Neuropilin-1* and *2*, and *FLT1*) were identified. The EBs also express genes involved in organ development, suggesting a heterogeneous mixture of cell types (*GATA2/4/5/6*, *BMP4*, *NCAM1*, *NOG*, *ISL2*, *NKX2.5*), and mesoderm genes (*HAND1*, *T-brachyury*, *MESP1*). Further examination of the data set identified multiple genes involved in the BMP signaling pathway in the differentiation of blood and endothelial cells. These genes include *BMPRI1A*, *BMP4*, *T-brachyury*, *KDR*, *GATA2*, and *TAL1/SCL* [26,28].

#### Genes enriched in BCs

We identified 2,735 genes that were up-regulated in BCs relative to breast epithelium after removing genes that were enriched in ESCs when compared to breast epithelium and genes enriched in BCs when compared to breast epithelium (Additional data file 11). When genes were clustered based on their GO, we identified processes characteristic of lymphocytic cells (response to stimulus, defense response,

immune response), erythrocytes (heme and porphyrin biosynthesis), coagulation, neurophysiology, development, and mesoderm and heart development (Table 3).

Genes that were up-regulated in BCs with respect to epithelia were characteristic of hematopoiesis (CD markers 5/6/9/38/41/48/55/71/74/84/244, *EPOR*, *GATA1/2/4/5*, *Tcr- $\alpha$* , *natural cytotoxicity triggering receptor-1, 2* and *3*), coagulation (*coagulation factor II, V, VII, XII, coagulation factor 2 receptor like 2,3/thrombin receptors, antithrombin 3, cyclooxygenase 1, plasminogen*), cardiac muscle (*NKX2.5, HAND1/2, GATA4, SOX6, TBX5*), smooth/skeletal muscle (*NOTCH1, smoothelin, acetylcholinesterase, desmin, SOX6*), synaptic markers (*cholinergic receptor, muscarinic 2 and 5, adrenergic  $\alpha$ -1A-receptor, dopamine receptor D2, serotonin receptor 1B and 4, glutamate receptor Nmda 1 and 2A/B and C, gaba A receptor  $\beta$ 1 and 2, purinergic receptor P2X 1 and 2*), and hemangioblasts (*GATA2* [26], *RUNX1* [29], *LMO2* [25] and *TAL1/SCL* [27,28]). Some of the genes identified in the coagulation ontology are not only involved in coagulation but also angiogenesis, such as *thrombin* [30], *plasminogen* [31], and possibly *coagulation factor 2 receptor like 2/3* [30].

This was also recapitulated by DAVID analysis, which identified the following pathways as statistically over-represented: porphyrin metabolism, acute myocardial infarction, hematopoietic cell lineage pathways and calcium signaling pathways. GenMapp was then used for pathway analysis. Each genetic signature was assigned a color: ESC, green; EBs, orange; and BCs, red. GenMapp analysis of pathways involved in whole blood, bone marrow, coagulation and complement, heme and porphyrin synthesis are indicative of hematopoietic cell types (Figure 2). Genes identified in GenMapp's heme biosynthesis pathway are indicative of erythroblasts (Figure 3). We also identified myogenic (cardiac and smooth) pathways, which correlates with the GO analysis.

### Level III analysis

#### Genes enriched in BCs relative to leukocytes

We identified 2,101 genes that were up-regulated in BCs relative to leukocytes (after removing genes that were enriched in both hESCs and BCs when compared to leukocytes (Additional data file 12). When these genes were clustered based on their GO, we identified processes involved in development, nervous system development, blood vessel development and angiogenesis, and erythrocytes (Table 4). The presence of development ontology not only indicates a 'progenitor' status of BCs, but contains genes involved in hemangioblast development, such as *LMO2*, *TAL1/SCL*, and *RUNX1*. This comparison identified genes that are characteristic of endothelia (*PECAM1*, *VE-Cadherin*, *CD34*, *vWF*, *EPOR* [32], *endothelin 1* [33], and *thrombin receptor*). There were also genes that indicate the presence of erythrocytes (*GATA1*, *spectrin* and *ankyrin*), blood vessel development (*neuropilin-1* and *2*, *stabilin 1* and *2*, *EGFR*, *FGF1* and *6*, *NOTCH4*), and neurons/neuronal junctions (*glutamate receptor 1/6*, *serotonin*

**Table 3****Gene ontologies of up-regulated processes in BCs versus epithelial cells**

Gene category	List hits	EASE score
Response to external stimulus (lymphocytic)	146	5.14E-09
Cellular process	490	4.74E-06
Response to biotic stimulus (lymphocytic)	94	4.87E-06
Ion transport	68	7.05E-06
Defense response (lymphocytic)	87	1.09E-05
Development (developmental)	165	5.05E-05
Hemostasis	19	1.15E-04
Blood coagulation (coagulation)	18	1.68E-04
Cell-cell signaling	60	2.23E-04
Response to abiotic stimulus (lymphocytic)	60	3.74E-04
Cyclic-nucleotide-mediated signaling	19	3.92E-04
Morphogenesis (developmental)	101	4.23E-04
Response to light	26	5.54E-04
Natural killer cell activation (lymphocytic)	6	7.08E-04
Second-messenger-mediated signaling	20	8.47E-04
Perception of light	25	9.29E-04
Synaptic transmission (neurophysiology)	31	9.63E-04
Cell communication	244	1.07E-03
Cation transport	46	1.46E-03
Transmission of nerve impulse (neurophysiology)	31	1.61E-03
Transport	159	1.65E-03
Response to radiation	26	1.73E-03
Organogenesis (developmental)	88	2.09E-03
Metal ion transport	36	2.15E-03
Perception of abiotic stimulus (lymphocytic)	35	2.34E-03
Vision	23	2.54E-03
Immune response (lymphocytic)	68	3.77E-03
Perception of external stimulus (lymphocytic)	37	4.60E-03
Cell surface receptor linked signal transduction	93	5.44E-03
Xenobiotic metabolism	10	6.19E-03
Sensory perception	34	6.34E-03
Heme biosynthesis (erythrocytic)	5	6.40E-03
Response to xenobiotic stimulus	10	8.00E-03
Skeletal development (development)	18	8.65E-03
Potassium ion transport	19	8.71E-03
G-protein signaling	15	9.32E-03
Pigment biosynthesis (erythrocytic)	6	1.10E-02
Monovalent inorganic cation transport	29	1.15E-02
Heme metabolism (erythrocytic)	5	1.16E-02
Negative regulation of natural killer cell activity (lymphocytic)	3	1.28E-02
Regulation of natural killer cell activity (lymphocytic)	3	1.28E-02
Anion transport	15	1.49E-02
Heart development (heart development)	6	1.63E-02
Histogenesis (developmental)	16	1.85E-02
Porphyrin biosynthesis (erythrocytic)	5	1.90E-02
Pigment metabolism (erythrocytic)	6	1.95E-02

**Table 3 (Continued)**

Gene ontologies of up-regulated processes in BCs versus epithelial cells		
Pigmentation (erythrocytic)	6	2.30E-02
Mesoderm development (mesoderm development)	7	2.37E-02
Cellular defense response (lymphocytic)	13	2.44E-02
Response to pest/pathogen/parasite (lymphocytic)	43	2.63E-02

When genes were clustered based on their GO, we identified processes characteristic of lymphocytic cells (response to stimulus, defense response, immune response), erythrocytes (heme and porphyrin biosynthesis), coagulation, neurophysiology, development, and mesoderm and heart development.

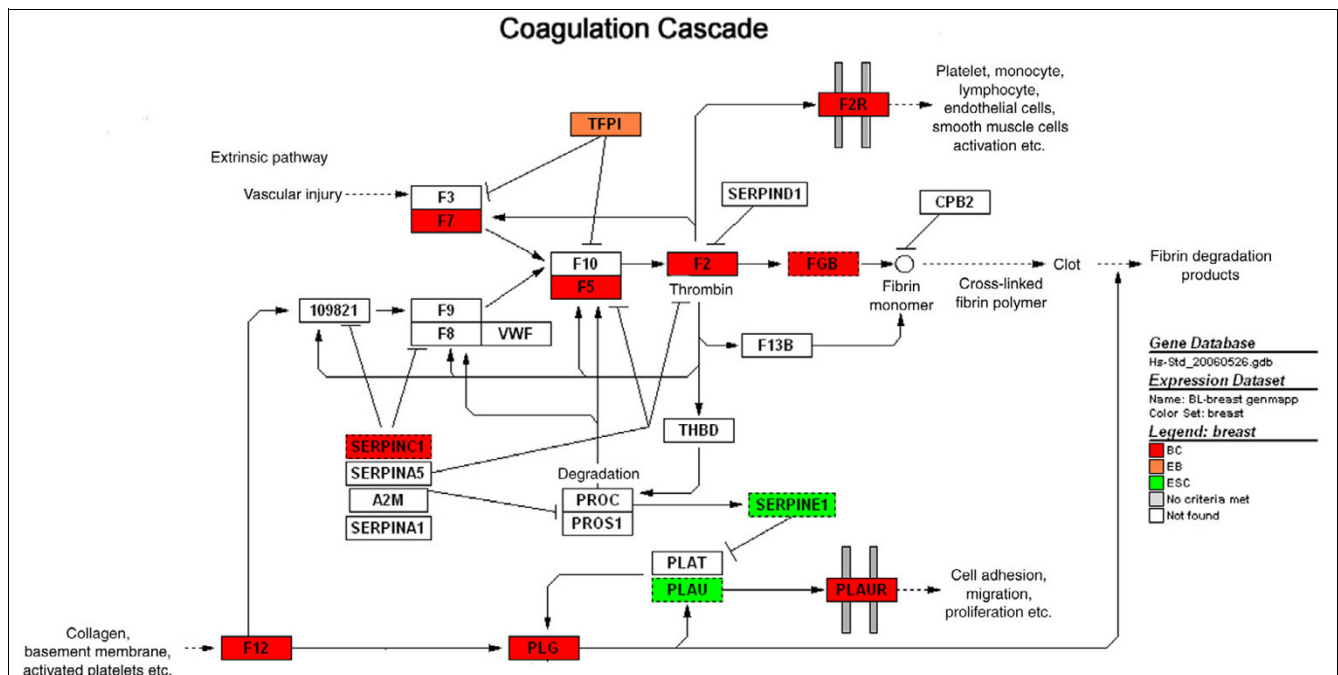
*receptor 1e/6, nestin, neurogenic differentiation 4, neurologin 2, myelin basic protein, peripheral myelin protein 22).*

Of particular note is the absence of leukocytic processes, such as response to stimulus and defense response identified in the level two analyses. Thus, this comparison allowed for the masking of the 'lymphocytic' signature and, thus, the identification of other endothelial and blood vessel development genes (*vWF, bradykinin receptor b1, and thrombin receptor*). When this data set was analyzed using GenMapp, more endothelial genes were mapped to the coagulation cascade pathway (*vWF, bradykinin receptor b1, thrombin receptor*; data not shown).

**Genes enriched in BCs relative to endothelial cells**

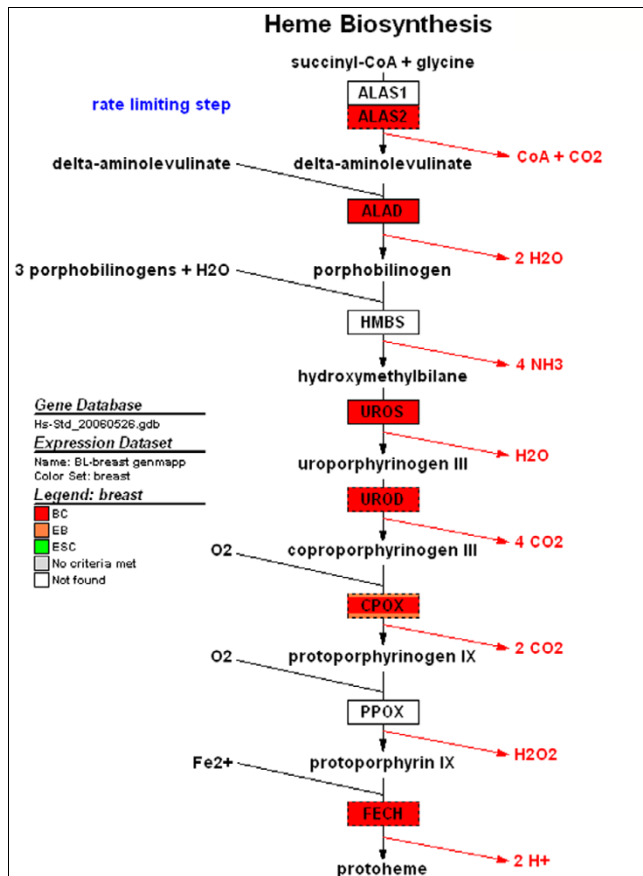
We identified 904 genes that were up-regulated in BCs relative to prostate-derived endothelium after filtering out those

genes that are enriched in ESCs relative to endothelium (Additional data file 13). Comparing BCs to endothelial cells identified fewer genes when compared to other comparisons of BCs in level II and III analyses, and, thus, identified fewer GO terms. However, it did identify more erythrocytic processes (nine in total; Table 5) than the other comparisons. Another predominant theme in this data set was development (Table 5). By comparing BCs to a more mature yet similar cell type (adult endothelium), we were able to mask the endothelial signature, thus identifying predominantly development genes (indicating stem/progenitor type signature), such as *caudal type homeobox transcription factor 2 (CDX2), delta-like homolog (DLK1), lamin A/C, secreted frizzled-related protein 5 (SFRP5), patched (PTCH), dishevelled 2 (DVL2), and even-skipped homeobox homolog 1 (EVX1).*



**Figure 2** GenMAPP of complement and coagulation. Genes that are up-regulated in BCs (red), EBs (orange), and ESCs (green) compared to breast epithelia as baseline were mapped onto a pre-existing pathway. This pathway contains genes that are mostly up-regulated in BCs relative to breast epithelia.





**Figure 3**  
GenMAPP of heme biosynthesis: Genes that are up-regulated in BCs (red), EBs (orange), and ESCs (green) compared to breast epithelia as baseline were mapped onto a heme biosynthesis pathway. This pathway contains genes that are up-regulated in BCs relative to breast epithelia.

*Genes enriched in BCs relative to stromal cells*

The BCs were then compared to prostate-derived stromal fibromuscular (CD49a immunoselected) tissue. This comparison identified the most number of genes (3,277 genes; Additional data file 14), and had the most diverse GO terms (lymphocytic, developmental, erythrocytic, coagulation, synapses and neurogenesis, and heart development; Table 6). This data set contained lymphocytic processes according to their GO (response to stimulus, defense response) and numerous lymphocytic markers (CD 6/38/41/43/48/55/61/71/84/244, immunoglobulin genes heavy constant  $\gamma$ , constant  $\kappa$ , constant  $\lambda$ 1, and CD 158A/B/D/F/H (killer cell immunoglobulin-like receptor)). This comparison identified genes involved in hemangioblast differentiation (TAL1/SCL, LMO2, RUNX1), endothelial genes (neuropilin 1 and 2), and coagulation genes (fibrinogen  $\alpha$  and  $\beta$  chain, coagulation factor 5, plasminogen, but not KDR, FLT1, CD4, PECAM, VE-Cadherin, vWF). Although EASE analysis for both epithelial and stromal comparisons identified similar heart GO terms, the stromal comparison identified different heart development genes, such as MEF2C, aortic preferentially

expressed (APEG1), POU6F1, TBX1, and ryanodine receptor 2 (cardiac).

When these genes were clustered for pathways analysis with DAVID, we identified Nfat, hypertrophy of the heart and Alk in cardiac myocytes as a statistically over-represented pathway (data not shown). GenMapp identified a similar pathway involved in myometrial contraction and calcium regulation in the cardiac cell (data not shown). These genetic signatures from this analysis would indicate the presence of progenitors (indicated by the number of developmental genes) of erythrocytes, leukocytes, neurons/neuronal-muscular junctions, and cardiomyocytes. Thus, comparing BCs to stromal cells masked a connective tissue-like signature, allowing for the identification of tissue-specific processes.

*Ingenuity analysis*

To identify signaling pathways involved in hemangioblast differentiation, each of the data sets was analyzed by Ingenuity. Ingenuity is a program that converts large data sets into networks containing direct and indirect relationships between genes based on known interactions in the literature. Genetic networks were created using the EB and BC data sets. The EB data set from the level 2 analysis contained a network of genes (VEGF, GATA4, BMP4) that are interconnected and involved in blood vessel development (VEGF), heart development (GATA4), and cellular development (BMP4) (Figure 4a). For example, Bmp-4 has been shown to promote blood vessel development by increasing VEGF production [34] and VEGF induces or binds to KDR FLT1, NRP1, and NRP2 [35-38]. This network suggests that BMP4 inhibits cardiac development by increasing HEY1, a transcriptional repressor of GATA4 and 6 (Figure 5a) [39,40], and inhibits heart development by inducing DKK1 (dickkopf homolog 1) [41], which then inhibits WNT11 mRNA expression, GATA4, and NKX2-5 [42-44]. In conclusion, this network suggests that BMP4 induces blood vessel development through VEGF signaling and inhibits cardiac differentiation through HEY1 and DKK1.

We then looked for these and other signaling pathways in the level 2 BC data set. Here we identified genes involved in cardiovascular development (SHH [45], RAR-B, TBX5 [46], WNT11 [43]) acting through GATA4 (Figure 4b). However, unlike the EB data set, we did not identify the cardiac repressor HEY1, VEGF and BMP4 in the BC data set. Instead, the BC network contained cardiac and skeletal genes, such as HAND2 and ANKRD1 [47-50], and HIF3A, an inhibitor of VEGF expression [51] (Figure 5b). Thus, these networks demonstrate that when EBs differentiate into BCs, we see some angiogenic and some cardiac pathways.

Another signaling network we identified as differentially expressed between EBs and BCs is the network containing GATA2. GATA2 has been shown to play a vital role in hemangioblast development [26] by up-regulating BMP4, KDR, and TAL1/SCL expression. In the EB data set, the GATA-2 net-



**Table 4****Gene ontologies for up-regulated processes in BCs versus leukocytes**

Gene category	List hits	EASE score
Development (developmental)	178	6.39E-09
Morphogenesis (developmental)	116	1.98E-08
Organogenesis (developmental)	105	5.04E-08
Cellular process	482	6.48E-08
Cell communication	257	4.78E-07
Cell adhesion	68	4.05E-06
Skeletal development (developmental)	25	4.44E-06
Cell-cell signaling	64	4.86E-06
Ion transport	62	8.10E-05
Circulation	20	9.47E-05
Metal ion transport	38	2.53E-04
Cartilage condensation	6	3.56E-04
Cation transport	46	5.97E-04
G-protein coupled receptor protein signaling pathway	61	8.03E-04
Regulation of blood pressure	9	9.06E-04
Neurogenesis (neuronal)	46	1.17E-03
Second-messenger-mediated signaling	19	1.32E-03
Cell surface receptor linked signal transduction	93	1.52E-03
Bone remodeling	9	2.01E-03
Ossification	9	2.01E-03
Cell homeostasis	15	2.07E-03
Cell ion homeostasis	14	2.27E-03
Ion homeostasis	14	2.27E-03
Transition metal ion transport	9	2.40E-03
Transport	151	2.97E-03
Homeostasis	15	3.10E-03
Transition metal ion homeostasis	7	3.37E-03
Receptor mediated endocytosis	7	4.20E-03
Metal ion homeostasis	11	4.48E-03
Homophilic cell adhesion	15	5.91E-03
Di-, tri-valent inorganic cation homeostasis	10	6.86E-03
Gas transport (erythrocytic)	5	7.51E-03
Oxygen transport (erythrocytic)	5	7.51E-03
Cation homeostasis	12	8.76E-03
Angiogenesis (blood vessel development)	8	1.10E-02
Cyclic-nucleotide-mediated signaling	15	1.13E-02
Transmission of nerve impulse (neurophysiology)	27	1.16E-02
Monovalent inorganic cation transport	28	1.19E-02
Di-, tri-valent inorganic cation transport	14	1.42E-02
Enzyme linked receptor protein signaling pathway	21	1.42E-02
Synaptic transmission (neurophysiology)	26	1.42E-02
Blood vessel development (blood vessel development)	8	1.44E-02
Iron ion homeostasis	5	1.64E-02
Blood coagulation (coagulation)	13	1.80E-02
Tyrosine kinase signaling pathway	15	1.99E-02
Signal transduction	177	2.18E-02
Potassium ion transport	17	2.47E-02

**Table 4** (Continued)

**Gene ontologies for up-regulated processes in BCs versus leukocytes**

Hemostasis	13	2.77E-02
Pregnancy	8	2.89E-02
Axon guidance (neurophysiology)	6	4.00E-02

When these genes were clustered based on their GO, we identified processes involved in development, nervous system development, blood vessel development and angiogenesis, and erythrocytes.

work contained *EPOR*, *TAL1/SCL*, *TCF3* and *PITX2* (Figure 6a). *Pitx2* is a homeobox gene involved in regulating the balance between proliferation and differentiation of progenitor cells [52] and is highly expressed in EBs (FC 24x) and is absent in BCs. *PITX2* is not only rapidly down-regulated upon hematopoietic stem cell differentiation [52], but may also promote hemangioblast differentiation by inducing *GATA2* expression [53]. In the BC data set, the *GATA2* net-

work contained the hemangioblastic and hematopoietic genes *TAL1/SCL* and *LMO2* [25,27,54], *FOG/Zfp1*[55], *CD41/GPIIB/Igta2B* [56] and *GATA1* [57] (Figure 6b).

A predominant network we identified in all three data sets of the level III analysis involved *GATA1*. *GATA1* is a globin transcription factor and is present in all BC but not EB data sets. For example, we see *GATA1* interacting with other nuclear

**Table 5**

**Gene ontologies for up-regulated processes in BCs versus endothelial cells**

Gene category	List hits	EASE score
Development (developmental)	84	2.10E-05
Oxygen transport (erythrocytic)	6	2.14E-05
Gas transport (erythrocytic)	6	2.14E-05
Heme biosynthesis (erythrocytic)	5	2.88E-04
Heme metabolism (erythrocytic)	5	5.56E-04
Morphogenesis (developmental)	51	6.08E-04
Organogenesis (developmental)	46	9.55E-04
Porphyrin biosynthesis (erythrocytic)	5	9.66E-04
Porphyrin metabolism (erythrocytic)	5	1.55E-03
Transport	77	1.89E-03
Pigment biosynthesis (erythrocytic)	5	2.83E-03
Pigment metabolism (erythrocytic)	5	4.70E-03
Pigmentation (erythrocytic)	5	5.46E-03
Hemostasis	9	9.09E-03
Learning and/or memory	4	1.67E-02
Skeletal development (developmental)	10	1.69E-02
Response to chemical substance	14	1.87E-02
Blood coagulation (coagulation)	8	2.05E-02
Hearing	7	2.16E-02
Perception of sound	7	2.29E-02
Synaptic transmission (neurophysiology)	14	2.69E-02
Ion transport	26	3.17E-02
Cell-cell signaling	25	3.20E-02
Cellular process	204	3.21E-02
Transmission of nerve impulse (neurophysiology)	14	3.36E-02
Response to abiotic stimulus (lymphocytic)	25	3.88E-02
Cartilage condensation	3	3.89E-02
Circulation	8	4.64E-02
Coenzyme and prosthetic group metabolism	9	4.98E-02

This comparison did not identify as many genes because of their similar origin and thus fewer gene ontologies. However, it did identify more erythrocytic processes than the other comparisons.

**Table 6****Gene ontologies for up-regulated processes in BCs versus stromal cells**

Gene category	List hits	EASE score
Cell communication	328	7.47E-07
Cellular process	617	1.25E-06
Cell-cell signaling	79	4.71E-06
Response to external stimulus (lymphocytic)	160	1.38E-05
Ion transport	74	4.07E-04
Response to abiotic stimulus (lymphocytic)	72	4.61E-04
Transport	199	8.97E-04
Cation transport	56	1.08E-03
Defense response (lymphocytic)	95	1.09E-03
Metal ion transport	44	1.30E-03
Response to biotic stimulus (lymphocytic)	101	1.35E-03
Development (developmental)	192	1.55E-03
Intracellular receptor-mediated signaling pathway	8	1.72E-03
Organogenesis (developmental)	108	1.98E-03
Cell adhesion	72	2.17E-03
Morphogenesis (developmental)	118	2.98E-03
Synaptic transmission (neurophysiology)	35	3.04E-03
Signal transduction	238	4.29E-03
Transmission of nerve impulse (neurophysiology)	35	5.07E-03
Regulation of blood pressure	9	5.42E-03
Monovalent inorganic cation transport	36	6.29E-03
Steroid hormone receptor signaling pathway	7	6.65E-03
Homeostasis	17	6.77E-03
Cell surface receptor linked signal transduction	113	8.56E-03
Perception of abiotic stimulus (lymphocytic)	39	1.09E-02
Hemostasis	17	1.10E-02
Second-messenger-mediated signaling	20	1.18E-02
Potassium ion transport	22	1.19E-02
Cyclic-nucleotide-mediated signaling	18	1.24E-02
Blood coagulation (coagulation)	16	1.39E-02
Response to chemical substance (lymphocytic)	31	1.45E-02
Heme biosynthesis (erythrocytic)	5	1.48E-02
Gas transport (erythrocytic)	5	2.00E-02
Oxygen transport (erythrocytic)	5	2.00E-02
Neurogenesis (neuronal)	51	2.02E-02
Cell homeostasis	15	2.31E-02
Perception of external stimulus (lymphocytic)	41	2.41E-02
Estrogen receptor signaling pathway	4	2.48E-02
Heme metabolism (erythrocytic)	5	2.61E-02
Homophilic cell adhesion	16	2.74E-02
Pigment biosynthesis (erythrocytic)	6	2.84E-02
Circulation (erythrocytic)	17	2.94E-02
Olfaction	9	3.02E-02
Response to pest/pathogen/parasite (lymphocytic)	52	3.19E-02
Acute-phase response (coagulation)	6	3.43E-02
Skeletal development	19	3.72E-02
Sensory perception	37	3.86E-02

**Table 6** (Continued)

**Gene ontologies for up-regulated processes in BCs versus stromal cells**

Disaccharide metabolism	3	3.86E-02
Enzyme linked receptor protein signaling pathway	24	3.96E-02
Heart development (heart development)	6	4.09E-02

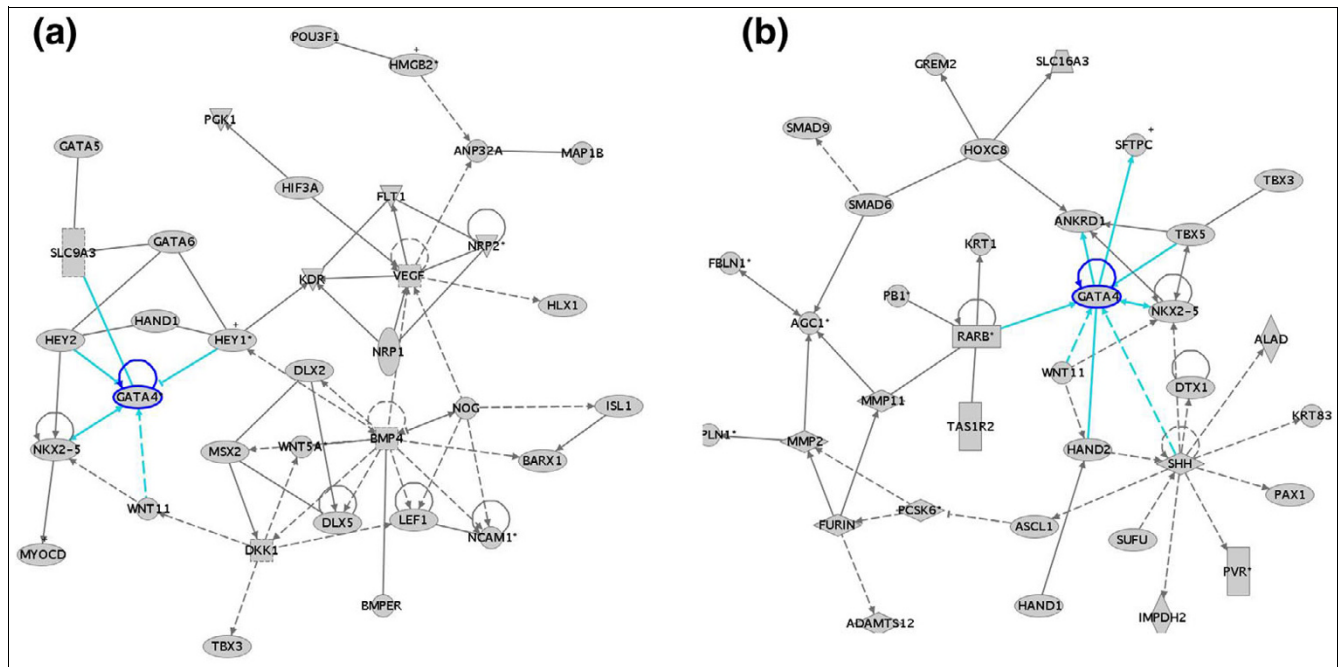
This comparison identified the most number of genes and had the most diverse gene ontologies.

genes, such as *TAL1/SCL*, *LMO2*, and *KLF1*, which induce erythropoietic genes such as the hemoglobin family (*HBG1*, *HBG2*, *HBE*, *HBB*, and *HBZ*), heme synthesis (*ALAS2*), and genes expressed on the cell surface of RBC (*Ankyrin 1*, *Rh blood group*, *glycophorin A*, *erythrocyte membrane protein band 4.2*) (Additional data file 2).

**Discussion**

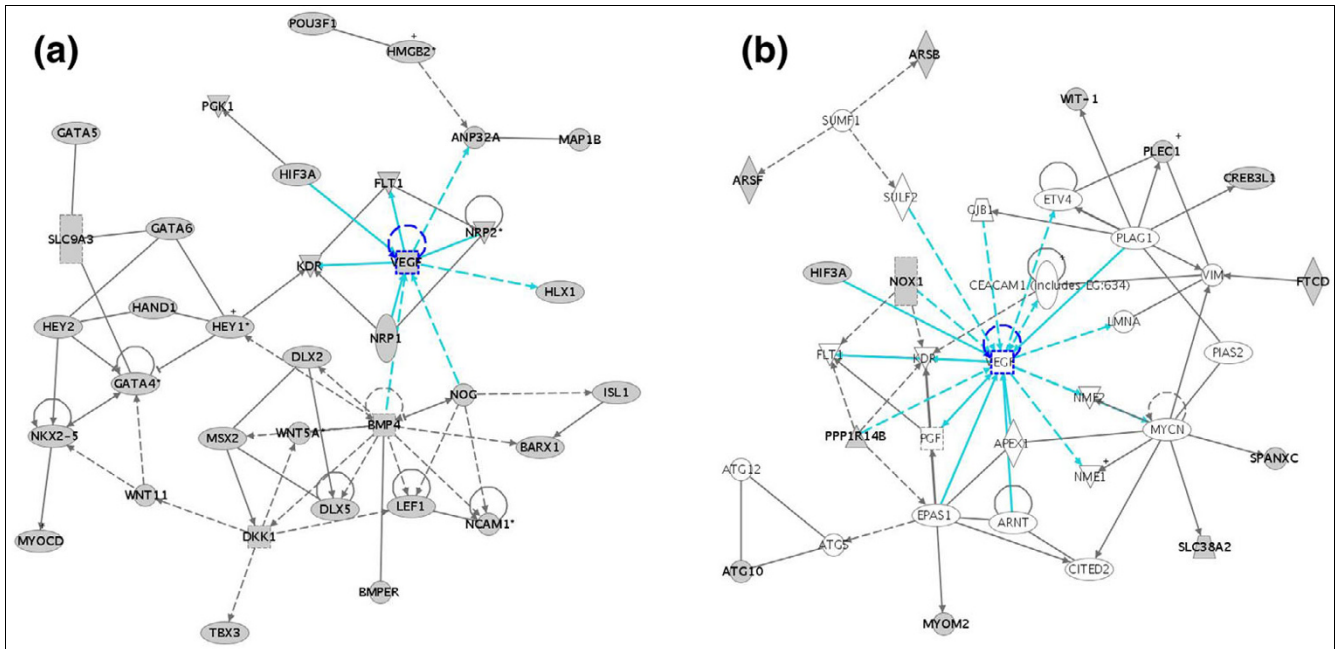
To delineate the mechanisms involved in the development of hemangioblasts from murine ES cells, Lugus *et al.* [26] performed global gene expression profiling of mouse Flk1+ cells that can form BL-CFCs. However, no such analysis has been reported for the human counterpart due to availability of such cells. In the present study, we carried out a large scale tran-

scriptional l analysis to profile undifferentiated hESCs, early stage EBs, and BCs as an initial effort to understand the differentiation program of hESCs towards hemangioblasts. The availability of a well-annotated genome database for hESCs, EBs and BCs will provide a foundation that allows one to map and identify genes involved in human hemangioblast development, and to optimize conditions for efficient generation of hemangioblasts from hESCs. Our studies in general are consistent with previous reports that a cluster of genes (*OCT4*, *NANOG*) expressed at significantly high levels in hESCs were down-regulated in early stage EBs and BCs [23,58-63], whereas some genes restricted to endothelial and hematopoietic cells (neuropilins, *LMO2*, *GATA1 and 2*, *TAL1/SCL* and globins) were up-regulated dramatically in BCs and even in early stage EBs [2,23,26,64].

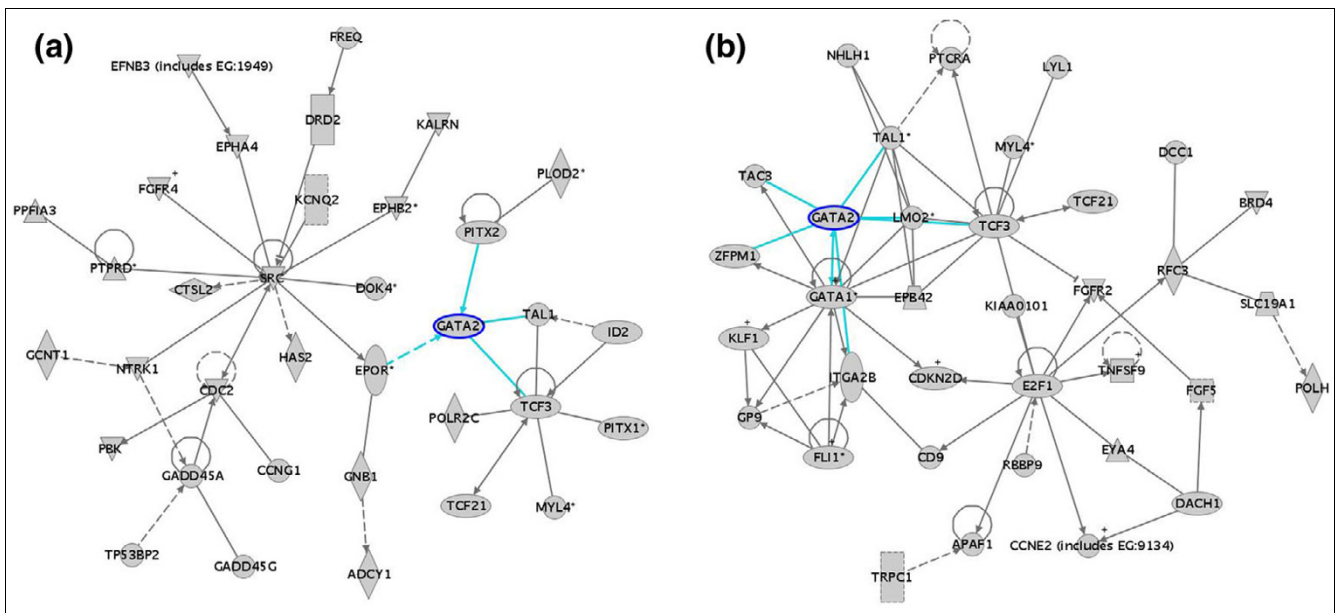


**Figure 4**

Ingenuity pathway analysis shows a network of genes expressed in EB and BC data sets from the level II analysis. This network contains nodes (genes/gene products) and edges (relationships between the nodes). The shaded genes, known as focus genes, were identified by microarrays and are the starting point to generate the network. The asterisks indicate that duplicates were identified in each dataset. The EB data set contains a network of genes (*VEGF*, *GATA4*, *BMP4*) that are interconnected and involved in blood vessel development (*VEGF*), heart development (*GATA4*), and cellular development (*BMP4*). The BC data set identified a network of genes involved in cardiovascular development (*SHH*, *RARB*, *TBX5*, *WNT11*) acting through *GATA4*. The intensity of the node color indicates the degree of expression. Nodes are displayed using various shapes that represent the functional class of the gene product (diamond-enzymes, ovals-transcription factors, triangles-kinase, circles-others). A solid line indicates a direct interaction while a dashed line indicates an indirect interaction. A line without an arrowhead indicates binding and a plus sign indicates that other networks contain this gene product.



**Figure 5**  
 Ingenuity pathway analysis identified a network of genes expressed in EB and BC data sets associated with VEGF. In the EB data set, VEGF is associated with 35 'focus' genes, including KDR, FLT1, NRP1, and NRP2. In the BC data set, 13 focus genes were associated with VEGF, even though it is not present in the data set. The intensity of the node color indicates the degree of expression. Nodes are displayed using various shapes that represent the functional class of the gene product (diamond-enzymes, ovals-transcription factors, triangles-kinase, circles-others). A solid line indicates a direct interaction while a dashed line indicates an indirect interaction. A line without an arrowhead indicates binding and a plus sign indicates that other networks contain this gene product.



**Figure 6**  
 Ingenuity pathway analysis of a network of genes expressed in EB and BC data sets associated with GATA2. In the EB data set, the GATA2 network contained EPOR, TAL1, TCF3 and PITX2. In the BC data set, the GATA2 network contained the hemangioblastic and hematopoietic genes TAL1, LMO2, FOG/ZFPM1, IGTA2B/CD41/GPIIb, and GATA1. The intensity of the node color indicates the degree of expression. Nodes are displayed using various shapes that represent the functional class of the gene product (diamond-enzymes, ovals-transcription factors, triangles-kinase, circles-others). A solid line indicates a direct interaction while a dashed line indicates an indirect interaction. A line without an arrowhead indicates binding and a plus sign indicates that other networks contain this gene product.

Our previous study has shown that BCs contain a mixed progenitor population of cells capable of forming hemangioblasts, and hematopoietic and endothelial cells [21]. To assess the heterogeneous populations in BCs, comparisons to publicly available data sets were performed *in silico*. Biologically relevant comparisons allowed us to mask particular genetic signatures within the heterogeneous population, allowing us to identify others, such as myogenic, vasculogenic, and hematopoietic progenitors within BCs. Our level I analysis compared hESCs to their differentiated progeny, providing a kinetic-like relationship of gene expression. When comparing the hESCs to BCs, many of the down-regulated genes were involved in development, cell differentiation, and morphogenesis. The predominant up-regulated genes identified in this level I analysis were those involved in the development of hemangioblasts and primitive erythroblasts. These genes include those encoding the transcription factors *TAL1/SCL*, *LMO2* and *GATA1* and *2*, heme synthesis genes, genes encoding erythroblast membrane proteins, and embryonic and fetal globin genes, but very low levels of adult globin gene, suggesting the yolk sac primitive status of BCs. This observation is consistent with findings obtained from both mouse and zebrafish models, in which hemangioblasts were first developed from primitive streaks and yolk sacs [65,66]. The results may also represent the fact that BC expansion medium contains several hematopoietic cytokines, which could push the developmental pathway towards the hematopoietic lineage, although BCs harvested at day 6 retain the potential of differentiating into endothelial cells under appropriate conditions [21]. Optimization of BC expansion conditions would, therefore, be valuable to keep these cells bipotential.

When applying this technique of multiple tissue type comparisons, some of the genes that are identified as 'up-regulated' in our tissue of interest could be 'under expressed' in the reference tissue. We controlled for this by filtering for those 'under expressed' genes with a comparison to a genotypically similar but different tissue type. For example, in our level II analysis, we identified those genes that are up-regulated in BCs relative to breast epithelial cells (3,700 genes) and then removed those genes that were up-regulated in hESCs relative to epithelial cells (2,735 genes). This removed 965 genes that could be thought of as being up-regulated in both BCs and hESCs relative to breast epithelium or as genes that are under-expressed in breast epithelium. GO analysis of this data set identified cell cycle genes as the predominant theme. This observation correlates with their biology because the BCs and hESCs are actively dividing cells *in vitro*, while the breast epithelium comprises relatively senescent cells freshly isolated from *in vivo*. Most importantly, we did not identify any tissue specific processes that we would expect to find in BCs.

GO analysis of the genes up-regulated in BCs with respect to epithelial cells after filtering out those that are up-regulated in hESCs relative to breast epithelial cells identified biological

themes involved in erythropoiesis, as in the level I analysis (heme and porphyrin biosynthesis), but also the angiogenic components of coagulation, and synapses. This analysis identified genetic signatures representative of not just erythrocytes, as in the above level I analysis, but also the other cellular components of the BC population, such as muscle, cardiac, and hematopoietic cells and hemangioblasts. This *in silico* comparison to a biologically distinct reference tissue and filtering allows one to identify statistically significant genes that might otherwise be missed within a heterogeneous population.

In the level III analysis, biologically relevant comparisons were made *in silico* to adjust for particular cell types represented in the BC population. As in the level II analysis, this analysis also identified genetic signatures of the erythrocytic population in addition to other cellular components of BCs. When BCs were compared to leukocytes *in silico*, we identified a genetic signature representative of vasculogenesis, endothelia, neurons/synapses, hemangioblasts, and erythrocytes, with the relative absence of leukocyte genes. When BCs were compared to endothelia *in silico*, we identified a signature of erythrocytic and developmental genes with the relative absence of the vasculogenic signature. When BCs were compared to stromal cells *in silico*, we identified processes involved in hematopoiesis, synapses, angiogenesis/endothelia, development and more genes involved in cardiomyogenesis. Although EASE analysis for both epithelial and stromal comparisons identified similar heart GO terms, the stromal comparison identified different heart development genes. We believe that the epithelial comparison revealed more of the mesodermal aspect of BCs while the stromal comparison masked the mesodermal components and revealed more cardiomyocytic genes.

It has been shown that murine BL-CFCs are able to differentiate into hematopoietic, endothelial and smooth muscle cells, but failed to give rise to cardiomyocytes [67]. Molecular analyses showed that BL-CFCs expressed genes indicative of the hematopoietic and endothelial lineages, but not cardiomyocytes [26]. Kattman *et al.* [68] recently identified a cardiovascular progenitor with the same phenotype as the BL-CFC progenitor, brachyury+ and Flk-1+ cells from day 4.25 EBs (one day later than the BL-CFC progenitor) derived from mouse ESCs. These studies strongly suggest that BL-CFCs and cardiomyocytes, at least in the mouse ESC system, are derived from two different progenitors. Two other studies demonstrate the existence of multipotential progenitors for cardiomyocytes and muscle cells, but with different surface markers [69,70]. In the present study, several genes restricted to cardiomyocytes or their progenitor were detected with relatively high levels of expression in BCs. This observation suggests that human BCs may possess the potential to differentiate into cardiomyocytes, which will need further investigation. Alternatively, the purified BCs from multiple colonies could contain dissimilar blast clones origi-

nated from different differentiation stages, some of which may have the potential to develop into cardiomyocytes, as demonstrated recently by three groups in the mouse ESC system [68-70]. Simultaneous isolation and characterization of functionally distinct colonies and analysis of gene expression in these colonies might serve to determine whether human BCs possess the potential to differentiate into hematopoietic, endothelial and cardiomyocyte lineages.

## Conclusion

The identification and characterization of cell types within a heterogeneous population will be of increasing importance in stem cell research since differentiation protocols require the formation of progenitors through a multi-stage approach. Our previous study has shown that BCs contain a mixed progenitor population of cells capable of forming hemangioblasts, and hematopoietic and endothelial cells [21]. To assess the heterogeneous populations in BCs, comparisons to publicly available data sets were performed *in silico*. Biologically relevant comparisons allowed us to mask particular genetic signatures within the heterogeneous population, allowing us to identify others, such as myogenic, vasculogenic, and hematopoietic progenitors, within BCs. The significance of this microarray study is in its ability to assess and identify cellular populations within a heterogeneous population through biologically relevant *in silico* comparisons of publicly available data sets. In conclusion, multiple *in silico* comparisons were necessary to characterize tissue-specific genetic signatures within a heterogeneous hemangioblast population.

## Materials and methods

### hESC culture and BC growth

Culture of hESCs and growth of BCs were as reported previously [21]. In brief, undifferentiated hESCs (H1, H9) were cultured with inactivated mouse embryonic fibroblast cells in complete hESC media until they reached 80% confluence. Undifferentiated hESCs were dissociated by 0.05% trypsin-0.53 mM EDTA (Invitrogen, Carlsbad, CA, USA) for 2-5 minutes and collected by centrifugation at 1,000 rpm for 5 minutes. To induce hemangioblast precursor (mesoderm) formation, hESCs ( $2-5 \times 10^5$  cells/ml) were plated on ultra-low dishes (Corning, Corning, NY, USA) in Stemline II media with the addition of BMP4 and VEGF<sub>165</sub> (50 ng/ml; R&D Systems (Minneapolis, MN, USA)) and cultured in 5% CO<sub>2</sub>. Forty eight hours later, half the media was removed and fresh medium was added with the same final concentrations of BMP4 and VEGF, plus SCF, Tpo and FLT3 ligand (20 ng/ml; R&D Systems), and PTD-HoxB4 (1.5 µg/ml) to expand out BCs and their precursor. After 3.5 days, EBs were collected and dissociated by 0.05% trypsin-0.53 mM EDTA (Invitrogen) for 2-5 minutes, and a single cell suspension was prepared by passing through a 22G needle 3-5 times. To expand BCs, a single cell suspension derived from differentiation of  $2-5 \times 10^5$  hESCs were mixed with 2 ml hemangioblast expansion

medium plated on ultra-low dishes and incubated at 37°C in 5% CO<sub>2</sub> for 6 days, and BCs were then collected and subjected to RNA isolation.

### Affymetrix GeneChip analysis

Total RNA was isolated from purified BCs, day 3.5 EBs and undifferentiated ESCs (from two hESC lines, H1 and H9) using the Qiagen (Valencia, CA, USA) RNeasy kit and amplified as previously described [21]. A total of six microarrays were performed (two biological replicates per time point using different hESC lines). Fragmented antisense cRNA was used for hybridizing with human U133 Plus 2.0 arrays (Affymetrix, Inc. Santa Clara, CA, USA) at the Core Genomic Facility of University of Massachusetts. The validation of differentially expressed genes was confirmed by immunocytochemistry in our previous studies [21] and by semi-quantitative RT-PCR analyses (Figure 1). The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (GEO) [71] and are accessible through GEO series accession numbers GSE8884 and GSE9196, in accordance with MIAME standards. The demographics of the publicly available GeneChip data sets are breast-derived epithelial cells ( $n = 7$ ), leukocytes ( $n = 6$ ), prostate-derived endothelial cells ( $n = 5$ ), and prostate-derived stromal cells ( $n = 5$ ). These samples were chosen based on their homogeneity (cells were immunoselected or enriched) and the number of replicates. These data sets were downloaded from the NCBI's GEO and are accessible through accession numbers GSE9086 (breast-derived epithelial cells), GSE9091 (leukocytes), GSE9090 (stromal), and GSE9089 (endothelia).

### Data analysis

Raw CEL files were provided by the Core Genome Facility of the University of Massachusetts and were then analyzed with a software package AffylmGUI (Affymetrix LIMMA, Linear Models for Microarray Data, Graphical User Interfaces) [72,73]. Within AffylmGUI, gene expression values were summarized with RMA. RMA adjusts for background noise, performs a quantile normalization, transforms the data into log base 2, and then summarizes the multiple probes into one intensity [74-76]. Quantification of relative differences in gene expression among the groups of interest was accomplished using AffylmGUI, the sister package of limmaGUI [72,73]. AffylmGUI reads the raw Affymetrix CEL files directly, summarizes the gene expression values using RMA, and then uses LIMMA to identify statistically significant differences in gene expression [77]. LIMMA fits a linear model for every gene (like ANOVA or multiple regression analysis), and adjusts *P* values for multiple testings [77]. Differentially expressed genes were identified with a *B* statistic >0. The *B* statistic, also known as a likelihood of odds (LOD) score is a moderated *t*-statistic with posterior residual standard deviations. Subsequent analyses were performed in Microsoft Excel and Microsoft Access.



## EASE

The application Expression Analysis Systematic Explorer (EASE) was used to determine biologically relevant themes in a list of differentially expressed genes. EASE identifies over-represented biological themes in terms of their GO [78]. GO was developed to provide consistent descriptions of genes in terms of biological processes and molecular function. Genes with a B value >0 are incorporated into EASE where each gene is matched to all possible GOs. The results of this analysis are compared to all possible GOs for all genes on the microarray platform and calculates a *P* value, based on a conservative variant of the Fischer's exact probability test. We selected those pathways/processes with *P* < 0.05.

## DAVID

DAVID (Database for Annotation, Visualization and Integrated Discovery) provides tools and statistical methods for uncovering enriched processes and pathways within diverse and disparate gene lists [79]. DAVID also identifies over-represented biological themes in terms of their GO [78] and provides tools to visualize the distribution of genes on BioCarta and KEGG pathway maps.

## GenMapp

GenMapp is designed to visualize gene expression data on maps representing biological pathways and grouping of genes. It consists of hundreds of pre-made pathway maps and is used to identify pathway level changes amongst multiple data sets [80]. In this program, data sets are assigned a color corresponding to cell type and the direction of gene expression changes.

## Ingenuity

Networks were constructed using Ingenuity Pathways Analysis (Ingenuity® Systems, Redwood City, CA, USA). A data set

containing gene identifiers of genes with a B > 0 was uploaded into the applications. These genes, called focus genes, were overlaid onto a global molecular network developed from information contained in the Ingenuity Pathway Knowledge Base. Networks of these focus genes were then algorithmically generated based on their connectivity. Each network is a graphical representation of the molecular relationships between genes/gene products. Genes or gene products are represented as nodes, and the biological relationship between two nodes is represented as an edge (line). All edges are supported by at least 1 reference from the literature stored in the Ingenuity Pathways Knowledge Base. The intensity of the node color indicates the degree of expression. Nodes are displayed using various shapes that represent the functional class of the gene product (diamond-enzymes, ovals-transcription factors, triangles-kinase, circles-others). A solid line indicates a direct interaction while a dashed line indicates an indirect interaction. A line without an arrowhead indicates binding and a plus sign indicates that other networks contain this gene product.

## RNA isolation and gene expression quantification by semi-quantitative PCR

Total RNA was isolated from hESCs, day 3 EBs and heman-gioblasts (BCs) using an RNAeasy Mini Kit (Qiagen) following the procedure recommended by the supplier with DNase I digestion, which eliminates the contamination of genomic DNA. RNA was subjected to first-strand cDNA synthesis with SMART II and CDS primers (Clontech, Mountain View, CA, USA), using Superscript II reverse transcriptase (Invitrogen), and cDNA pools were constructed using the SMART cDNA synthesis kit (Clontech) as described previously [2,81]. Complementary DNA pools generated by the SMART procedure have been shown to preserve the relative abundance relationship of the original mRNA populations [82-84]. The

**Table 7**

### Primer sequences

Gene	Primer	Sequence	Annealing T <sub>m</sub> (C)	MgCl <sub>2</sub> conc. (μM)	Product size (bp)
HPRT	Sense	5'-CTTGCGACCTTGACCATCTTTGGA-3'	58	1.5	467
	Antisense	5'-GGCGTCGTGATTAGTGATGATGAACC-3'			
GATA-2	Sense	5'-TATGTGCCGGCGGCTGCCACGACTACA-3'	60	1.5	280
	Antisense	5'-GGCTCTTCTGGCGGCCGACAGTCTT-3'			
SCL	Sense	5'-GAAGTGCTCCCCTCTGAAAGTT-3'	58	1.5	319
	Antisense	5'-GGCTATCTCTCCTCTGACCTCG-3'			
β-Cluster globin genes	Sense	5'-GTYTACCCHTGGACCCAGA-3'	56	2	190
	Antisense	5'-GCAGCTTGTCACAGTGCA-3'			
OCT4	Sense	5'-GAAGGTATTGAGCCAAACGAC-3'	55	2	315
	Antisense	5'-GTTACAGAACCACACTCGGA-3'			
Nanog	Sense	5'-TGCAAATGCTTCTGCTGAGAT-3'	55	2	285
	Antisense	5'-GTTACAGGATGTTGGAGAGTTC-3'			
Rex-1	Sense	5'-TGACAGGCAAGAAGCTTCCG-3'	55	2	350
	Antisense	5'-GCGTACGCAAATTAAGTCCAGA-3'			

DNA templates in cDNA pools of human ES cells, EBs and hemangioblasts were adjusted to equal amounts based on the relative expression level of the hypoxanthine phosphoribosyl-transferase gene (HPRT) and gene expression quantification by semi-quantitative PCR was performed as described previously [2,81]. The sense and anti-sense primer sequences, and the corresponding cDNA PCR product sizes, are shown in Table 7. The conditions for PCR amplification were as described with annealing temperatures and concentrations of MgCl<sub>2</sub> for each specific gene as shown below. PCR products (10 µl) were separated on 1.5 to 2.0% agarose gel and visualized by ethidium bromide staining. The relative expression levels in cDNA from hESCs, EBs and hemangioblasts were estimated visually.

### Direct and indirect comparison of microarray data of differentially expressed genes

The direct analysis (level I) consists of making comparisons between ESCs, EBs, and BCs. Since there are two possible comparisons for each cell type, for example, ESCs relative to EBs or ESCs relative to BCs, fold changes were determined by comparing each to a cell type that did not detect the gene in question. Fold change levels for *Oct-4* and *Nanog* were determined by comparing ESCs to BCs and EBs to BCs. Fold change levels for *Gata-2* were determined by comparing EBs to ESCs and BCs to ESCs. Fold change levels for *SCL/Tal1* were determined by comparing BCs to ESCs. Fold change levels for  $\gamma/\epsilon$ -globin were determined by comparing BCs to EBs. For the indirect (level II) analysis, fold changes were simply determined by comparing each cell type to breast epithelia. If more than one probe-set was identified as differentially expressed, fold changes were averaged.

### Abbreviations

BC, blast cell; BL-CFC, blast colony forming cell; DAVID, Database for Annotation, Visualization and Integrated Discovery; EASE, Expression Analysis Systematic Explorer; EB, embryoid bodie; ESC, embryonic stem cell; GEO, Gene Expression Omnibus; GO, gene ontology; h, human; RMA, robust multi-chip average.

### Authors' contributions

SJL and QF designed and performed the cell culture, differentiation and RNA isolation. JAH and JDH performed the microarray analysis. This study was conceived by JDH, SJL, and JAH. SJL and JAH wrote the manuscript with input from JDH, RL and AA.

### Additional data files

The following additional data are available with the online version of this paper. Additional data file 1 is a figure showing a GenMAPP generated embryonic stem cell pathway that corresponds to genes enriched in ESCs (green), EBs (orange) and

BCs (red) with epithelial cells as a baseline. Additional data file 2 is a figure showing a predominant Ingenuity generated network that was identified in all four data sets of the level III analysis. *GATA1* interacts with other nuclear genes such as *TAL1*, *LMO2*, and *KLF1*, which induce erythropoietic genes such as the hemoglobin family (*HBG1*, *HBG2*, *HBE*, *HBB*, and *HBZ*), and those involved in heme synthesis (*ALAS2*). Additional data files 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 are tables listing genes that were up-regulated in ESCs, EBs, or BCs relative to each other or a specific reference tissue. Additional data file 3 contains a list of genes that are down-regulated upon differentiation of ESCs into EBs. Additional data file 4 contains a list of genes that are down-regulated upon differentiation of ESCs into BCs. Additional data file 5 contains a list of genes that are up-regulated upon differentiation of ESCs into EBs. Additional data file 6 contains a list of genes that are down-regulated upon differentiation of EBs into BCs. Additional data file 7 contains a list of genes that are up-regulated upon differentiation of EBs into BCs. Additional data file 8 contains a list of genes that are up-regulated upon differentiation of ESCs into BCs. Additional data file 9 contains a list of genes that are up-regulated in ESCs when compared to breast epithelia. Additional data file 10 contains a list of genes that are up-regulated in EBs when compared to breast epithelia. Additional data file 11 contains a list of genes that are up-regulated in BCs when compared to breast epithelia. Additional data file 12 contains a list of genes that are up-regulated in BCs when compared to leukocytes. Additional data file 13 contains a list of genes that are up-regulated in BCs when compared to endothelial cells. Additional data file 14 contains a list of genes that are up-regulated in BCs when compared to stromal cells. Additional data files 15, 16, 17, 18, 19, 20 are six CEL files of data from human U133 Plus 2.0 arrays (Affymetrix, Inc.) hybridized to RNA from purified BCs, day 3.5 EBs and undifferentiated ESCs (from two hESC lines, H1 and H9). Additional data file 15 contains a CEL file of undifferentiated ESCs, H1-GFP labeled. Additional data file 16 contains a CEL file of day3.5 EBs-H1 and additional data file 17 contains a CEL file of BC derived from H1. Additional data file 18 contains a CEL file from undifferentiated ESCs, H9. Additional data file 19 contains a CEL file from day3.5 EBs-H9. Additional data file 20 contains a CEL file of BC derived from H9.

### Acknowledgements

We would like to thank Dr Phyllis L Spatrick from the Genomics Core Facility, University of Massachusetts Medical School, for performing the hybridization procedures.

### References

1. Kaufman DS, Hanson ET, Lewis RL, Auerbach R, Thomson JA: **Hematopoietic colony-forming cells derived from human embryonic stem cells.** *Proc Natl Acad Sci USA* 2001, **98**:10716-10721.
2. Lu S-J, Li F, Vida L, Honig GR: **CD34+CD38- hematopoietic precursors derived from human embryonic stem cells exhibit an embryonic gene expression pattern.** *Blood* 2004,

- 103:4134-4141.
3. Wang L, Li L, Shojaei F, Levac K, Cerdan C, Menendez P, Martin T, Rouleau A, Bhatia M: **Endothelial and hematopoietic cell fate of human embryonic stem cells originates from primitive endothelium with hemangioblastic properties.** *Immunity* 2004, **21**:31-41.
  4. Wang L, Menendez P, Shojaei F, Li L, Mazurier F, Dick JE, Cerdan C, Levac K, Bhatia M: **Generation of hematopoietic repopulating cells from human embryonic stem cells independent of ectopic HOXB4 expression.** *J Exp Med* 2005, **201**:1603-1614.
  5. Chadwick K, Wang L, Li L, Menendez P, Murdoch B, Rouleau A, Bhatia M: **Cytokines and BMP-4 promote hematopoietic differentiation of human embryonic stem cells.** *Blood* 2003, **102**:906-915.
  6. Woll PS, Martin CH, Miller JS, Kaufman DS: **Human embryonic stem cell-derived NK cells acquire functional receptors and cytolytic activity.** *J Immunol* 2005, **175**:5095-5103.
  7. Wang L, Menendez P, Cerdan C, Bhatia M: **Hematopoietic development from human embryonic stem cell lines.** *Exp Hematol* 2005, **33**:987-996.
  8. Zambidis ET, Peault B, Park TS, Bunz F, Civin CI: **Hematopoietic differentiation of human embryonic stem cells progresses through sequential hematoendothelial, primitive, and definitive stages resembling human yolk sac development.** *Blood* 2005, **106**:860-870.
  9. Levenberg S, Golub JS, Amit M, Itskovitz-Eldor J, Langer R: **Endothelial cells derived from human embryonic stem cells.** *Proc Natl Acad Sci USA* 2002, **99**:4391-4396.
  10. Choi K: **The hemangioblast: a common progenitor of hematopoietic and endothelial cells.** *J Hematother Stem Cell Res* 2002, **11**:91-101.
  11. Choi K, Kennedy M, Kazarov A, Papadimitriou JC, Keller G: **A common precursor for hematopoietic and endothelial cells.** *Development* 1998, **125**:725-732.
  12. Kennedy M, Firpo M, Choi K, Wall C, Robertson S, Kabrun N, Keller G: **A common precursor for primitive erythropoiesis and definitive haematopoiesis.** *Nature* 1997, **386**:488-493.
  13. Cogle CR, Wainman DA, Jorgensen ML, Guthrie SM, Mames RN, Scott EV: **Adult human hematopoietic cells provide functional hemangioblast activity.** *Blood* 2004, **103**:133-135.
  14. Guo H, Fang B, Liao L, Zhao Z, Liu J, Chen H, Hsu SH, Cui Q, Zhao RC: **Hemangioblastic characteristics of fetal bone marrow-derived Flk1(+)/CD31(-)/CD34(-) cells.** *Exp Hematol* 2003, **31**:650-658.
  15. Loges S, Fehse B, Brockmann MA, Lamszus K, Butz M, Guckenbiehl M, Schuch G, Ergun S, Fischer U, Zander AR, Hossfeld DK, Fiedler W, Gehling UM: **Identification of the adult human hemangioblast.** *Stem Cells Dev* 2004, **13**:229-242.
  16. Pelosi E, Valtieri M, Coppola S, Botta R, Gabbianelli M, Lulli V, Marziali G, Masella B, Muller R, Sgadari C, Testa U, Bonanno G, Peschle CI: **Identification of the hemangioblast in postnatal life.** *Blood* 2002, **100**:3203-3208.
  17. Grant MB, May WVS, Caballero S, Brown GA, Guthrie SM, Mames RN, Byrne BJ, Vaught T, Spoerri PE, Peck AB, Scott EV: **Adult hematopoietic stem cells provide functional hemangioblast activity during retinal neovascularization.** *Nat Med* 2002, **8**:607-612.
  18. Bailey AS, Jiang S, Afentoulis M, Baumann CI, Schroeder DA, Olson SB, Wong MH, Fleming WH: **Transplanted adult hematopoietic stem cells differentiate into functional endothelial cells.** *Blood* 2004, **103**:13-19.
  19. Umeda K, Heike T, Yoshimoto M, Shinoda G, Shiota M, Suemori H, Luo HY, Chui DH, Torii R, Shibuya M, Nakatsuji N, Nakahata T: **Identification and characterization of hemoangiogenic progenitors during cynomolgus monkey embryonic stem cell differentiation.** *Stem Cells* 2006, **24**:1348-1358.
  20. Kennedy M, D'Souza SL, Lynch-Kattman M, Schwantz S, Keller G: **Development of the hemangioblast defines the onset of hematopoiesis in human ES cell differentiation cultures.** *Blood* 2007, **109**:2679-2687.
  21. Lu SJ, Feng Q, Caballero S, Chen Y, Moore MAS, Grant MB, Lanza R: **Generation of functional hemangioblasts from human embryonic stem cells.** *Nat Methods* 2007, **4**:501-509.
  22. Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J, et al.: **The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells.** *Nat Genet* 2006, **38**:431-440.
  23. Wang J, Rao S, Chu J, Shen X, Levasseur DN, Theunissen TW, Orkin SH: **A protein interaction network for pluripotency of embryonic stem cells.** *Nature* 2006, **444**:364-368.
  24. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, et al.: **Core transcriptional regulatory circuitry in human embryonic stem cells.** *Cell* 2005, **122**:947-956.
  25. Gering M, Yamada Y, Rabbitts TH, Patient RK: **Lmo2 and Scl/Tal1 convert non-axial mesoderm into haemangioblasts which differentiate into endothelial cells in the absence of Gata1.** *Development* 2003, **130**:6187-6199.
  26. Lugas JJ, Chung YS, Mills JC, Kim SI, Grass JA, Kyba M, Doherty JM, Bresnick EH, Choi K: **GATA2 functions at multiple steps in hemangioblast development and differentiation.** *Development* 2007, **134**:393-405.
  27. D'Souza SL, Elefanty AG, Keller G: **SCL/Tal-1 is essential for hematopoietic commitment of the hemangioblast but not for its development.** *Blood* 2005, **105**:3862-3870.
  28. Park C, Afrikanova I, Chung YS, Zhang WJ, Arentson E, Fong GG, Rosendahl A, Choi K: **A hierarchical order of factors in the generation of FLK- and SCL-expressing hematopoietic and endothelial progenitors from embryonic stem cells.** *Development* 2004, **131**:2749-2762.
  29. Lacaud G, Gore L, Kennedy M, Kouskoff V, Kingsley P, Hogan C, Carlsson L, Speck N, Palis J, Keller G: **Runx1 is essential for hematopoietic commitment at the hemangioblast stage of development in vitro.** *Blood* 2002, **100**:458-466.
  30. Shawber CJ, Das I, Francisco E, Kitajewski J: **Notch signaling in primary endothelial cells.** *Ann NY Acad Sci* 2003, **995**:162-170.
  31. Shiose S, Hata Y, Noda Y, Sassa Y, Takeda A, Yoshikawa H, Fujisawa K, Kubota T, Ishibashi T: **Fibrinogen stimulates in vitro angiogenesis by choroidal endothelial cells via autocrine VEGF.** *Graefes Arch Clin Exp Ophthalmol* 2004, **242**:777-783.
  32. Beleslin-Cokic BB, Cokic VP, Yu X, Weksler BB, Schechter AN, Noguchi CT: **Erythropoietin and hypoxia stimulate erythropoietin receptor and nitric oxide production by endothelial cells.** *Blood* 2004, **104**:2073-2080.
  33. Wang GX, Cai SX, Wang PQ, Ouyang KQ, Wang YL, Xu SR: **Shear-induced changes in endothelin-1 secretion of microvascular endothelial cells.** *Microvasc Res* 2002, **63**:209-217.
  34. Deckers MM, van Bezooijen RL, van der HG, Hoogendam J, van Der BC, Papapoulos SE, Lowik CW: **Bone morphogenetic proteins stimulate angiogenesis through osteoblast-derived vascular endothelial growth factor A.** *Endocrinology* 2002, **143**:1545-1553.
  35. Joukov V, Kumar V, Sorsa T, Arighi E, Weich H, Saksela O, Alitalo K: **A recombinant mutant vascular endothelial growth factor-C that has lost vascular endothelial growth factor receptor-2 binding, activation, and vascular permeability activities.** *J Biol Chem* 1998, **273**:6599-6602.
  36. Wang D, Donner DB, Warren RS: **Homeostatic modulation of cell surface KDR and Flt1 expression and expression of the vascular endothelial cell growth factor (VEGF) receptor mRNAs by VEGF.** *J Biol Chem* 2000, **275**:15905-15911.
  37. Soker S, Takashima S, Miao HQ, Neufeld G, Klagsbrun M: **Neuropilin-1 is expressed by endothelial and tumor cells as an isoform-specific receptor for vascular endothelial growth factor.** *Cell* 1998, **92**:735-745.
  38. Gluzman-Poltorak Z, Cohen T, Herzog Y, Neufeld G: **Neuropilin-2 is a receptor for the vascular endothelial growth factor (VEGF) forms VEGF-145 and VEGF-165.** *J Biol Chem* 2000, **275**:29922.
  39. Fischer A, Klattig J, Kneitz B, Diez H, Maier M, Holtmann B, Englert C, Gessler M: **Hey basic helix-loop-helix transcription factors are repressors of GATA4 and GATA6 and restrict expression of the GATA target gene ANF in fetal hearts.** *Mol Cell Biol* 2005, **25**:8960-8970.
  40. Fischer A, Schumacher N, Maier M, Sendtner M, Gessler M: **The Notch target genes Hey1 and Hey2 are required for embryonic vascular development.** *Genes Dev* 2004, **18**:901-911.
  41. Grotewold L, Ruther U: **The Wnt antagonist Dickkopf-1 is regulated by Bmp signaling and c-Jun and modulates programmed cell death.** *EMBO J* 2002, **21**:966-975.
  42. Chu EY, Hens J, Andl T, Kairo A, Yamaguchi TP, Brisken C, Glick A, Wysolmerski JJ, Millar SE: **Canonical WNT signaling promotes mammary placode development and is essential for initiation of mammary gland morphogenesis.** *Development* 2004, **131**:4819-4829.
  43. Schulze M, Belema-Bedada F, Technau A, Braun T: **Mesenchymal stem cells are recruited to striated muscle by NFAT/IL-4-mediated cell fusion.** *Genes Dev* 2005, **19**:1787-1798.
  44. Kathiriyi IS, King IN, Murakami M, Nakagawa M, Astle JM, Gardner

- KA, Gerard RD, Olson EN, Srivastava D, Nakagawa O: **Hairy-related transcription factors inhibit GATA-dependent cardiac gene expression through a signal-responsive mechanism.** *J Biol Chem* 2004, **279**:54937-54943.
45. Gianakopoulos PJ, Skerjanc IS: **Hedgehog signaling induces cardiomyogenesis in P19 cells.** *J Biol Chem* 2005, **280**:21022-21028.
  46. Hiroi Y, Kudoh S, Monzen K, Ikeda Y, Yazaki Y, Nagai R, Komuro I: **Tbx5 associates with Nkx2-5 and synergistically promotes cardiomyocyte differentiation.** *Nat Genet* 2001, **28**:276-280.
  47. Charite J, McFadden DG, Olson EN: **The bHLH transcription factor dHAND controls Sonic hedgehog expression and establishment of the zone of polarizing activity during limb development.** *Development* 2000, **127**:2461-2470.
  48. McFadden DG, Barbosa AC, Richardson JA, Schneider MD, Srivastava D, Olson EN: **The Hand1 and Hand2 transcription factors regulate expansion of the embryonic cardiac ventricles in a gene dosage-dependent manner.** *Development* 2005, **132**:189-201.
  49. Zou Y, Evans S, Chen J, Kuo HC, Harvey RP, Chien KR: **CARP, a cardiac ankyrin repeat protein, is downstream in the Nkx2-5 homeobox gene pathway.** *Development* 1997, **124**:793-804.
  50. Ishiguro N, Baba T, Ishida T, Takeuchi K, Osaki M, Araki N, Okada E, Takahashi S, Saito M, Watanabe M, et al.: **Carp, a cardiac ankyrin-repeated protein, and its new homologue, Arpp, are differentially expressed in heart, skeletal muscle, and rhabdomyosarcomas.** *Am J Pathol* 2002, **160**:1767-1778.
  51. Makino Y, Cao R, Svensson K, Bertilsson G, Asman M, Tanaka H, Cao Y, Berkenstam A, Poellinger L: **Inhibitory PAAS domain protein is a negative regulator of hypoxia-inducible gene expression.** *Nature* 2001, **414**:550-554.
  52. Degar BA, Baskaran N, Hulspas R, Quesenberry PJ, Weissman SM, Forget BG: **The homeodomain gene Pitx2 is expressed in primitive hematopoietic stem/progenitor cells but not in their differentiated progeny.** *Exp Hematol* 2001, **29**:894-902.
  53. Suh H, Gage PJ, Drouin J, Camper SA: **Pitx2 is required at multiple stages of pituitary organogenesis: pituitary primordium formation and cell specification.** *Development* 2002, **129**:329-337.
  54. Gering M, Rodaway AR, Gottgens B, Patient RK, Green AR: **The SCL gene specifies haemangioblast development from early mesoderm.** *EMBO J* 1998, **17**:4029-4045.
  55. Freson K, Thys C, Wittewrongel C, Vermeylen J, Hoylaerts MF, Van GC: **Molecular cloning and characterization of the GATA1 cofactor human FOG1 and assessment of its binding to GATA1 proteins carrying D218 substitutions.** *Hum Genet* 2003, **112**:42-49.
  56. Sevinsky JR, Whalen AM, Ahn NG: **Extracellular signal-regulated kinase induces the megakaryocyte GPIIb/CD41 gene through MafB/Kreisler.** *Mol Cell Biol* 2004, **24**:4534-4545.
  57. Yokomizo T, Takahashi S, Mochizuki N, Kuroha T, Ema M, Wakamatsu A, Shimizu R, Ohneda O, Osato M, Okada H, et al.: **Characterization of GATA-1(+) hemangioblastic cells in the mouse embryo.** *EMBO J* 2007, **26**:184-196.
  58. Bhattacharya B, Miura T, Brandenberger R, Mejido J, Luo Y, Yang AX, Joshi BH, Ginis I, Thies RS, Amit M, et al.: **Gene expression in human embryonic stem cell lines: unique molecular signature.** *Blood* 2004, **103**:2956-2964.
  59. Bhattacharya B, Cai J, Luo Y, Miura T, Mejido J, Brimble SN, Zeng X, Schulz TC, Rao MS, Puri RK: **Comparison of the gene expression profile of undifferentiated human embryonic stem cell lines and differentiating embryoid bodies.** *BMC Dev Biol* 2005, **5**:22.
  60. Dvash T, Mayshar Y, Darr H, McElhaney M, Barker D, Yanuka O, Kotkoff KJ, Rubin LL, Benvenisty N, Eiges R: **Temporal gene expression during differentiation of human embryonic stem cells and embryoid bodies.** *Hum Reprod* 2004, **19**:2875-2883.
  61. Liu Y, Shin S, Zeng X, Zhan M, Gonzalez R, Mueller FJ, Schwartz CM, Xue H, Li H, Baker SC, et al.: **Genome wide profiling of human embryonic stem cells (hESCs), their derivatives and embryonic carcinoma cells to develop base profiles of U.S. Federal government approved hESC lines.** *BMC Dev Biol* 2006, **6**:20.
  62. Li H, Liu Y, Shin S, Sun Y, Loring JF, Mattson MP, Rao MS, Zhan M: **Transcriptome coexpression map of human embryonic stem cells.** *BMC Genomics* 2006, **7**:103.
  63. Sun Y, Li H, Liu Y, Shin S, Mattson MP, Rao MS, Zhan M: **Cross-species transcriptional profiles establish a functional portrait of embryonic stem cells.** *Genomics* 2007, **89**:22-35.
  64. Zhong JF, Zhao Y, Sutton S, Su A, Zhan Y, Zhu L, Yan C, Gallaher T, Johnston PB, Anderson WF, et al.: **Gene expression profile of murine long-term reconstituting vs. short-term reconstituting hematopoietic stem cells.** *Proc Natl Acad Sci USA* 2005, **102**:2448-2453.
  65. Huber TL, Kouskoff V, Fehling HJ, Palis J, Keller G: **Haemangioblast commitment is initiated in the primitive streak of the mouse embryo.** *Nature* 2004, **432**:625-630.
  66. Vogeli KM, Jin SW, Martin GR, Stainier DY: **A common progenitor for haematopoietic and endothelial lineages in the zebrafish gastrula.** *Nature* 2006, **443**:337-339.
  67. Kouskoff V, Lacaud G, Schwantz S, Fehling HJ, Keller G: **Sequential development of hematopoietic and cardiac mesoderm during embryonic stem cell differentiation.** *Proc Natl Acad Sci USA* 2005, **102**:13170-13175.
  68. Kattman SJ, Huber TL, Keller GM: **Multipotent flk-1+ cardiovascular progenitor cells give rise to the cardiomyocyte, endothelial, and vascular smooth muscle lineages.** *Dev Cell* 2006, **11**:723-732.
  69. Wu SM, Fujiwara Y, Cibulsky SM, Clapham DE, Lien CL, Schultheiss TM, Orkin SH: **Developmental origin of a bipotential myocardial and smooth muscle cell precursor in the mammalian heart.** *Cell* 2006, **127**:1137-1150.
  70. Moretti A, Caron L, Nakano A, Lam JT, Bernshausen A, Chen Y, Qyang Y, Bu L, Sasaki M, Martin-Puig S, et al.: **Multipotent embryonic isl1+ progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification.** *Cell* 2006, **127**:1151-1165.
  71. **Gene Expression Omnibus** [<http://www.ncbi.nlm.nih.gov/geo>]
  72. Wettenhall JM, Smyth GK: **limmaGUI: a graphical user interface for linear modeling of microarray data.** *Bioinformatics* 2004, **20**:3705-3706.
  73. Wettenhall JM, Simpson KM, Satterley K, Smyth GK: **afflymGUI: a graphical user interface for linear modeling of single channel microarray data.** *Bioinformatics* 2006, **22**:897-899.
  74. Bolstad BM, Irizarry RA, Astrand M, Speed TP: **A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.** *Bioinformatics* 2003, **19**:185-193.
  75. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: **Exploration, normalization, and summaries of high density oligonucleotide array probe level data.** *Biostatistics* 2003, **4**:249-264.
  76. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP: **Summaries of Affymetrix GeneChip probe level data.** *Nucleic Acids Res* 2003, **31**:e15.
  77. Smyth GK: **Linear models and empirical bayes methods for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol Biol* 2004, **3**:Article3.
  78. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al.: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25**:25-29.
  79. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA: **DAVID: Database for Annotation, Visualization, and Integrated Discovery.** *Genome Biol* 2003, **4**:3.
  80. Dahlquist KD, Salomonis N, Vranizan K, Lawlor SC, Conklin BR: **GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways.** *Nat Genet* 2002, **31**:19-20.
  81. Lu S-J, Li F, Vida L, Honig GR: **Comparative gene expression in hematopoietic progenitor cells derived from embryonic stem cells.** *Exp Hematol* 2002, **30**:58-66.
  82. Zhumabayeva B, Diatchenko L, Chenchik A, Siebert PD: **Use of SMART-generated cDNA for gene expression studies in multiple human tumors.** *Biotechniques* 2001, **30**:158-163.
  83. Seth D, Gorrell MD, McGuinness PH, Leo MA, Lieber CS, McCaughan GW, Haber PS: **SMART amplification maintains representation of relative gene expression: quantitative validation by real time PCR and application to studies of alcoholic liver disease in primates.** *J Biochem Biophys Methods* 2003, **55**:53-66.
  84. Zheng J, Shen W, He DZ, Long KB, Madison LD, Dallos P: **Prestin is the motor protein of cochlear outer hair cells.** *Nature* 2000, **405**:149-155.