

**The following resources related to this article are available online at
www.sciencemag.org (this information is current as of November 10, 2009):**

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/302/5644/393d>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/cgi/content/full/302/5644/393d#related-content>

This article **cites 8 articles**, 3 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/302/5644/393d#otherarticles>

This article appears in the following **subject collections**:

Development

<http://www.sciencemag.org/cgi/collection/development>

Technical Comments

http://www.sciencemag.org/cgi/collection/tech_comment

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

Response to Comments on " 'Stemness': Transcriptional Profiling of Embryonic and Adult Stem Cells" and "A Stem Cell Molecular Signature"

Today, biologists can define biological processes in molecular language by acquiring global gene expression profiles. Rigorous comparison of data sets from numerous laboratories is necessary to extract biological insights. Often, this is not without complications. The comments by Fortunel *et al.* (1) and Evsikov and Solter (2) highlight this issue. Both ask an important question: Why do independently identified sets of "stemness" genes (1, 3, 4) display very little overlap? For example, only 6 out of 82 genes (7.3% overlap) in the stem cell (SC) signature of genes overexpressed by three SC populations as defined by Ivanova *et al.* (3) are shared by the stemness profile defined by Ramalho-Santos *et al.* (4, 5). Similarly, only 27 out of the 295 genes (9.2% overlap) overexpressed by at least two SC populations identified by Ivanova *et al.* are identified in similar comparisons with Ramalho-Santos *et al.* To better understand these discrepancies, we have compared and analyzed all procedures used Fortunel *et al.* (1), Ivanova *et al.* (3), and Ramalho-Santos *et al.* (4), including SC definition and isolation and the technical and analytical aspects of the microarray studies (Table 1).

It is important to stress that gene expression monitoring technologies are only as powerful as the biological definition of the starting cell populations. Hematopoietic stem cells (HSCs) produce all blood cells in mouse and man. These are the only somatic SCs that have been isolated to near-homogeneity and whose functions have been rigorously quantified in vitro and in vivo. The studies of Ivanova *et al.* (3) and Ramalho-Santos *et al.* (4) are grounded in establishing the gene expression profiles of the HSC. In the case of neural stem cells (NSCs), the in vitro propagated neurosphere contains cells that can produce neurons, glia, and additional primitive cells. Therefore, one can say that there are progenitor cells within these structures. However, the exact relationship of neurosphere cells to "true" NSCs is poorly defined. The major cell populations found in neurospheres are heterogeneous, and the resident SC

compartment is very minor (6). It is likely that a neurosphere contains a mixture of SCs, progenitor cells, and more mature cells developmentally similar to those found in an unfractionated hematopoietic system. Microarray technology allows the detection of genes expressed at low levels, but if only a small proportion of an analyzed cell preparation comprises SCs, many important SC regulatory gene products, often expressed at low levels, will go undetected. Similar concerns apply to the retinal progenitor/stem cell (RPC) population utilized by Fortunel *et al.* (1); RPCs have not been fully characterized. Finally, embryonic SCs (ESCs), although extremely useful, do not necessarily represent a naturally occurring SC population. In the light of these considerations, we question whether the studies performed by Fortunel *et al.* (1) are grounded in adequate definitions of SCs and, therefore, whether the data represent an analysis of gene expression in bona fide SCs. To highlight this point, among the ostensibly RPC-specific gene products defined by Fortunel *et al.*, one encounters immunoglobulin kappa light chain and six hemoglobin transcripts. This raises serious doubts regarding the nature and the purity of their "RPCs."

It is also crucial to highlight the importance of analyzing primary, freshly purified SC populations. NSCs (as defined by neurosphere generation) and ESCs are in vitro cultured cells, and in this sense their regulatory mechanisms may be in vitro adaptations. In primary cells, self-renewal and differentiation may depend on multiple signaling pathways, whereas in cultured cells, these processes may be controlled by a single in vitro "selected" pathway (FGF/EGF for NSCs and LIF for ESCs). Monitoring "natural" SC gene products responsible for regulating normal SC behavior is challenging and will be acutely sensitive to the definition and purity of SC fractions, the choice of comparative baselines, analytical algorithms, and other variables. Because of the ranges of these parameters, it is likely that some SC genes may be detectable only by a combination of different approaches.

We nevertheless asked whether differences in data analyses could partially contribute to

observed discrepancies. All three groups used different analytical approaches (Table 1). For example, Fortunel *et al.* (1) and Ramalho-Santos *et al.* (4) relied on the statistical significance of gene expression changes observed between "SCs" and differentiated cells, and essentially ignored the fold change values. Ivanova *et al.* (3), in addition to statistical-significance analyses, included a twofold-or-greater fold change criterion for designating a gene product to be enriched in SCs. Applying this additional criterion to the 2230 nonredundant "RPC-specific" genes listed by Fortunel *et al.* (1) yields only 1135 genes that show more than a twofold expression increase when compared with the corresponding baseline sample. Housekeeping genes such as *GAPDH* can display subtle expression level variations in different cells. Thus, ~50% of the "RPC-specific" genes of Fortunel *et al.* could equally well be described as "broadly expressed." Although a possible role for these genes in SCs cannot be ruled out at present, they can equally well be defined as "housekeeping" genes. Finally, we note that gene products previously implicated in stem cell regulation are identified in the studies of Ivanova *et al.* and Ramalho-Santos *et al.* These include *Edr1*, *Hes1*, *Tcf3*, *Mdr1*, *ThrombinR*, *Integrin b1*, *p21*, *Cyclin D1*, and *Eif4g2/Nat1*, as well as numerous members of the *Jak/Stat*, *Wnt*, and *Notch* signaling pathways.

To investigate the effects of different analytical methods more directly, we reanalyzed the data of Ivanova *et al.* and Ramalho-Santos *et al.* using identical statistical methods for designating or scoring gene products as enriched in SCs. These analyses show a total of 605 gene products identified in common between the HSC populations studied by the two laboratories, a 65.7% overlap (Fig. 1).

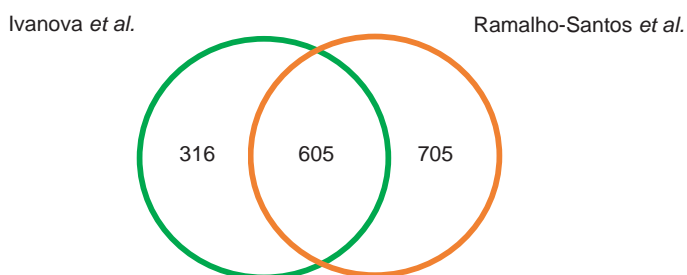
We stress that individual genes whose products are components of widely utilized stem cell regulatory pathways could be expressed in multiple, though not necessarily all, SC types. Related homologs or members of protein families may perform similar functions in different SC populations. We have noticed, for example, that the expression patterns of individual *Wnt* signaling pathway components are distinct in different SC populations (3). Nevertheless, each SC type expresses at least one member of the *Frizzled* receptor family, at least one *Tcf/Lef* transcriptional mediator family member, and at least one member of the *Wnt* ligand family. The importance of the *Wnt* pathway in multiple stem cell types of the hematopoietic, gut, and epidermal systems is well established (7–9). It is very likely that more sophisticated analyses of the different data sets from the point of view of gene ontologies and pathways will reveal commonalities and shared features.

TECHNICAL COMMENT

Table 1. Comparative analysis of experimental conditions.

	Ivanova <i>et al.</i> (3)	Ramalho-Santos <i>et al.</i> (4)	Fortunel <i>et al.</i> (1)
Stem cell populations	Fluorescence-activated cell sorting (FACS) purified long-term (LT) and short-term (ST) HSC (both embryonic and adult), cultured d13.5 NSC, ES cell line CCE	FACS-purified adult HSC, cultured adult NSC, C57Bl/6 ES cell line	Unfractionated embryonic retinas (RPC), cultured d13.5 NSC, ES cell line E14
Baseline cell populations	FACS-purified lineage-committed progenitors (LCPs) and mature blood cells (both embryonic and adult)	Unfractionated BM cell, unfractionated lateral ventricular brain (LVB) cells	Unfractionated D10 postpartum retinas, unfractionated LVB cells
Functional analysis of cell populations	In vivo competitive repopulation assay for all HSC populations; colony-forming cell (CFC) assay for HSC and LCP populations; secondary neurosphere formation and multilineage differentiation capacity for NSC	None	None
Analysis algorithm	<ol style="list-style-type: none"> 1. Probe sets are assigned into hematopoietic clusters that correlate with functions of hematopoietic subsets analyzed. 2. Probes that show "conflicting" expression profiles in BM and fetal liver (FL) are removed. 3. Hematopoietic clusters are filtered to remove low-fold change and low-confidence probe sets. 4. Only HSC-enriched genes are scored for enrichment in NSC and ESC. 5. Genes enriched in all three SC are defined by individual scores rather than an average score. 	<ol style="list-style-type: none"> 1. Probe sets that show increased expression in SC compare to baseline sample within 90% confidence interval are defined as SC-enriched. 2. Genes enriched in all three SC are defined on the basis of average score. 	Exact parameters of the algorithm are not clearly described.
Microarrays used	MG-U74 A, B, C	MG-U74 A only	MG-U74 A only
Data verification	Semi-quantitative reverse transcription-polymerase chain reaction (RT-PCR) analysis for selected HSC-specific genes	None	None

Fig. 1. Venn diagram showing similarities between HSC populations isolated by Ivanova *et al.* (3) and Ramalho-Santos *et al.* (4). Bone marrow (BM) HSCs from both studies were compared with BM Lin⁺ cells used by Ivanova *et al.* using MAS5.0 software. HSC-specific probe sets were scored as HSC-enriched if they were defined as "present" and twofold up-regulated with 90% confidence. 605 probe sets (65.7%) were found that satisfied these criteria in both datasets. 705 probe sets were up-regulated in the Ramalho-Santos *et al.* data set, but not in the Ivanova *et al.* data set; 316 probe sets were up-regulated in the Ivanova *et al.* data set, but not in the Ramalho-Santos *et al.* dataset. Among these 316 probe sets were several genes that were PCR-confirmed as HSC-enriched by Ivanova *et al.*, such as MDR1, Fzd4, Nr4a2, Sox4, Hgf1. The "nonshared" uniquely present in either of the two data sets are likely to be a reflection of cell population differences in HSC subsets that are differentially purified by Rhodamine 123 and SP-based protocols. The biological similarities of the three HSC populations suggest that the shared gene products are the most interesting for functional studies. These analyses underscore the value of comparative studies.



The most important point, raised by all authors, is that the validity and utility of these reports ultimately lies with in vivo and in vitro assays for the functions of the gene products identified in the SCs. In this respect, we are encouraged that among the gene products identified as enriched in ESCs by both Ivanova *et al.* (3) and Ramalho-Santos *et al.* (4), one (Nanog) has recently been shown to be required for the maintenance of the undifferentiated state (10, 11).

Natalia B. Ivanova
John T. Dimos
Christopher Schaniel
Jason A. Hackney
Kateri A. Moore
Department of Molecular Biology
Princeton University
Washington Road
Princeton, NJ 08544, USA

Miguel Ramalho-Santos
 Department of Molecular and
 Cellular Biology and
 Howard Hughes Medical Institute
 Harvard University
 Cambridge, MA 02138, USA

Soonsang Yoon
Yumi Matsuzaki
Richard C. Mulligan
 Department of Genetics
 Harvard Medical School
 Boston, MA 02115, USA

Douglas A. Melton
 Department of Molecular and
 Cellular Biology and
 Howard Hughes Medical Institute
 Harvard University

Ihor R. Lemischka
 Department of Molecular Biology
 Princeton University
 E-mail: ilemischka@molbio.princeton.edu

References and Notes

1. N. O. Fortunel *et al.*, *Science* **302**, 393 (2003); www.sciencemag.org/cgi/content/full/302/5644/393b.

2. A. V. Evsikov, D. Solter, *Science* **302**, 393 (2003); www.sciencemag.org/cgi/content/full/302/5644/393c.
3. N. B. Ivanova *et al.*, *Science* **298**, 601 (2002).
4. M. Ramalho-Santos, S. Yoon, Y. Matsuzaki, R. C. Mulligan, D. A. Melton, *Science* **298**, 597 (2002).
5. These comparisons are restricted to the genes represented by the Affymetrix MG-U74A microarray, the set of genes analyzed in both (3) and (4). Because of the presence of redundant probe sets on the microarray, Unigene 119 identifiers were utilized for these calculations rather than the the Affymetrix probe IDs.
6. B. A. Reynolds, S. Weiss, *Dev. Biol.* **175**, 1 (1996).
7. T. Reya *et al.*, *Nature* **423**, 409 (2003).
8. V. Korinek *et al.*, *Nature Genet.* **19**, 379 (1998).
9. L. Alonzo, E. Fuchs, *Genes Dev.* **17**, 1189 (2003).
10. M. K. Tokuzawa *et al.*, *Cell* **113**, 631 (2003).
11. I. Chambers *et al.*, *Cell* **113**, 643 (2003).

19 June 2003; accepted 23 September 2003