

The following resources related to this article are available online at www.sciencemag.org (this information is current as of November 9, 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/302/5644/393b>

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/302/5644/393b#related-content>

This article **cites 6 articles**, 4 of which can be accessed for free:

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

This article has been **cited by 2 articles** hosted by HighWire Press; see:

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

TECHNICAL COMMENT

Comment on “ ‘Stemness’: Transcriptional Profiling of Embryonic and Adult Stem Cells” and “A Stem Cell Molecular Signature” (I)

Ramalho-Santos *et al.* (1) and Ivanova *et al.* (2), comparing the same three “stem cells”—embryonic stem cells (ESCs); neural stem cells (NSCs), referred to as neural progenitor/stem cells (NPCs) in the present study; and hematopoietic stem cells (HSCs)—with their differentiated counterparts, each identified a list of commonly expressed “stemness” genes, proposed to be important for conferring the functional characteristics of stem cells. The ability to capture expression profiles of cells using microarrays offers the possibility of defining a stem cell by its constellation of active genes. An intriguing question, however, is whether the functional commonalities (self-renewal and pluripotency) (3) among stem cells can be defined at the genetic level. Do all stem cells express a similar set of “stemness” genes necessary for their unique properties, or do different stem cells express different sets of genes that confer stemness?

We have independently carried out gene expression profiling of three types of stem or immature progenitor cells: ESCs, NPCs, and retinal progenitor/stem cells, or RPCs (Fig. 1) (4, 5). The intersection of ESC-, NPC- and RPC-enriched genes defined a list of 385 genes that are collectively expressed by all three stem cells (6). It can be inferred that these genes may represent or include putative “stemness” genes. For the approach taken here to be

able to define and support the notion of “stemness” genes, however, would also require that very similar sets of genes can be identified regardless of the type of stem cells used. To test the validity of this notion, we have collectively analyzed our results along with those from the studies of Ramalho-Santos *et al.* (1) and Ivanova *et al.* (2) (Figs. 2 and 3). To our surprise, a comparison of the three independently derived lists of “stemness” genes showed only one gene (integrin alpha-6) commonly identified in the three studies (Figs. 2A and 3A) (6). This finding raised serious concerns about the conclusions reported in (1) and (2), as was also critically highlighted by Burns and Zon (7).

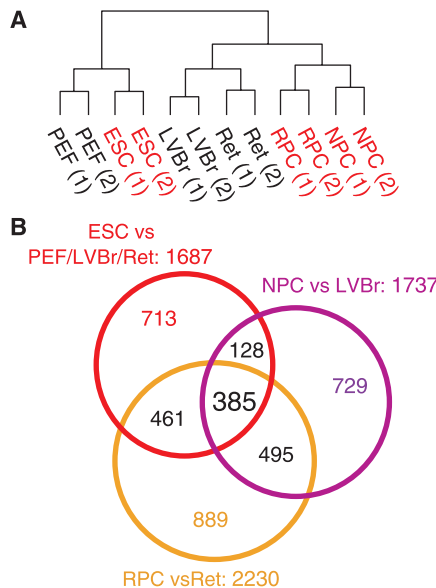


Fig. 1. (A) Hierarchical clustering of stem cells and their differentiated progenies (6), as identified in this study. The clustering of NPCs and RPCs with their differentiated progenies [tissues adjacent to lateral ventricles of adult brain (LVB) and mature retina (Ret)] suggest that somatic stem cells are already primed to become specific lineages. (B) Venn diagram showing the number of genes enriched in ESCs, NPCs and RPCs, as identified in this study. The overlap of these genes identified a common list of 385 stemness genes enriched in all three “stem cells” (6).

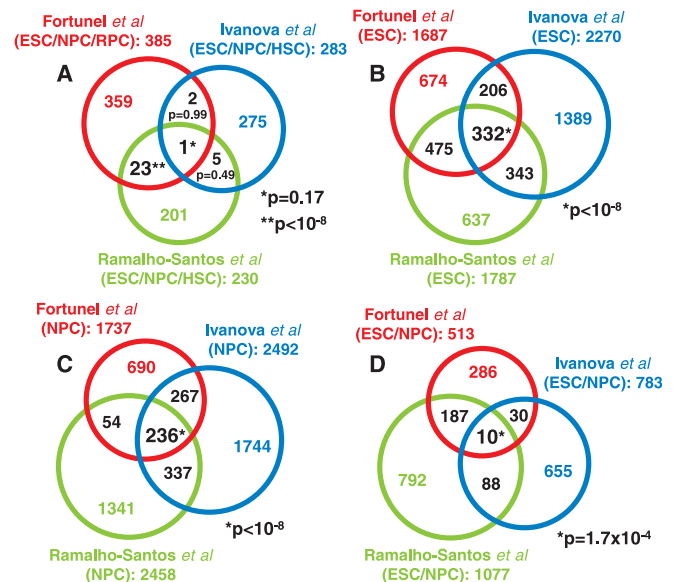


Fig. 2. Venn diagrams showing overlap of “stemness” genes and stem cell-enriched genes among studies by Ramalho-Santos *et al.* (1), Ivanova *et al.* (2), and Fortunel *et al.* (this study). Ivanova *et al.* used three different Affymetrix chips (U74v2 A, B, and C); Fortunel *et al.* and Ramalho-Santos *et al.* used only the U74v2 A chip (Fig. 3 shows same comparison, limiting Ivanova *et al.* results to the A chip). (A) “Stemness” genes found by the three groups overlap by only one gene. (B) ESC-enriched genes identified by each study overlap by 332 genes; the probability that such overlap occurs by chance is extremely low ($P < 10^{-8}$). (C) NPC-enriched genes overlapping by 236 genes between the three groups ($P < 10^{-8}$). (D) Overlap of “stemness” genes—two types of stem cell (ESC/NPC)-enriched genes—is limited to 10 genes. The probability of this number of genes overlapping by chance is greatly increased. $P > 10^{-4}$ is not significant because there are more than 10^4 genes studied (8).

TECHNICAL COMMENT

change (FC) value or lower confidence bound (LCB) score] (9). It is quite apparent that as one increases the number of stem cell types for comparison, the genes that occur in the intersection are genes

that show progressively lower differential expression between stem and differentiated cells (Fig. 4). For example, we found that the log of the significance scores for individual ESC or NPC-specific transcripts are

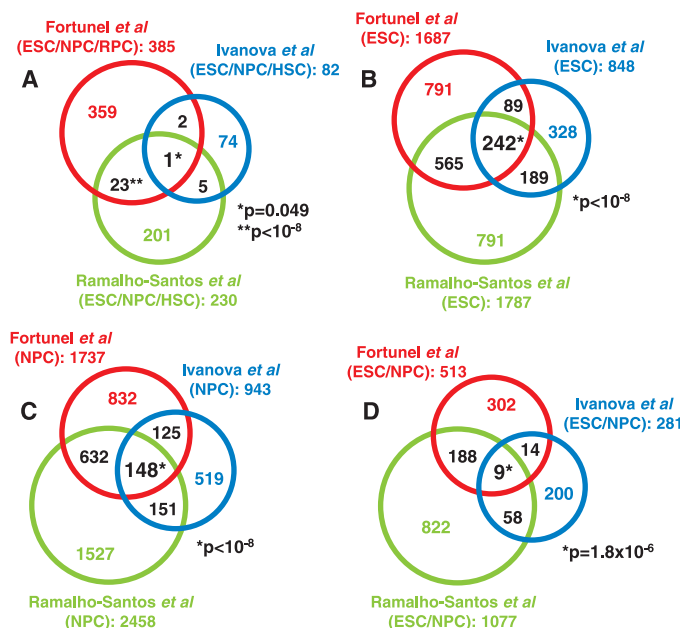


Fig. 3. (A to D) Same as Fig. 2, but with data limited to the Affymetrix U74v2 A chip for all three groups. Although the total number of genes identified by Ivanova *et al.* (2) is reduced, the number of genes found in the intersecting regions among the three studies are not changed significantly: The same single “stemness” gene common among the three studies remains in (A); the number of ESC- and NPC-enriched genes in (B) and (C), though it has declined, remains in the same order of magnitude; and the number of ESC/NPC-enriched genes in (D) has declined by only one, from 10 to 9. The decrease in the number of genes found in the intersecting regions of ESC- and NPC-enriched genes (panels B and C) is due to the fact that probes for some genes are present in more than one chip and that some genes that were detected in chip A by Ramalho-Santos *et al.* (1) and Fortunel *et al.* (this study) were detected by Ivanova *et al.* in either chip B or chip C but not in chip A. The *P* values remain low in both Fig. 2 and Fig. 3.



Fig. 4. (A to C) Comparison of relative expression levels of “stemness” genes by fold changes. The graphs depict significance scores for log FC or log LCB changes in the expression levels of the top 80 genes from the list of ESC-enriched genes (blue), NPC-enriched genes (orange), ESC/NPC-enriched genes (red), and “stemness” genes (green). “Stemness” genes here are (A) the top 80 genes from the 385 genes identified in this study (Fig. 1), (B) the 230 genes identified in (1), and (C) the 283 genes identified in (2), as published; values in (B) and (C) were derived from published data. Fold changes are highest for genes enriched in a single stem cell type; with two stem cell types (NPC and ESC), the top 80 genes have much lower fold change compared to one stem cell type. With three stem cells, the fold changes compared with individual stem cells are markedly reduced.

tion, amplified by subtle differences in experimental conditions between different studies. We propose this as the most important basis for the extreme discrepancies in the lists of putative “stemness” genes.

Our observation does not rule out the possibility that genes unique to all stem cells which are expressed at low levels may exist. However, it is clear that no one single study can confidently identify the bona fide genes that specify “stemness,” and cross-validation of lists generated independently by different investigators is crucial. For instance, the total of 24 genes commonly identified in two of the studies (Figs. 2A and 3A) is statistically significant and warrants further investigation (6). It is possible also that there are “stemness” genes that have not yet been identified and are not represented in the chips used. Genes that may be important for stem cell functions such as self-renewal but that are also expressed in non-stem cells (for example, *Stat3*, *gp130*) are unlikely to be identified by a comparative microarray approach. Another possibility is that different stem cell types may use different gene networks to achieve self-renewal or multipotency (10). Finally, “stemness” genes may only be transiently expressed, so that they are easily missed by comparing two homeostatic states. Further efforts to identify these genes will require different strategies.

In summary, speculations made from independent studies (1, 2) about identity of stemness genes do not hold up when the studies are compared. Our explanation for this also demonstrates the inherent problem of testing the stemness hypothesis using a profiling approach.

Nicolas O. Fortunel*
 Department of Medicine
 Beth Israel Deaconess Medical Center
 Harvard Institutes of Medicine
 77 Avenue Louis Pasteur
 Boston, MA 02115, USA and
 Genome Institute of Singapore
 60 Biopolis Street
 Singapore 138672 and
 Laboratoire de Biologie des Cellules
 Souches Humaines
 CNRS UPR 9045
 Villejuif, France

Hasan H. Otu*
 Department of Medicine
 Beth Israel Deaconess Medical Center

Huck-Hui Ng*
 Genome Institute of Singapore and
 Department of Biological Science
 National University of Singapore
 Singapore

Jinhui Chen
 Genome Institute of Singapore and
 Departments of Physiology and

Biochemistry
National University of Singapore

Xiuqian Mu

Department of Biochemistry and Molecular

Biology

The University of Texas M. D. Anderson

Cancer Center

Houston, TX 77030, USA

Timothy Chevassut

Xiaoyu Li

Marie Joseph

Charles Bailey

Department of Medicine

Beth Israel Deaconess Medical Center

Jacques A. Hatzfeld

Antoinette Hatzfeld

Laboratoire de Biologie des Cellules

Souches Humaines

Fatih Usta

Department of Medicine

Beth Israel Deaconess Medical Center

Vinsensius B. Vega

Philip M. Long

Genome Institute of Singapore

Towia A. Libermann

Department of Medicine

Beth Israel Deaconess Medical Center

Bing Lim

Department of Medicine

Beth Israel Deaconess Medical Center and

Genome Institute of Singapore

E-mail: blim@caregroup.harvard.edu

*These authors contributed equally to this work.

References and Notes

1. M. Ramalho-Santos, S. Yoon, Y. Matsuzaki, R. C. Mulligan, D. A. Melton, *Science* **298**, 597 (2002).
2. N. B. Ivanova *et al.*, *Science* **298**, 601 (2002).

3. J. E. Till and E. A. Mc Culloch, *Biochim. Biophys. Acta* **605**, 431 (1980).
4. E14 pluripotent ESCs were cultured following established methods on embryonic fibroblast feeder layer in the presence of leukemic inhibitory factor. NPCs were cultured as neurospheres derived from E12 embryos following an established protocol (11). Although the frequency of stem cells cannot be ascertained, this population consisted largely of cells that exhibit stem cell properties such as self-renewal and the capacity to generate the major cell types found in the central nervous system. RPCs were obtained from manually dissected from E14.5 mouse embryo retinas, a stage in which more than 90% of cells in the retina are undifferentiated neuroblasts (12). Mature retinas were dissected from day 10 postpartum (P10) mice. Lateral ventricular brain (LVbr) came from adult mice. Primary embryonic fibroblasts (PEF) were isolated from E13 embryo. All RNAs were extracted using Trizol.
5. Duplicate samples of total RNAs from stem/progenitor cells and differentiated cells were probed using the Affymetrix U74Av2 chip following established manufacturer's protocol. Samples that passed a priori quality control criteria were scaled using a 2% trimmed mean to perform comparative analyses. The high reproducibility of the samples for the replicate arrays was indicated by the high correlation coefficient values (between 0.95 and 0.98). An unsupervised-learning technique was applied by constructing an unweighted pair group method with arithmetic-mean (UPGMA) tree using Pearson's correlation as the metric of similarity (13). This tree represented the inherent grouping in the sample set and the degree of closeness between samples. To find genes enriched in a given cell type, we followed the supervised-learning approach as described by Tusher *et al.* (14). The statistical significance of the results was assessed using permutation testing. Genes that showed at most a 5% false discovery rate were reported as differentially expressed. Programs used in the analyses can be obtained through www.biostage.org. We have also used dChip software, used by Ramalho-Santos *et al.* (1), to compute for transcripts enriched in stem cells (data not shown). The results for the list of genes obtained were highly similar.
6. Raw data and gene lists are at http://giscompute.gis.a-star.edu.sg/suppdata_stemness
7. C. E. Burns, L. I. Zon, *Dev. Cell* **3**, 612 (2002).
8. The significance of the number of overlapping genes after intersection of gene lists was measured by *P* values, the probabilities that such overlaps could occur by chance alone. Monte Carlo simulations were

employed to approximate the *P* values of the three-group comparisons (6, 15). Ivanova *et al.* (2) used Affymetrix U74v2 A, B, and C chips, each containing ~12,000 probes. Since at most 12,000 genes [the number in this study and Ramalho-Santos *et al.* (1), using U74Av2 chip] from each study could be in any list of overlapping genes, our Monte Carlo analysis provides a conservative estimate of the significance of overlaps, including the study by Ivanova *et al.* (2). The *P* values arising from our Monte Carlo analysis are estimates of how unlikely it is that a degree of overlap in the gene lists of the different studies would arise by chance if the gene lists were chosen independently at random. The fact that *P* is small should be interpreted as suggesting that the processes through which these gene lists were generated have something in common (for example, if there was a small-to-moderate number of truly differentially expressed genes, these were very likely to be discovered in each of the studies, even if each of the studies also had a large number of false positives). In particular, the *P* values should not be interpreted as a measure of the confidence that all or even most of the overlapping genes are the correct ones."

9. ESC- and NPC-enriched genes were listed with their significance scores, as evaluated in the three studies by fold changes (FC) value or lower confidence bound (LCB) score, and tabulated for their log value. A graph is plotted as shown for the first 80 scores of ESC- and NPC-enriched genes (blue and orange curves in Fig. 4). To measure the extent to which the data implicate a gene as being differentially expressed in more than one stem cell type (for example, ESC and NPC), we took the average of the scores for differential expression between the two stem cells. A graph is plotted for ESC/NPC-enriched genes in each of the three studies (red curves in Fig. 4). A similar approach was taken to get the curves for the ESC/NPC/RPC (this study) or ESC/NPC/HSC (other two studies) "stemness" genes for the intersection of three stem cells (green curves in Fig. 4).
10. T. S. Burdon *et al.*, *Trends Cell Biol.* **9**, 432 (2002).
11. C. C. Morshead *et al.*, *Neuron* **5**, 1071 (1994).
12. X. Mu *et al.*, *Nucleic Acids Res.* **29**, 4983 (2001).
13. P. H. Sneath *et al.*, *Numerical Taxonomy* (Freeman, San Francisco, 1973).
14. V. G. Tusher *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5116 (2001).
15. A. C. Davison and D. Hinkley, *Bootstrap Methods and Their Application* (Cambridge Univ. Press, Cambridge, 1997).

2 May 2003; accepted 28 July 2003