A Stem Cell Molecular Signature: Are There Hallmark Properties That are Shared by all Stem Cells?

Ute Schepers*[a]

KEYWORDS:

bioinformatics · genetic profiling · genomics · stem cells · tissue replacement

Stem cells have long been regarded as a pool of undifferentiated cells that have the potential to proliferate, self-renew, and produce a variety of differentiated progeny, which gives them the capacity to regenerate several types of tissue. The classical paradigm describes unidirectional, hierarchical lineage proceedings, which follow step-wise differentiation of totipotent or pluripotent stem cells through intermediate, more restricted progenitor cells, and finally to differentiated cells. A long-accepted tenet of stem cell biology proposed a model in which only embryonic stem cells (ESC) are pluripotent and able to differentiate into several lineages of progeny, whereas adult stem cells are restricted in their differentiative potential to the tissue in which they reside. However, a recent series of studies has challenged these fundamental concepts by suggesting that stem cells derived from adult tissues can differentiate outside their tissue of origin. Thus stem cells from the blood system, termed hematopoietic stem cells (HSC), were suggested to be capable of producing mature liver cells, muscle tissue, or even neurons. Likewise, central nervous system stem cells or muscle stem cells have been considered capable of generating mature blood cell (MBC) populations.[1] The ability of tissue-specific stem cells to acquire the fate of cell types different from that in the tissue of origin has been termed adult stem cell plasticity.[2] These reports have generated considerable excitement since this discovery has raised the therapeutic possibility of using these stem cells for tissue repair and regeneration. Several mechanisms, all supported by published data, may underlie this apparent plasticity. It still remains unclear if multiple tissue-specific stem cells with different potential to differentiate coexist, or whether plasticity is the result of fusion of a donor cell with resident cells in an organ.[3, 4] Further models propose de- and redifferentiation (transdifferentiation)[5] of cells or the persistence of true multi- or pluripotent stem cells in postnatal life.[6] Nevertheless, little is known about the cellular and molecular prerequisites. Some doubts have even emerged about the biological significance of transdifferentiation, fusion, or finally adult stem cell plasticity. While Weissman and colleagues have reported failure to replicate transdifferentiation experiments,[7] Ihor Lemischka stressed the need for stronger evidence for adult stem cell plasticity and experimental rigor and caution.[8] In his opinion, the traditional paradigms should still not be abandoned. Moreover, the models are in need of mechanistic explanation. He claims that stem cell plasticity will most likely be directed by cell-autonomous mechanisms mediated by the panel of gene products expressed in stem cells. Since little is known about genetic programs in stem cells, it is necessary to identify the complete gene expression profiles that define all kinds of stem cells. There are still many open questions, such as, what are the mechanisms that regulate the fundamental decision that certain types of stem cells have to make, that is, whether to self-renew, or to make a commitment to a pathway of differentiation to produce a large population of mature cells? In the last few months, a tremendous step forward has been made by several groups towards answering this kind of question. By using high-throughput screening, the studies have applied one of the most powerful techniques of modern molecular biology to a question that has largely resisted investigation up until now. A series of publications that have revolutionized the field started last year with the *Science* paper, "A Stem Cell Molecular Signature", by Lemischka and co-workers, who determined the global gene expression profiles for mouse and human hematopoietic stem cells and other stages of the hematopoietic hierarchy.[9] This series is now climaxing with the first discoveries of several key players in the process of stem cell self-renewal in normal hematopoiesis and human leukemia.[10–12]

Since hematopoietic stem cells are so far the best-characterized stem cells, they were used as prototypes for these genetic profiling studies. These cells are defined by their tremendous ability to self-renew and differentiate into all types of mature blood cells. Continuous replacement of the mature cell populations is dictated by their limited lifetime. To cover the daily demands, the rate of mature blood cell production has to be very high. In humans, many billions of new cells are produced per day.[13] Stem cells from the hematopoietic system originate in the midgestation fetal liver. From the liver, these cells migrate into the bone marrow, where they reside to contribute to mature blood cells throughout adult life. This is

[a] *Dr. U. Schepers*
Universität Bonn
Kekulé-Institut für Organische Chemie und Biochemie
Gerhard-Domagk-Strasse 1
53121 Bonn (Germany)
Fax: (+49) 228-737778
E-mail: schepers@uni-bonn.de

  DOI: 10.1002/cbic.200300661

where the decision between self-renewal and commitment to differentiation finally takes place. The hematopoietic system is organized in a hierarchy. Stem cells follow a differentiation pathway from long-term functional hematopoietic stem cells, which are absolutely essential for the renewal of the whole hematopoietic system, through short-term functional hematopoietic stem cells, which still show some self-renewing ability and thus can differentiate to form lineage-committed progenitors, and further to produce mature blood cells (Figure 1 A). In many cases, these progenitor cells are still multipotential cells in terms of their differentiation capacities but have significantly reduced self-renewal ability. A hematopoietic stem cell not only gives rise to billions of mature blood cells but also to further copies of the stem cell that will proceed to differentiate later in life. An additional feature of the hematopoietic stem/progenitor hierarchy is that progenitor cell numbers are directly correlated with the degree of commitment of the cells. Mice stem cells divide at least once during a period of 3 – 4 weeks, which suggests that all stem cells participate in steady-state blood cell production. In young mice, only 10% of the LT-HSCs enter the cell cycle each day. When more hematopoietic stem cells differentiate into multipotent progenitors their number in the cell cycle increases until almost all cells are in cycle in old adult mice. The decisions whether, when, and how to differentiate are supposed to be mediated by certain gene products, most likely regulatory molecules, that are preferentially expressed in the stem cell itself.

Lemischka and co-workers have definitely taken a major step toward identifying the "genetic blueprint" or the "molecular signature" of stem cells.[9, 14] When analyzing several kinds of stem cells beside those of the blood, including embryonic stem cells and those of the nervous system, they identified a core set of genes that is shared by all of these cells. In a second comprehensive study they compared the molecular profile of a fetal hematopoietic stem cell to that of its adult counterpart and asked: 1) which part of the entire genetic program of a mouse is turned on in each of the different populations of hematopoietic stem cells, primarily in long-term repopulating cells (LT-HSC) and secondarily in various, putatively distinct stages of the hematopoietic developmental hierarchy (ST-HSC and LCP)? and 2) which genes are responsible for the unique properties of these cells?

To isolate and enrich different types of stem cells, Lemischka and co-workers performed a selective depletion of the tissue by sequential cell sorting based on surface antigen markers specific to each

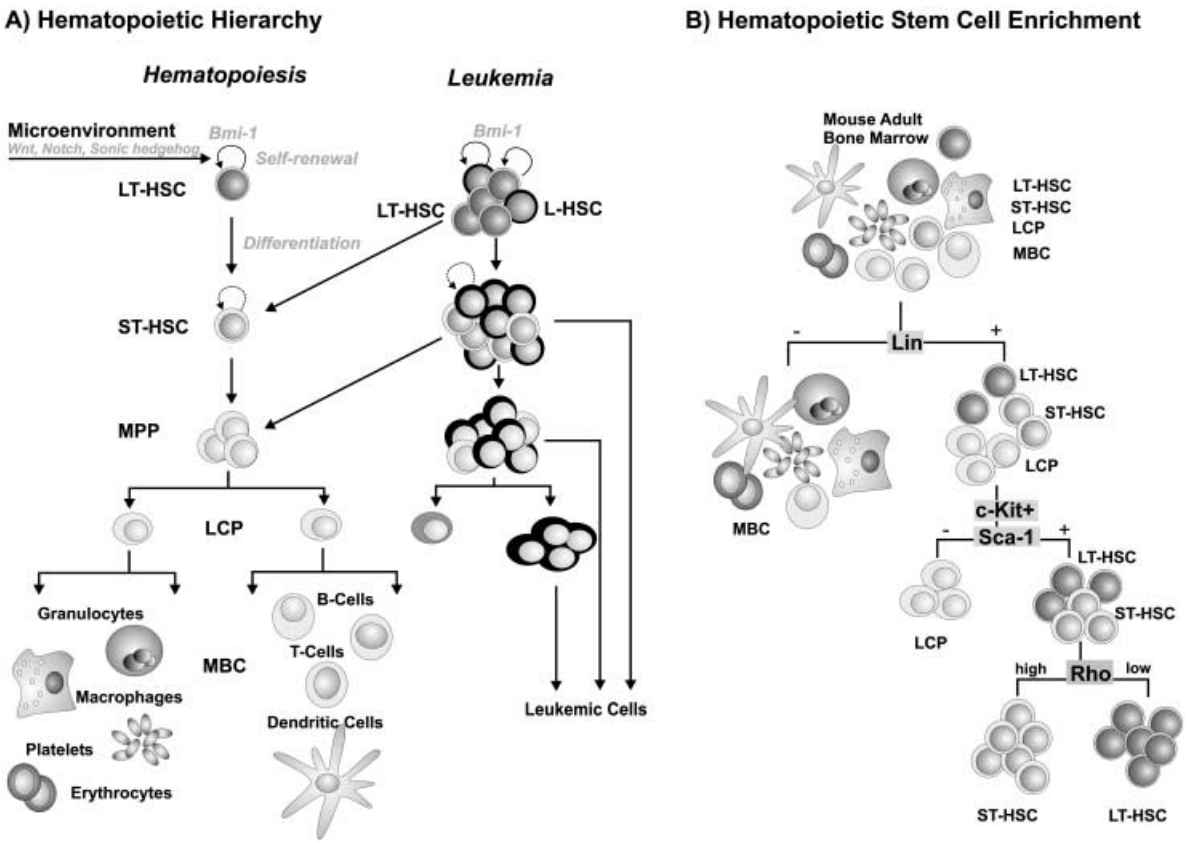

Figure 1. *A) Scheme of the hematopoietic and leukemic stem cell hierarchy. Long-term repopulation hematopoietic stem cells (LT-HSC) as well as leukemic stem cells (L-HSC) are characterized by their high capacity to self-renew, whereas all differentiated stem cells, such as the short-term hematopoietic stem cells (ST-HSC), the multipotent progenitors (MPP), and the lineage-committed progenitors (LCP) are restricted in their self-renewal abilities. Only some of the mature blood cells (MBC), which are generated from LCPs, are listed in this scheme. B) Flowchart of stem cell isolation from adult bone marrow by FACS Sorting using stem-cell-specific surface markers. Important antigenic markers are Sca-1 (stem cell antigen) and the c-Kit tyrosine kinase receptor. Sorting of fetal-liver stem cells benefits from their expression of the AA4 antigen, and bone marrow stem cells have been shown to take up the dye Rhodamine-123 rather poorly (Rho^{low}).*

stem cell population (Figure 1B). Gene expression (complementary DNA) libraries were prepared by reverse transcription and linear amplification of all mRNAs from fetal murine liver and from the other murine and human tissues mentioned above. These libraries were hybridized with Affymetrix oligonucleotide arrays, which so far allow screening of about 80% of HSC-related gene products (Figure 2). The approach is far more comprehensive than the common methods, which usually rely on gene mutations in patients or animals with stem cell disorders.

Simultaneously with Douglas Melton and co-workers at Harvard, who published their study on the molecular fingerprint of hematopoietic stem cells in the same *Science* issue,[15] Lemischka and coworkers identified more than 4000 genes that are upregulated specifically in stem cells. After sophisticated bioinformatic analysis and comparison of the hybridization profiles, these genes were assigned to distinct clusters according to their expression patterns within the hematopoietic hierarchy. By using this kind of DNA microchip technology it was possible to identify virtually all of the different possible combinations and overlaps of genetic expression upregulated products. The products of several hundred differentially expressed genes were assigned to a putative function, mostly as regulatory molecules like transcription factors, DNA binding proteins, and cell surface molecules. In addition, Lemischka and coworkers uncovered a lot of potentially regulatory molecules that communicate signals from the outside of the cell to the inside, including many molecules involved in known signal transduction pathways (Figure 2). Since many of the signals are secretory molecules from the surrounding mature cells, one might expect that they are the key players that mediate the decision for commitment or self-renewal. Nevertheless, the majority of genes have not been described previously and their potential roles as candidate players for stem cell biology have still to be determined.

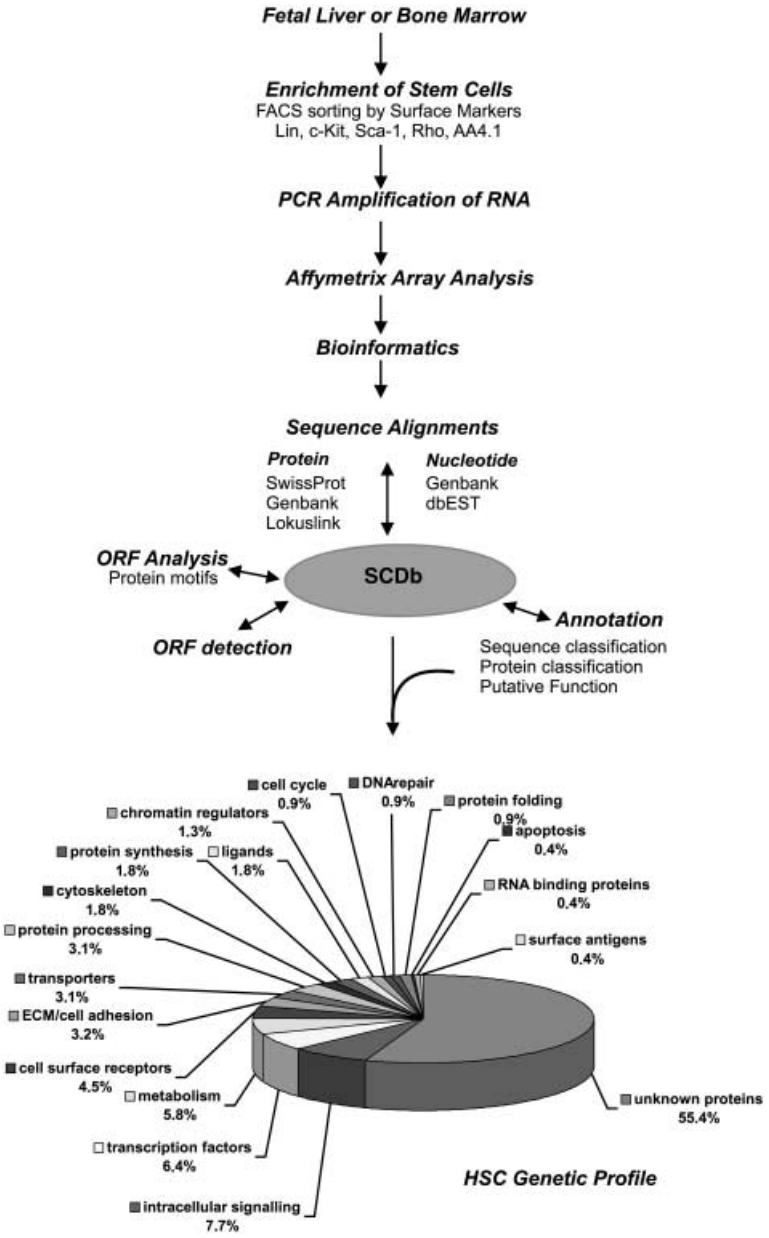

Figure 2. *Flowchart depicting the large-scale genetic profiling of hematopoietic stem cells. Microarray sequence results were evaluated for their encoding open reading frames (ORFs) and were stored in an annotated database. Novel ORFs were analyzed for protein motifs and both novel and known ORFs were assigned to a putative function. The pie diagram shows the categorization of the informative sequences by their putative function. SCDb, Princeton Stem Cell Database; dbEST, database of expressed sequence tags.*

About 56% of the genes were previously identified by a different screening strategy for hematopoietic stem cell specific gene products, which is strong evidence for their functionality in stem cell biology.[14] Comparison of fetal versus adult hematopoietic stem cells revealed more than a 70% overlap of the genetic fingerprint. However, the largest discrepancy is found by comparison of cell-cycle-related genes. It looks like fetal stem cells specifically express more than 30% additional genes grouped into this category. Despite the enormous overlap, one still has to unravel the function of these genes before one can start to talk about the equivalence of adult versus fetal or embryonic stem cells. The same kind of analysis was done to compare human and mouse fetal hematopoietic stem cells. One could argue that there is no need to work with human embryonic stem cells when the mouse, in terms of the hematopoietic system, shows the same genetic background or properties. Moreover, ethical considerations make one hesitate to use human embryos for stem cell research. Surprisingly, only 39% of the

human homologues to the murine hematopoietic stem cell related genes were enriched in human cells. These genes might be responsible for the regulatory aspects of human and mouse, whereas the remaining 60% are probably unique to the mouse system. As a result of these observations, research on human stem cells is far from becoming obsolete. The most remarkable result however is the fact that only 283 genes are shared by all analyzed types of stem cells, such as hematopoietic, neural, and embryonic cells (Figure 3). This common genetic profile, which seems to give stem cells their unique properties, is termed the "stem cell molecular signature" (Lemischka) or "stemness" (Melton). The big challenge is now to assemble these comprehensive genetic programs into complex regulatory pathways and functional networks. Rather the similarities than the differences in these programs need to be understood to explain the properties that define a stem cell.

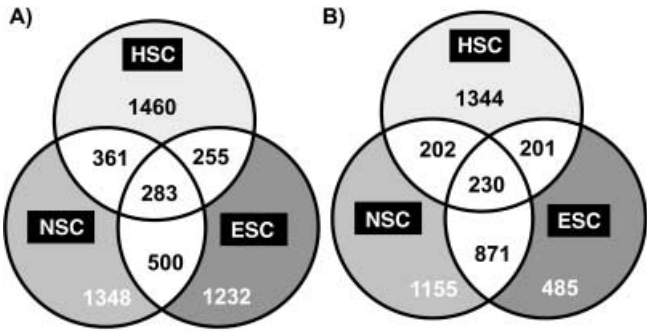

Figure 3. *Venn diagrams of the two large-scale microarray hybridization approaches published in* Science: *A) Lemischka and co-workers, B) Melton and co-workers.[9, 15] These diagrams illustrate the overlapping gene expression profiles in hematopoietic (HSC), neural (NSC), and embryonic stem cells (ESC) obtained by the two groups.*

Nevertheless, understanding stem cells will not give the answers to all questions. What is at least as important is the local microenvironment, which influences their fate. In this context, one also has to consider the role of the surrounding tissue, which communicates with the different populations of stem cell by secretion of signaling molecules. The composition of these molecules allows the cells to make decisions about whether to differentiate, to self-renew, or even to become a certain type of stem cell such as hematopoietic or neural. Another Princeton group led by Kateri Moore focused on stem cell surrounding cells, known as the stem cell microenvironment. They analyzed more than 200 cell lines derived from mouse fetal liver and documented the activity of more than 4000 genes in the stem cell supportive cells, but not in the others.[16] By combining these two complementary approaches, the analysis of gene expression profiles of the stem cells and the supportive cells will help to unravel the mysteries of how stem cells behave and what influences plasticity. However, it will take a lot of interactive experiments to uncover the function of each gene and its role in regulatory pathways.

Most of the genetic information that Lemischka and co-workers have collected is published online in the Princeton Stem Cell Database (SCDb) as an interactive web resource[17]. This database is freely available to anyone in the scientific community and covers various types of analytical tools to look up the properties of each stem cell or microenvironment-specific gene product. The classification of proteins into functional groups was done by homology comparisons with a wide variety of protein families. The decision as to whether a clone is a novel homologue of a known protein or a novel member of a particular protein family depends on the degree of homology outside the conserved protein motifs.

In the meantime, many groups have used these data or started their own high-throughput genetic profiling analyses to find genes involved either in blood stem cell differentiation and self-renewal or in certain types of cancer such as leukemia or breast cancer, which are now thought to be stem-cell related. Therefore, understanding the factors that regulate between self-renewal and differentiation means getting new evidence about the proliferation of cancer, where self-renewal is assumed to get out of control. Very recently, some factors that influence self-renewal not only in embryonic development but also in mature tissue have been found.[11, 12, 18, 19] These include either factors from the microenvironment that can act on the gene expression in stem cells, such as the Wnt proteins,[18, 19] Sonic hedgehog,[20] and proteins of the notch family,[21] or genes in the stem cell itself like *Bmi-1*.[11, 12] In contrast to most of the other growth factors that have been used so far in adult hematopoietic stem cell culture, Wnt proteins can induce unlimited expansion and prevent differentiation, an effect which was also observed on stem cells of other origin, such as the skin,[22] the gut,[23, 24] and the brain.[25] This discovery is expected to revolutionize adult stem cell biology since the low self-renewal rate of adult stem cells hinders research and their application in therapeutic approaches. An understanding of the mechanism underlying self-renewal would possibly allow expansion of adult stem cells, which would make the use of embryonic stem cells unnecessary.

Surprisingly, Lemischka and co-workers did not pick up *Bmi-1*, a proto-oncogene that is highly expressed in hematopoietic stem cells and was very recently shown to be responsible for stem-cell renewal in the stem cell itself. This candidate gene was found in another genetic profiling screen performed by Clarke and co-workers,[12, 26] which shows that the data set of one screen is far from being complete. Simultaneously to Lessard and Sauvageau,[11] Clarke and co-workers found that in *Bmi-1* deficient mice, hematopoietic stem cells differentiate but do not self-renew. By transplanting stem cells from these mice into wild-type mice they could detect a normal pattern of blood cells derived from *Bmi-1* deficient stem cells, but their presence was only temporary because of an inability to maintain the LT-HSC population. Moreover, Sauvageau and co-workers showed that *Bmi-1* plays an important role in certain types of leukemia, such as acute myeloid leukemia. Although leukemia is the most common malignancy in children aged 15 years and younger, most cases of leukemia are diagnosed in adults, which gives rise to the speculation that leukemia is a cancer that correlates with the remaining number of self-renewing stem cells.[27] It

has been proposed that the initial, cancer-causing mutations in leukemia, which often originate after exposure to chemotherapy, radiation, solvents, or alkylating agents, occur in the hematopoietic stem cell population rather than in mature cells, which leads to leukemic hematopoietic stem cells that still retain remnants of the normal stem cell properties (Figure 1 A). Differentiation of this rare L-HSC population to generate mature leukemic cells then follows a similar hierarchy to that of normal blood cells.[27] Since self-renewal and differentiation is out of balance in leukemic cells, L-HSCs can even overgrow normal hematopoietic stem cells. As a result, fewer mutations would be required to generate fully leukemic cells if they originated from primitive stem cells as opposed to progenitors or mature cells. Lessard and Sauvageau showed with their transplantation assays with *Bmi-1* deficient mice that *Bmi-1* is not necessary for the onset and progress of leukemia but is absolutely essential for the maintenance and self-renewal of this rare population of leukemic stem cells, which expand and replenish the tumor. L-HSCs lacking *Bmi-1* can still give rise to many mature leukemic blood cells but fail to renew themselves, which leads to a depletion of the leukemia-replenishing stem cell pool. These and other recently published data support the model in which the cellular targets for transformation reside neither in the lineage-committed progenitors nor in the mature blood cells but in the most primitive stem cells. Heterogeneity of symptoms or the progression observed in human leukemia supposedly depends on the direct influence of the microenvironment.[28] Similar results were observed in studies on breast cancer. There are first hints that the solid tumors of breast cancer may be initiated by transformations in mammary gland stem cells that lead to breast cancer stem cells that follow a similar hierarchical differentiation to that of normal blood cells to produce the mature breast cancer tumor cells.[29]

Despite the stunning characterization of these genes involved in stem cell renewal, there are still a huge number of genes that appear to be common to or differentially expressed in different stem cell populations; such differential expression on single stem cell clones has not been validated so far. The group of Harvey Lodish at the Whitehead Institute in Cambridge even claims that no applicable markers have resulted from these screens that could be used to ensure the purification of a certain stem cell subpopulation, such as long-term repopulating hematopoietic stem cells.[30] The genetic screens have yet to be completed so finding surface markers is still not becoming obsolescent. Purification of single stem cells from tissue and their clonal proliferation is still the best way to finally analyze stem cell plasticity, especially when one looks at bone-marrow-derived stem cells, which have often been observed to be contaminated with cells that came from somewhere else.

In addition, there is still an enormous amount of work to be done in terms of the functional biology of the stem cell genes. We are still far away from addressing this issue and even further away from the conclusion that fetal and adult cells share equivalent properties and are clinically compatible. Gene expression profiling is only the beginning, but combined with proteomics and histochemical staining it will ultimately lead to the unravelling of multiple pathways and networks in stem cell biology. This approach will certainly offer scientists a powerful resource for understanding some of the stem-cell-related diseases and will become a basis for new experiments on the therapeutic application of stem cells, such as clonal proliferation of purified stem cells, tissue replacement, in vivo transplantation, or gene therapy.

[1] C. R. Bjornson, R. L. Rietze, B. A. Reynolds, M. C. Magli, A. L. Vescovi, *Science* **1999**, *283*, 534.
[2] C. M. Verfaillie, M. F. Pera, P. M. Lansdorp, *Hematology* **2002**, 369.
[3] N. Terada, T. Hamazaki, M. Oka, M. Hoki, D. M. Mastalerz, Y. Nakano, E. M. Meyer, L. Morel, B. E. Petersen, E. W. Scott, *Nature* **2002**, *416*, 542.
[4] Q. L. Ying, J. Nichols, E. P. Evans, A. G. Smith, *Nature* **2002**, *416*, 545.
[5] C. N. Shen, J. M. Slack, D. Tosh, *Nat. Cell. Biol.* **2000**, *2*, 879.
[6] Y. Jiang, B. N. Jahagirdar, R. L. Reinhardt, R. E. Schwartz, C. D. Keene, X. R. Ortiz-Gonzalez, M. Reyes, T. Lenvik, T. Lund, M. Blackstad, J. Du, S. Aldrich, A. Lisberg, W. C. Low, D. A. Largaespada, C. M. Verfaillie, *Nature* **2002**, *418*, 41.
[7] A. J. Wagers, R. I. Sherwood, J. L. Christensen, I. L. Weissman, *Science* **2002**, *297*, 2256.
[8] I. Lemischka, *Nat. Biotechnol.* **2002**, *20*, 425.
[9] N. B. Ivanova, J. T. Dimos, C. Schaniel, J. A. Hackney, K. A. Moore, I. R. Lemischka, *Science* **2002**, *298*, 601.
[10] J. E. Dick, *Nature* **2003**, *423*, 231.
[11] J. Lessard, G. Sauvageau, *Nature* **2003**, *423*, 255.
[12] I. K. Park, D. Qian, M. Kiel, M. W. Becker, M. Pihalja, I. L. Weissman, S. J. Morrison, M. F. Clarke, *Nature* **2003**, *423*, 302.
[13] J. Domen, I. L. Weissman, *Mol. Med. Today* **1999**, *5*, 201.
[14] R. L. Phillips, R. E. Ernst, B. Brunk, N. Ivanova, M. A. Mahan, J. K. Deanehan, K. A. Moore, G. C. Overton, I. R. Lemischka, *Science* **2000**, *288*, 1635.
[15] M. Ramalho-Santos, S. Yoon, Y. Matsuzaki, R. C. Mulligan, D. A. Melton, *Science* **2002**, *298*, 597.
[16] J. A. Hackney, P. Charbord, B. P. Brunk, C. J. Stoeckert, I. R. Lemischka, K. A. Moore, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 13061.
[17] http://stemcell.princeton.edu
[18] T. Reya, A. W. Duncan, L. Ailles, J. Domen, D. C. Scherer, K. Willert, L. Hintz, R. Nusse, I. L. Weissman, *Nature* **2003**, *423*, 409.
[19] K. Willert, J. D. Brown, E. Danenberg, A. W. Duncan, I. L. Weissman, T. Reya, J. R. Yates, R. Nusse, *Nature* **2003**, *423*, 448.
[20] G. Bhardwaj, B. Murdoch, D. Wu, D. P. Baker, K. P. Williams, K. Chadwick, L. E. Ling, F. N. Karanu, M. Bhatia, *Nat. Immunol.* **2001**, *2*, 172.
[21] B. Varnum-Finney, L. Xu, C. Brashem-Stein, C. Nourigat, D. Flowers, S. Bakkour, W. S. Pear, I. D. Bernstein, *Nat. Med.* **2000**, *6*, 1278.
[22] U. Gat, R. DasGupta, L. Degenstein, E. Fuchs, *Cell* **1998**, *95*, 605.
[23] V. Korinek, N. Barker, P. Moerer, E. van Donselaar, G. Huls, P. J. Peters, H. Clevers, *Nat. Genet.* **1998**, *19*, 379.
[24] M. van de Wetering, E. Sancho, C. Verweij, W. de Lau, I. Oving, A. Hurlstone, K. van der Horn, E. Batlle, D. Coudreuse, A. P. Haramis, M. Tjon-Pon-Fong, P. Moerer, M. van den Born, G. Soete, S. Pals, M. Eilers, R. Medema, H. Clevers, *Cell* **2002**, *111*, 241.
[25] A. Chenn, C. A. Walsh, *Science* **2002**, *297*, 365.
[26] I. K. Park, Y. He, F. Lin, O. D. Laerum, Q. Tian, R. Bumgarner, C. A. Klug, K. Li, C. Kuhr, M. J. Doyle, T. Xie, M. Schummer, Y. Sun, A. Goldsmith, M. F. Clarke, I. L. Weissman, L. Hood, L. Li, *Blood* **2002**, *99*, 488.
[27] T. Reya, S. J. Morrison, M. F. Clarke, I. L. Weissman, *Nature* **2001**, *414*, 105.
[28] J. E. Dick, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3547.
[29] M. Al-Hajj, M. S. Wicha, A. Benito-Hernandez, S. J. Morrison, M. F. Clarke, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3983.
[30] C. Z. Chen, M. Li, D. de Graaf, S. Monti, B. Gottgens, M. J. Sanchez, E. S. Lander, T. R. Golub, A. R. Green, H. F. Lodish, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 15468.

Received: May 16, 2003 [H661]

