The molecular signature of MDS stem cells supports a stem cell origin of 5q- myelodysplastic syndromes

The molecular signature of MDS stem cells supports a stem cell origin of 5q- myelodysplastic syndromes

Lars Nilsson¹, Patrik Edén²*, Eleonor Olsson³*, Robert Månsson¹, Ingbritt Åstrand-Grundström¹, Bodil Strömbeck⁴, Kim Theilgaard-Mönch⁵, Kristina Anderson¹, Robert Hast⁶, Eva Hellström-Lindberg⁷, Jan Samuelsson⁸, Gösta Bergh⁹, Claus Nerlov¹,¹⁰, Bertil Johansson⁴, Mikael Sigvardsdson¹, Åke Borg³, and Sten Eirik W. Jacobsen¹

*These authors contributed equally to this work.

¹Hematopoietic Stem Cell Laboratory, Lund Strategic Research Center for Stem Cell Biology and Cell Therapy, Sweden

²Complex Systems Division, Department of Theoretical Physics, Lund University, Sweden

³DNA Microarray Resource Center, Department of Oncology, Lund University, Sweden

⁴Department of Clinical Genetics, Lund University, Sweden

⁵Granulocyte Research Laboratory, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

⁶Hematology Center, Karolinska University Hospital, Stockholm, Sweden

⁷Department of Medicine, Division of Hematology, Karolinska University Hospital (Huddinge), Stockholm, Sweden

⁸Department of Medicine, South Hospital, Karolinska Institute, Stockholm, Sweden

⁹Department of Medicine, Helsingborg Hospital, Sweden

¹⁰EMBL, Mouse Biology Unit, Via Ramarini 32, 00015 Monterotondo, Italy

Short title: Gene profiling of 5q- syndrome stem cells

Corresponding author: Sten Eirik W. Jacobsen, Lund Strategic Research Center for Stem Cell Biology and Cell Therapy, BMC B10, Sweden, Ph+46 46 2223082, Fax+46 46 2223600 E-mail: sten.jacobsen@med.lu.se

ABSTRACT

Global gene expression profiling of highly purified 5q deleted CD34⁺CD38⁻Thy1⁺ cells in 5q- myelodysplastic syndromes (MDS) supported that they might originate...
from and outcompete normal CD34^+CD38^-Thy1^+ hematopoietic stem cells. Few but distinct differences in gene expression distinguished MDS and normal stem cells. Expression of *BMI1*, encoding a critical regulator of self-renewal was up-regulated in 5q- stem cells. Whereas multiple previous MDS genetic screens failed to identify altered expression of the gene encoding the myeloid transcription factor *CEBPA*, stage specific and extensive down-regulation of *CEBPA* was specifically observed in MDS progenitors. These studies establish the importance of molecular characterization of distinct stages of cancer stem and progenitor cells to enhance the resolution of stage-specific dysregulated gene expression.
Introduction

Although the existence of cancer stem cells (CSCs), in particular leukemic stem cells (LSCs), has been established for more than a decade\(^1\), fundamentally important questions remain unanswered, relating to the exact identity and normal cellular origin of human CSCs, knowledge likely to have major impact towards a better understanding of the evolution, prognosis and therapeutic targeting of CSCs.

Hematopoietic stem cells (HSCs) possess life long self-renewal capacity, and are therefore at high risk of acquiring the multiple mutations thought to be required for ultimate leukemic transformation. However, conclusive evidence for a normal HSC origin of LSCs would require proof of clonal involvement of all blood cell lineages derived from the multipotent HSCs. However, as yet, it has only been possible to obtain direct evidence for multilineage involvement in myeloproliferative disorders\(^3\),\(^4\). Although this could mean that leukemias rarely originate in normal HSCs or multipotent progenitors, it is as possible that it in most cases is not feasible to prove through this approach, as most leukemias are lineage restricted in nature. Specifically, it is likely that a transforming event in a HSC will not only promote the development of the dominating leukemic lineage, but simultaneously also suppress the development of other lineages\(^5\).

An alternative and novel approach would rather be to apply genomics to better identify the origin of LSCs. As normal HSCs have been highly purified and found to have a distinct gene expression profile when compared to normal progenitors\(^6\),\(^7\), a similar approach should be possible to better establish whether LSCs have a stem or progenitor cell identity. It would however require the identity of the LSC population to be well established, as well as conclusive evidence for a high clonal involvement of the purified LSCs to be investigated. So far, these criteria would be fulfilled only in
rare cases, one being myelodysplastic syndromes (MDS) or preleukemia, a group of clonal malignant hematopoietic disorders, which frequently progress to acute myeloid leukemia (AML)\(^8\). In patients with 5q- syndrome, a distinct clinical subgroup of MDS, characterized by an isolated deletion of the long arm of chromosome 5 (del(5q) or 5q\(^-\))\(^9\), we recently demonstrated that the 5q- stem cells resides in the minor CD34\(^+\)CD38\(^-\)Thy1\(^+\) compartment, with definitive evidence for clonal myeloid as well as B lymphoid involvement in some patients, suggesting that 5q- syndrome at least in part originates at a multipotent stem/progenitor cell level\(^10,11\). However, in most patients no evidence was obtained for B cell involvement and T cell involvement could never be established.

In the present studies we purified and performed global gene expression profiling of normal and 5q- CD34\(^+\)CD38\(^-\)Thy1\(^+\) cells\(^12,13\), typically representing 0.1% of the total MDS bone marrow (BM) cells, and CD34\(^+\)CD38\(^-\)Thy1\(^-\) progenitors. This approach allowed us in all investigated cases to implicate the 5q- syndrome as a likely true HSC disease, and to establish stage specific dysregulated gene expression.

**Patients, material, and methods**

**Patient samples**

BM samples from 11 MDS patients with typical 5q- syndrome\(^8\) (Table 1) and from 10 healthy subjects were collected at the Hematology Departments at Karolinska (Solna), Karolinska (Huddinge), South Hospital, Helsingborg, and Lund University Hospitals, Sweden. The investigation was approved by the Research Ethics Committees at the respective University Hospitals, and informed consent was provided according to the Declaration of Helsinki. The MDS patients were chosen based on the finding of an isolated del(5q) by conventional cytogenetic analysis in combination with clinical and morphologic characteristics typical for the 5q-syndrome (an indolent clinical course with macrocytic anemia and a variable need for red blood cell transfusions, normal to elevated platelet counts, normal to slightly
reduced WBC counts, hypolobulated megakaryocytes, and less than 5% blasts in the BM) resulting in MDS RA (FAB) and a 5q- syndrome (WHO) diagnosis and a low risk score according to the International Prognostic Scoring System (IPSS), for included patients. Patients #2, #6 and #9 had been treated unsuccessfully with EPO, patient #10 and #11 had ongoing EPO-treatment and six of 11 patients needed erythrocyte transfusions regularly (Table 1).

**Purification of BM cell populations**

BM MNCs were isolated by Lymphoprep® (Nycomed, Oslo, Norway) gradient centrifugation. Positive selection of CD34+ BM cells was performed using a MACS (magnetically activated cell sorting) CD34 isolation kit (Miltenyi Biotec, Bergish Gladbach, Germany) as previously described11. The mean purity of enriched CD34+ cells was 81%. Normal and 5q- MNC or CD34+ cells were cryopreserved in 10% dimethylsulfoxid (DMSO) (Merck, Darmstadt, Germany) and 50% fetal calf serum (FCS; BioWhittaker, Walkersville, MD), and thawed swiftly in 37°C water bath, washed twice in Dulbecco’s phosphate buffered saline (PBS; PAA Laboratories, Pasching, Austria), and 5% FCS before staining. The CD34 enriched cells were incubated with CD38-allophycocyanin (APC), CD34-fluorescein isothiocyanate (FITC), and Thy-1 (CD90)-phycoerythrin (PE) monoclonal antibodies. The 30% CD34+ cells expressing the highest levels of CD38 and lacking expression of Thy-1 (CD34+CD38+Thy-1-) and the 5% CD34+ cells with the lowest expression of CD38 and co-expression of Thy-1 (CD34+CD38-Thy-1+) were sorted on a FACSDiVa (Becton Dickinson). All samples were stained with 7-amino-actinomycocin D (7-AAD; Sigma, St Louis, MO) to exclude non-viable cells. Both sorted populations reproducibly had a purity of more than 98% with regard to all three antigens. All antibodies were from BD Pharmingen (San Jose, CA) unless otherwise indicated.

**Hematopoietic growth factors**

Recombinant human (rh) granulocyte colony-stimulating factor (G-CSF), rh stem cell factor (SCF), rh interleukin-3 (IL-3), and rh granulocyte-macrophage colony-stimulating factor (GM-CSF) were generously provided by Amgen Corp. (Thousand Oaks, CA). Rh erythropoietin (Epo) was supplied by Boehringer Mannheim Corp. (Mannheim, Germany), thrombopoietin (Tpo) by Genentech (San Francisco, CA), and rh flt3 ligand (FL) by Immunex (Seattle, WA).
Long-term culture-initiating cell (LTC-IC) assay

Murine stromal feeders engineered to produce human growth factors (M2-10B4 and Sl/Sl mixed 1:1; kindly provided by Dr D.E. Hogge, Vancouver, Canada) were used to support growth of LTC-ICs as previously described\textsuperscript{11}. Cultures were established in 96-well collagen-coated microtiter plates with 5,000 cells/well of each cell line, after irradiation with 8000 cGy, and cultured in long-term culture medium (MyeloCult H5100; Stem Cell Technologies) with $10^{-6}$ M hydrocortisone 21-hemisuccinate. 75-750 CD34$^+$/CD38$^+$Thy-1$^-$ and 75-750 CD34$^+$/CD38$^-$Thy-1$^+$ cells from MDS patients and healthy subjects were added to the stroma layers, and co-cultures were maintained at 37°C in high humidity and with 50% medium exchange every week, as previously described (4). After 6 weeks, non-adherent and adherent cells were plated in methylcellulose cultures, supplemented with SCF, GM-CSF, G-CSF, FL, IL-3 (all at 10 ng/mL), and Epo (5 U/mL). Colony-forming cells (CFCs; read-out of LTC-IC assay) were scored after an additional 12 days in culture. Individual colonies were then picked and transferred to slides for subsequent FISH analysis.

FISH probes and analyses

Interphase FISH analyses, using probes hybridizing to 5q31 [SpectrumOrange LSI EGRI] and 5p15.2 [SpectrumGreen LSI D5S721:D5S23] were performed essentially as described previously\textsuperscript{10}. All probes were obtained from Abbott (Stockholm, Sweden), and the signals were analyzed with the Chromofluor System (Applied Imaging, Newcastle, UK). The number of nuclei analyzed varied depending on the number of available cells, but whenever possible at least 200 nuclei were analyzed per cell population. In the nuclei of normal cells, the probes appear as 4 distinct signals, 2 orange and 2 green, whereas patients with del(5q) typically show 1 orange and 2 green signals. Based on FISH analyses on control cytospin preparations, the cut-off value (median + 2SD) for del(5q) was 6.2%. However, taking the purity of the sorted populations into account (>98%), findings of less than 10% del(5q) were considered to be negative or inconclusive.

Purification of total RNA

Normal and MDS CD34$^+$/CD38$^+$Thy-1$^-$ and CD34$^+$/CD38$^-$Thy-1$^+$ cells were sorted directly into RLT lysis buffer (Qiagen, Hilden, Germany) and snap frozen at -80°C immediately after the sorting procedure. Samples were then thawed and, after homogenization by vortexing, total RNA was purified.
from $1 \times 10^4$ CD34$^+$CD38$^-$Thy-1$^-$ and CD34$^+$CD38$^+$Thy-1$^+$ cells with RNeasy Micro Kit (Qiagen) following the manufacturer’s protocol for isolation of RNA from animal cells including the optional on-column DNase treatment.

**Double linear amplification of total RNA and probe preparation**

Before amplification, 200 ng of Poly (dI-dC) (Sigma, St Louis, MO) was added as nucleic acid carrier to each sample of total RNA. The total RNA isolated from $1 \times 10^4$ CD34$^+$CD38$^-$Thy-1$^-$ and CD34$^+$CD38$^+$Thy-1$^+$ cells was double linear amplified with a combination of RiboAmp OA RNA Amplification Kit (Arcturus, Mountain View, CA) for the first round of amplification and the following generation of double stranded cDNA, and Low RNA Input Flourescent Linear Amplification Kit (Agilent, Palo Alto, CA) to generate the second round amplified labeled aRNA, following the manufacturer’s protocols. A pool of double linear amplified labeled Universal Human Reference RNA (Stratagene, La Jolla, CA) was used as reference to all samples at the hybridization. Samples were labeled with Cy3-CTP (PerkinElmer, Boston, MA) and reference reactions with Cy5-CTP (PerkinElmer). A nanodrop Spectrophotometer (NanoDrop Technologies, DE) was used to determine the dye concentration, and appropriate amounts of sample and reference aRNA were combined, lyophilized to dryness in an Eppendorf Concentrator (Eppendorf AG, Germany) and stored at -80°C.

**Preparation and printing of oligonucleotide microarrays**

Oligonucleotide microarrays were produced by the Swegene DNA Microarray Resource Center, Department of Oncology, Lund University, Sweden (http://swegene.onk.lu.se; Swegene Center Home page). Array ready oligolibraries Human Genome Oligo Version 2.1 (containing 21,329 70mer probes, Cat. No. 810518) and Human Genome Oligo Set Version 2.1 Upgrade (containing 5,462 70mer probes, Cat. No. 810518) were obtained from Operon (Biotechnologies Inc., Alabama). Lyophilized probes were re-suspended in Pronto! Universal Spotting solution (Corning Incorporated, Corning, NY) to a concentration of 24 mM. The entire probe set, in addition to a number of positive and negative controls, was printed in duplicate on aminosilane coated UltraGAPS slides (Corning Incorporated) using a BioRobotics MicroGrid2 R600 robot (Genomic Solutions, MI) equipped with MicroSpot 10K quill pins (Genomic Solutions, MI). Printing was performed in a temperature (18-20°C) and humidity (44-
49% RH) controlled area, and the printed slides were left in a vacuum dessicator to dry for at least 48 hours before use.

**Hybridization, image processing and image acquisition**

Microarray slides were re-hydrated over steaming water for 1-2 seconds, snap dried on a hot plate (98°C), and then UV cross-linked (800mJ/cm²) using a Stratalinker (Stratagene). Pre-hybridization treatment, hybridization, and post-hybridization washes of slides were performed according to the manufacturers’ protocols provided with the Universal Microarray Hybridization Kit (Corning Incorporated). In short, the array slides were pre-soaked with a sodium borohydride reducing solution, pre-hybridized with a BSA containing solution, washed, and dried by centrifugation. Prepared labeled aRNA was re-suspended in hybridization solution and incubated at 65°C for 5 minutes and then cooled to ambient temperature before it was applied to the array and covered with a glass cover slip. The array slides were hybridized in Corning hybridization chambers (Corning Incorporated) at 42°C during 17-20 hours, washed, and finally dried by centrifugation.

The Agilent G2565AA Microarray Scanner (Agilent Technologies) was used to measure the fluorescence intensities at PMT gain 100%, and at 10 um pixel resolution, data were collected at two different wavelengths (for Cy3 and Cy5) and stored as multi TIFF images and analyzed by using the GenePix™ Pro 4.1.1 software (Axon Instruments Inc., Foster City, CA) with standard flagging criteria. Quantified data matrix from GenePix was saved as a GenePix Results File (gpr) and loaded into the Bio Array Software Environment (BASE) for further data analysis. Background subtraction for Cy3 and Cy5 intensities were calculated using the median spot pixel intensity and median local background intensity provided in the GenePix result file.

**Data extraction**

In an initial spot quality filter, spots flagged by the criteria above, with a diameter of 40 pixels or less, an intensity of 0 units or less in either channel, or 10% or more saturated pixels in either channel, were removed. Intensity dependent lowess\(^{14}\) fits were used to normalize intensity ratios on each assay. For each spot, the uncertainty of the expression value was estimated as 

\[ u = (SNR_1)^2 + (SNR_2)^2 \]

where \( SNR_i \) is the signal to background noise ratio for channel \( i \). Replicate measurements of the same reporter...
on an assay were merged and represented by a weighted mean. The weighted mean of a set of values $x_i$ was defined as $m = \frac{\sum w_i x_i}{\sum w_i}$ where the weight $w_i$ is $\exp(-3u_i^{1/2}/|x_i-m|)$. This set of equations was solved numerically by simple iteration. The error of the merged value was defined as $U = \frac{1}{\sum (1/u_i) + \sum w_i^2 (x_i-m)^2/(\sum w_i)^2}$. All reporters with GenBank accession number or RefSeq specified were associated to genes using ACID 15. The four different types of assays (normal and 5q-CD34+CD38+Thy-1- progenitors and CD34+CD38-Thy-1+ stem cells) were compared pair-wise in different ways, and for each comparison, the error model, presence filter (requiring 85% presence), and variation filter described above were applied to the assays involved, resulting in an optimal data extraction for each comparison. Association was based on UniGene Homo Sapiens build 176. Expression values for reporters representing the same gene were merged in the weighted fashion described above. Reporters without association to a known gene were excluded from the analysis. The merged data were transformed again to a modified expression value $x_i'=w_i (x_i-m)$. A presence filter and a variation filter across assays were also applied to the data, keeping only spots with expression in both channels in at least 21 of 23 assays, and with a standard deviation of modified expression values greater than 0.3. The 23 assays represented 18 different samples, of which three were done in duplicate and one in triplicate. Hierarchical clustering revealed that replicate assays ended up in the same cluster (data not shown). This means that the experimental variability was small compared to differences between assay types, though not significantly smaller than variations among assays of the same type, which may indicate large homogeneity within the assay types. As the rest of the study concerns comparisons of different types, replicates were merged in the weighted fashion described above before continued analysis.

**Statistical analysis**

As a measure of difference between two types of assays, we used the false discovery rate in ranked gene lists. In each of the pair-wise comparisons described above, the genes that passed the filters were ranked according to the Fisher linear discriminant; $F = (m_1 - m_0) / (\sigma_1^2 + \sigma_0^2)^{1/2}$, where $m_1$ and $m_0$ are the (unweighted) mean values for subgroups 1 and 0, respectively, whereas $\sigma_1$ and $\sigma_0$ are the standard deviations for the same subgroups. A permutation test with all possible permutations of sample labels was performed, and for each score, the average number of genes in a permutation list
above that score was divided by the number of genes in the true list above to get the false discovery rate.

**Quantitative-RT-PCR (Q-PCR)**

10,000 BM CD34"CD38-Thy-1" and CD34"CD38+Thy-1" cells were sorted directly into 330 µL RLT lysis buffer (Qiagen) and snap frozen at -80°C. RNA extraction and DNase treatment were performed with the RNeasy® Micro kit (Qiagen) according to the manufacturer’s instructions for samples containing ≤10⁵ cells. Eluted RNA samples were reverse transcribed using SuperScript II™ and random hexamers (Invitrogen) according to the protocol supplied by the manufacturer. Newly synthesized cDNA was diluted to approximately contain cDNA from 50 cells/µl and frozen at -20°C. Q-PCR reactions were performed by mixing 2×TaqMan® universal PCR master mix, 20× Assays-on-Demand™ (primer/MGB-probe mix), RNase-free H2O, and 5µl of cDNA to a final reaction volume of 20 µL. The TaqMan® Assays-on-Demand™ probes used are described in Supplementary methods. All experiments were performed in triplicates, and differences in cDNA input were compensated by normalizing against HPRT expression levels. The TaqMan® Assays-on-Demand™ probes used were AREG: Hs00155832_m1, BMI1: Hs00180411_m1, CDC42SE2: Hs00184113_m1, CEBPA: Hs00269972_s1, CTNNA1: Hs00426996_m1, CTNNB1: Hs00170025_m1, DLK1: Hs00171584_m1, HLF: Hs00171406_m1, HPRT1: Hs99999909_m1, IFITM1: Hs00705137_s1, and TAF7: Hs00538821_s1, (all from Applied Biosystems, California)
Results

Extensive clonal involvement of the CD34+CD38-Thy-1+ HSC compartment and replacement of normal HSCs in 5q- syndrome

The size of the CD34+CD38-Thy1+ HSC compartment in 11 patients with 5q-syndrome9, represented on average 0.14% of total BM cells, only slightly increased over healthy individuals (mean 0.04%; Table 1 and Figure 1). Notably, as much as 92-100% of the CD34+CD38-Thy1+ cells were part of the 5q- clone.

The content of functionally defined normal and MDS long-term culture-initiating cells (LTC-IC), as an assay for stem cell activity16,17, was next investigated in 4 patients. Although MDS BM cells, reflecting MDS being a disease of inefficient hematopoiesis, in almost all cases fail to long-term reconstitute MDS in immune-deficient mice in vivo18, we have previously demonstrated that MDS-initiating LTC-IC activity, when detected, is exclusively contained within the minor CD34+CD38-HSC compartment10,11,18. In agreement with this, in two of the 4 investigated patients with 5q- syndrome we detected LTC-IC activity, and this was exclusively derived from CD34+CD38-Thy1+ cells (Table 2), despite investigating up to 10-fold more CD34+CD38-Thy-1- cells. Noteworthy, all LTC-CFC generated from 5q-CD34+CD38-Thy1+ BM cells were shown to harbor the del(5q) (Table 2). Thus, neither CD34+CD38-Thy1+ nor CD34+CD38-Thy1- cells from 5q- patients displayed detectable normal HSC activity, supporting that the normal HSC compartment in 5q-syndrome patients has been replaced by a clonally expanded 5q- compartment with the same CD34+CD38-Thy1+ phenotype.
Global gene expression profiling implicates that the 5q- syndrome initiates in normal CD34⁺CD38⁻Thy1⁺ HSCs.

Although MDS stem cells share a CD34⁺CD38⁻Thy1⁺ phenotype with normal HSCs\(^1,10,13,19\), this does not necessarily mean the malignancy initiated in normal HSCs, as it might rather reflect acquisition of a stem cell surface phenotype by a transformed MDS progenitor\(^{19}\). Thus, we here used a novel global gene profiling strategy, in an attempt to distinguish better between a stem and progenitor cell identity of the 5q- stem cells, taking advantage of the distinct gene expression patterns of normal HSCs and progenitors\(^6\).

While no previous studies have performed a global gene expression profiling of CD34⁺CD38⁻Thy1⁺ BM cells, Georgantas \textit{et al.} (2004) reported genes significantly up- or down-regulated in normal BM CD34⁺CD38⁻Lin⁻ HSCs, compared to CD34⁺CD38⁺Lin⁺ progenitor cells. We extracted data from our normal CD34⁺CD38⁻Thy-1⁻ and CD34⁺CD38⁺Thy-1⁺ datasets, ranked genes using Fisher linear discriminant scores\(^20\) and found a high degree of agreement with the data set of Georgantas \textit{et al.} (2004) (Supplementary Figure 1), in which 1,190 reporters were found to be significantly up-regulated in CD34⁺CD38⁻Lin⁻ cells. We updated all gene identities for these reporters according to UniGene build 176 [http://www.ncbi.nlm.nih.gov/UniGene/], and found that 276 genes were present also in our list after filtration. Of those, 197 had positive Fisher scores in our data set. Eight of the 276 genes were found among our 20 most up-regulated genes, whereas the other twelve were not mentioned. To further evaluate the overall agreement, using our whole ranked list, we calculated the area under the Receiver Operating Characteristic (ROC)-curve\(^21\), which is a linear transformation of the Wilcoxon rank sum\(^22\). The ROC area for the 276 genes up-regulated in CD34⁺CD38⁺Lin⁻ cells was...
0.73 (Supplementary Fig. 1), with a $p$-value of $10^{-37}$ [random gene permutations]. Similarly, the 1,159 reporters found down-regulated by Georgantas et al. (2004), were associated to genes using UniGene build 176, and 374 of the genes were found in our ranked list as well. Turning our ranked list upside-down, so that down-regulated genes received top-ranks, we found that eleven of our top 20 down-regulated genes were down-regulated in the studies of Georgantas et al. (2004) as well, whereas the other nine were not mentioned. The ROC area was 0.82, with a $p$-value of $10^{-93}$ (Supplementary Fig. 1). The complete list of genes differentially expressed between normal CD34$^+$CD38$^+$Thy-1$^+$ and CD34$^+$CD38$^+$Thy-1$^-$ cells is available in Supplementary Table 1.

We next compared the global expression profiles of CD34$^+$CD38$^+$Thy-1$^+$ MDS stem cells from four 5q- syndrome patients (#1, #2, #3, and #5), with those of CD34$^+$CD38$^+$Thy-1$^+$ and CD34$^+$CD38$^+$Thy-1$^-$ cells from 5 healthy subjects. Notably, the number of differentially expressed genes was considerably less when comparing 5q- CD34$^+$CD38$^+$Thy-1$^+$ cells to normal CD34$^+$CD38$^+$Thy-1$^+$ cells, than when comparing to normal or 5q- CD34$^+$CD38$^+$Thy-1$^-$ progenitors (Figure 2). Furthermore, the Fisher scores of 5q- CD34$^+$CD38$^+$Thy-1$^+$ compared to normal CD34$^+$CD38$^+$Thy-1$^+$ cells were comparable to the random expectation, represented by permutation test results (Methods), demonstrating a very close identity between normal and 5q- CD34$^+$CD38$^+$Thy-1$^+$ cells (Figure 2a-b).

The false discovery rate was high when comparing 5q- CD34$^+$CD38$^+$Thy-1$^+$ cells with normal CD34$^+$CD38$^+$Thy-1$^+$ cells, but essentially zero when compared to normal CD34$^+$CD38$^+$Thy-1$^-$ progenitors (Figure 1c), further reflecting the high similarity between 5q- and normal CD34$^+$CD38 Thy-1$^+$ cells.
Differentially expressed genes in 5q deleted CD34+CD38-Thy1+ stem cells

To confirm the expected down-regulation of genes located at 5q31q32 we investigated 37 such genes found in our lists, of which 34 were down-regulated in CD34+CD38-Thy-1+ cells in at least 3 of the 4 investigated patients (Supplementary Table 2).

Based on the array data, a total of ten potentially interesting genes were investigated in further detail using Q-PCR analysis, in part to confirm and quantify better some of the array findings on CD34+CD38-Thy-1+ cells (patients #1 and #2), and in part to extend the analysis to additional patients. Table 3 shows array data on selected genes that were up- or down-regulated with high Fisher scores in 5q-CD34+CD38-Thy-1+ cells (complete list of genes in Supplementary Table 3).

Importantly, the Q-PCR analyses of patients #1 and #2 confirmed all array findings. In all patients investigated by Q-PCR, all three examined genes situated on the involved region of 5q (TAF7 at 5q31, CDC42SE2 at 5q23, and CTNNA1 at 5q31) were, as expected, down-regulated in CD34+CD38-Thy-1+ cells (Figure 3). Thus, TAF723, CDC42SE224, and CTNNA1 (alpha catenin; implicated in the Wnt signaling pathway as a tumor suppressor gene)25, were all down-regulated by approximately 50% (46%,46%, and 55% respectively) (Figure 3). However, the functionally related CTNNB1 (beta catenin)26, implicated in the regulation of normal and leukemic HSC self-renewal27 (not located at 5q), was not significantly affected in investigated patients (Figure 3).

HLF, a transcription factor involved in t(17;19)/TCF3-HLF (previously E2A-HLF)-positive acute lymphoblastic leukemia (ALL)28, reported to be specifically expressed in normal human HSCs6, was down-regulated in all 7 patients investigated (5 by Q-PCR and 2 additional by array) by a mean of 3.2-fold. Interferon-induced transmembrane protein-1 (IFITM1), suggested to play a role in the antiproliferative
activity of interferons\textsuperscript{29}, was up-regulated in all 7 investigated patients, in the five investigated by Q-PCR by a mean of 5.3-fold, whereas Amphiregulin (AREG), an apoptosis inhibitor\textsuperscript{30}, was down-regulated in all 7 patients, by as much as 10-fold or more (Figure 3).

Delta–like homolog (\textit{DLK1}), involved in Notch signaling\textsuperscript{31}, was by Q-PCR analysis found to be up-regulated 1.5–12.6 fold in 6 of 9 patients investigated by Q-PCR (Figure 3).

Of particular interest, \textit{BMI1}, critically involved in regulation of HSC self-renewal\textsuperscript{32,33} was up-regulated in all but one of nine patients analyzed by Q-PCR, by a mean of 2.5-fold (Figure 3).

CCAAT enhancer binding protein-alpha (\textit{CEBPA}), essential for normal myeloid development and implicated in the transformation of AML\textsuperscript{34,35}, was in the array analysis up-regulated in all 4 patients (#1, #2, #3, and #5). This was confirmed by Q-PCR for two patients (#1 and #2) and extended to an additional 2 patients (#4 and #6). However, in four other patients (#8, #9, #10, and #11), the expression was not significantly different from normal CD34\textsuperscript{+}CD38\textsuperscript{-}Thy-1\textsuperscript{+} cells and in one patient (#7) rather down-regulated.

Enhanced resolution of dysregulated gene expression in purified CD34\textsuperscript{+}CD38\textsuperscript{-} Thy-1\textsuperscript{+} stem and CD34\textsuperscript{+}CD38\textsuperscript{-}Thy-1\textsuperscript{-} progenitor cells with del (5q)

To evaluate the potential benefit of global gene profiling of purified and distinct MDS stem and progenitor cell populations, rather than mixed populations\textsuperscript{36–39}, and thereby also identification of stage-specific changes in gene expression, we next compared differentially expressed genes between purified 5q- and normal CD34\textsuperscript{+}CD38\textsuperscript{-}Thy-1\textsuperscript{+} stem cells and between 5q- and normal CD34\textsuperscript{+}CD38\textsuperscript{-}Thy-1\textsuperscript{-}
progenitors (Figure 4). Whereas 5q- and normal CD34+CD38-Thy-1 progenitors showed a relatively high number of differentially expressed genes, much fewer and smaller differences were observed when comparing the profiles of normal and 5q-CD34+CD38-Thy-1+ stem cells (Figure 4a-b).

Strikingly, expression of CEBPA, which was either up-regulated or normal in CD34+CD38-Thy-1+ cells in most 5q- patients (Figure 3 and 4c-d), was rather consistently and dramatically down-regulated in 5q- CD34+CD38-Thy-1 progenitors (Figure 4c-d), as determined by microarray analysis. CTNNA1, BMI1 and DLK1 were also differentially affected in the 5q- CD34+CD38-Thy-1+ stem and CD34+CD38-Thy-1 progenitor populations (Figure 4c), whereas other investigated genes (TAF7, CDC42SE2, HLF, IFITM1 and AREG), showed a similar pattern of expression between CD34+CD38-Thy-1+ and CD34+CD38-Thy-1 cells (Figure 4c).

In contrast to recent studies we did not find CTNNA1 to be downregulated more than the expected 50%, perhaps reflecting that the reported hypermethylation of CTNNA1 in patients with 5q deletions, might be a late event not typically observed in patients with 5q-syndrome.

Nine patients were investigated by Q-PCR to confirm and extend the contrasting expression of CEBPA in 5q- CD34+CD38-Thy-1+ stem and CD34+CD38-Thy-1 progenitor cells. Noteworthy, when compared to their normal counterparts, CEPBA was dramatically down-regulated in all nine investigated patients (range 2.2–1236 fold) in 5q- CD34+CD38-Thy-1 progenitors, but not in CD34+CD38-Thy-1+ stem cells (Figure 4d). Although CEBPA was in one case (#7) down-regulated (4.3-fold) in 5q- CD34+CD38-Thy-1+ cells, its expression was reduced even more in 5q- CD34+CD38-Thy-1 progenitors (as much as 1236-fold) when compared to normal CD34+CD38-Thy-1 progenitors (Figure 4d).
Discussion

The present studies exemplify the importance of performing global gene expression profiling of purified cancer stem and progenitor cell populations. Not only did the molecular fingerprinting of 5q- CD34+CD38-Thy-1+ cells provide stronger support for 5q- syndrome potentially originating from normal HSCs, it also resulted in a much higher resolution of specific gene expression changes when comparing MDS and normal stem and progenitor cells, rather than more heterogeneous populations36-39.

Although it has been postulated that LSCs and other CSCs frequently may originate in the corresponding rare normal multipotent stem cell populations19,41, conclusive evidence for such a model has only been obtained for myeloproliferative disorders3,4. In other leukemias the evidence has at best been circumstantial, typically limited to the leukemic and normal stem cells sharing a CD34+CD38- cell surface phenotype42. However, the expression of CD34 and CD38 antigens, although instrumental for identification of normal HSCs in steady state hematopoiesis13, has in mice as well as in humans been demonstrated to fluctuate considerably43,44.

In the current studies, we adopted a novel approach applying genomics to better determine the identity of the 5q- stem cell population. In all investigated patients, >92% of CD34+CD38-Thy1+ cells were shown to harbor the del(5q), demonstrating that virtually all the cells in the CD34+CD38-Thy1+ compartment of these patients are part of the MDS clone. The 5q- CD34+CD38-Thy1+ compartment represented 0.14% of all BM cells, being only slightly expanded when compared to the CD34+CD38-Thy-1+ HSC compartment in healthy subjects, demonstrating that the normal CD34+CD38-Thy1+ HSC compartment had almost completely been replaced with 5q- stem cells, a conclusion further substantiated by the lack of detectable normal LTC-IC activity in all investigated patients. Thus, 5q- CD34+CD38-Thy1+
cells appear to largely outcompete and hence deplete their normal counterparts, a finding also supported by up to 99% of granulocytes in these patients being 5q deleted\textsuperscript{10}. However, the mechanisms or niches that are in place to limit the size of the normal HSC pool\textsuperscript{45,46} appear also to be able to restrict the size of the 5q- HSC compartment at this stage of the disease (up to 8 years after diagnosis).

In two of the four investigated 5q- patients we observed 5q- LTC-CFC activity from CD34\textsuperscript{+}CD38\textsuperscript{-}Thy1\textsuperscript{+} but not CD34\textsuperscript{+}CD38\textsuperscript{+}Thy1\textsuperscript{-} cells, supporting that the infrequent 5q- CD34\textsuperscript{+}CD38\textsuperscript{-}Thy1\textsuperscript{+} cells are the MDS stem cells in these patients. However, in two other patients in which we also observed a virtually complete replacement of the normal CD34\textsuperscript{+}CD38\textsuperscript{-}Thy1\textsuperscript{+} compartment with 5q- CD34\textsuperscript{+}CD38\textsuperscript{-}Thy1\textsuperscript{+} cells, we failed to detect not only normal but also 5q- LTC-CFC activity. Although we could show also in these patients that normal HSCs have been out-competed by 5q- MDS stem cells, the functional data failed in these cases to conclusively show that 5q- CD34\textsuperscript{+}CD38\textsuperscript{-}Thy1\textsuperscript{+} cells are in fact the MDS stem cells. The most likely interpretation of the lack of 5q- LTC-CFC activity in these two patients is that the readout of this assay requires efficient myeloid differentiation from investigated HSC populations, which is variably affected in MDS.

Strikingly, 5q- and normal CD34\textsuperscript{+}CD38\textsuperscript{-}Thy1\textsuperscript{+} stem cells showed an almost perfect match in global gene expression patterns, as the Fisher scores were low and comparable to the random expectation. In contrast, 5q- CD34\textsuperscript{+}CD38\textsuperscript{-}Thy1\textsuperscript{+} cells displayed extensive differences to normal as well as 5q- CD34\textsuperscript{+}CD38\textsuperscript{-}Thy1\textsuperscript{-} progenitors. Thus, the gene expression pattern of 5q- CD34\textsuperscript{+}CD38\textsuperscript{-}Thy1\textsuperscript{+} cells further supports that the 5q- syndrome might initiate in the normal CD34\textsuperscript{+}CD38\textsuperscript{-}Thy1\textsuperscript{+} HSC compartment, even though an origin in an even more primitive, not yet identified cell population can not be ruled out. Although previous studies suggested the possibility
of a HSC origin of 5q- syndrome, by demonstrating involvement of B cells in a few
patients\(^{10,47}\), the present results are compatible with most, if not all, 5q-syndromes
potentially originating in the HSC pool. As LSCs and other CSCs typically will be
biased in their differentiation towards one particular lineage, and inefficient or
incapable of generating other lineages, we predict that global gene expression
profiling will become an important complementary tool for determining the cellular
origin of LSCs as well as other CSCs. Although our studies unequivocally establish a
very close identity between normal and MDS stem cells, it can however not be ruled
out that this could result from reacquisition of a stem cell phenotype of a targeted
normal progenitor cell population, rather than origin in a normal HSC.

The most salient finding in the present study, was the many and pronounced
differences observed in gene expression between 5q- and normal CD34\(^+\)CD38\(^-\)Thy1\(^-\)
progenitors, compared to the few but distinct differences detected between 5q- and
normal CD34\(^+\)CD38\(^-\)Thy1\(^+\) stem cells. Previous global expression profiling studies
on MDS cases have, at most been performed on CD34\(^+\) or CD133\(^+\) cells\(^{36-39}\),
containing both CD34\(^+\)CD38\(^-\)Thy1\(^+\) stem and CD34\(^+\)CD38\(^-\)Thy1\(^-\) progenitor cells.
The present approach allowed us to identify stage specific changes in gene expression.

Multiple previous gene profiling studies of MDS (including 5q-) patients, had
failed to find consistent changes in expression of \textit{CEBPA}\(^{36-39}\), despite of mouse
studies implicating the potential involvement of reduced \textit{CEBPA} expression in
development of MDS\(^{48}\). In contrast, in the present studies we found in all 9
investigated 5q- patients that \textit{CEBPA} was down-regulated in the MDS progenitors, at
a minimum 2.2-fold. Our finding of consistently reduced \textit{CEBPA} expression in 5q-
progenitors, support an involvement of dysregulated \textit{CEBPA} expression in the
inefficient granulopoiesis characteristic of MDS\(^{48}\).
BMI1 was preferentially up-regulated in 5q- CD34+CD38-Thy1+ cells. Identified as an essential regulator of normal HSC self renewal, and involved in leukemogenesis if deregulated32, the up-regulation of BMI1 could potentially explain the competitive advantage of 5q- CD34+CD38-Thy1+ stem cells over normal HSCs.

In conclusion, the present studies highlight the importance of molecular characterization of distinct stages of cancer stem and progenitor cells to identify and characterize stage-specific dysregulated gene expression in cancer.

Acknowledgments

The authors declare that they have no competing financial interests. The expert advice and assistance of Kees-Jan Pronk, Anna Fossum, Zhi Ma, Klas Raaschau-Jensen and Gunilla Gärdebring for enrichment procedures and cell sorting are highly appreciated. We thank Associate Professor Thoas Fioretos and members of the Lund Stem Cell Center, especially Professor Carsten Peterson, for helpful discussions and Ramin Tehranchi for help with figures. We thank all patients and BM volunteers and the staff at the Departments of Hematology for assistance with aspirations, and Amgen and Genentech for generous contribution of cytokines. This work was supported by grants from The Swedish Cancer Society, The Ingabritt and Arne Lundberg Foundation, The Knut and Alice Wallenberg Foundation via the SWEGENE program, The Georg Danielsson Foundation, The Gunnar Nilsson`s Cancer Foundation, The Crafoordska Foundation, The Åke Wiberg Foundation, The Tobias Foundation, Government Public Health (ALF) Grants, Region Skåne, and the Medical Faculty, University of Lund. The Lund Stem Cell Center is supported by a Center of Excellence grant in Life Sciences from the Swedish Foundation for Strategic Research.
(SSF). PE has an Assistant Professor position supported by a SSF Center of Excellence grant.

**Author Contributions:**

Lars Nilsson designed research, performed research, collected data, analyzed and interpreted data, and wrote the manuscript. Patrik Edén performed research, analyzed and interpreted data, performed statistical analysis and drafted the manuscript. Eleonor Olsson performed research, collected data, analyzed and interpreted data. Robert Månsson performed research, collected data, analyzed and interpreted data, and drafted the manuscript. Ingbritt Åstrand-Grundström performed research, collected data, analyzed and interpreted data. Bodil Strömbeck performed research, collected data, analyzed and interpreted data. Kim Theilgaard-Moench designed research, performed research, collected data, analyzed and interpreted data. Kristina Anderson analyzed and interpreted data. Robert Hast provided clinical material, collected data, analyzed and interpreted data. Eva Hellström-Lindberg provided clinical material, collected data, analyzed and interpreted data. Jan Samuelsson provided clinical material, collected data. Gösta Bergh provided clinical material, collected data. Claus Nerlov designed research, analyzed and interpreted data, and contributed to writing of the manuscript. Bertil Johansson designed research, analyzed and interpreted data, and contributed to writing of the manuscript. Mikael Sigvardsson designed research, analyzed and interpreted data, and contributed to writing of the manuscript. Åke Borg designed research, analyzed and interpreted data, and contributed to writing of the manuscript. Sten Eirik W. Jacobsen designed research, analyzed and interpreted data, and wrote the manuscript.
References


renewal of hematopoietic stem cells mediated by the polycomb gene product Bmi-1. Immunity. 2004;21:843-851


42. Passegue E, Jamieson CH, Ailles LE, Weissman IL. Normal and leukemic hematopoiesis: are leukemias a stem cell disorder or a reacquisition of stem cell characteristics? Proc Natl Acad Sci U S A. 2003;100 Suppl 1:11842-11849


Table 1. Clinical, hematologic, and cytogenetic characteristics of the 5q- syndrome patients

<table>
<thead>
<tr>
<th>No</th>
<th>Sex/age</th>
<th>Years since diagnosis</th>
<th>Hb g/l</th>
<th>WBC $10^9$/l</th>
<th>Plt $10^9$/l</th>
<th>Ongoing treatment</th>
<th>CD34⁺CD38⁺ cells</th>
<th>CD34⁺CD38⁻Thy-1⁺ cells</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>% of MNC</td>
<td>%FISH 5q-</td>
</tr>
<tr>
<td>1</td>
<td>F/83</td>
<td>3</td>
<td>96</td>
<td>3.3</td>
<td>374</td>
<td>E</td>
<td>1.50</td>
<td>100 (216)</td>
</tr>
<tr>
<td>2</td>
<td>F/77</td>
<td>8</td>
<td>100</td>
<td>3.8</td>
<td>130</td>
<td>E</td>
<td>4.00</td>
<td>99 (262)</td>
</tr>
<tr>
<td>3</td>
<td>F/54</td>
<td>1</td>
<td>108</td>
<td>6.1</td>
<td>231</td>
<td>E</td>
<td>0.60</td>
<td>100 (224)</td>
</tr>
<tr>
<td>4</td>
<td>F/89</td>
<td>2</td>
<td>85</td>
<td>5.3</td>
<td>236</td>
<td>E</td>
<td>0.45</td>
<td>100 (221)</td>
</tr>
<tr>
<td>5</td>
<td>F/60</td>
<td>1</td>
<td>120</td>
<td>9.2</td>
<td>1234</td>
<td>--</td>
<td>1.60</td>
<td>100 (244)</td>
</tr>
<tr>
<td>6</td>
<td>F/58</td>
<td>4</td>
<td>122</td>
<td>3.3</td>
<td>448</td>
<td>E</td>
<td>3.20</td>
<td>99 (140)</td>
</tr>
<tr>
<td>7</td>
<td>F/64</td>
<td>3</td>
<td>100</td>
<td>4.9</td>
<td>267</td>
<td>--</td>
<td>3.20</td>
<td>98 (259)</td>
</tr>
<tr>
<td>8</td>
<td>F/85</td>
<td>1</td>
<td>102</td>
<td>3.6</td>
<td>309</td>
<td>--</td>
<td>0.50</td>
<td>93 (323)</td>
</tr>
<tr>
<td>9</td>
<td>F/85</td>
<td>7</td>
<td>76</td>
<td>1.3</td>
<td>95</td>
<td>P</td>
<td>0.43</td>
<td>100 (47)</td>
</tr>
<tr>
<td>10</td>
<td>F/82</td>
<td>3</td>
<td>103</td>
<td>6.4</td>
<td>443</td>
<td>EPO, P</td>
<td>0.34</td>
<td>96 (24)</td>
</tr>
<tr>
<td>11</td>
<td>F/86</td>
<td>8</td>
<td>129</td>
<td>4.7</td>
<td>204</td>
<td>EPO</td>
<td>1.50</td>
<td>nd</td>
</tr>
</tbody>
</table>

MDS mean [SEM]  | 104 [4.7] | 4.7 [0.6] | 361 [93] | 1.57 [0.40] | 98.5 [0.73] | 0.14 [0.04] | 97.4 [0.79] |
Normal mean† [SEM] | 135 | 7.0 | 270 | 1.80 [0.20] | 6.2* | 0.04 [0.01] | 6.2* |

BM morphology and blood values were investigated in connection with BM aspiration for the present studies. All patients were females and had a verified diagnosis of 5q- syndrome according to the WHO classification, with a typical BM morphology including hypolobular megakaryocytes and a low BM blast count (<5%; RA according to FAB classification), an isolated 5q deletion encompassing q13-q15 to q31-q35, normal to elevated platelet counts and normal to slightly reduced WBC. The relative frequencies of CD34⁺CD38⁺ and CD34⁺CD38⁻Thy-1⁺ cells were calculated from the number of BM MNC and the FACS profiles. Number of nuclei evaluated by FISH is given in parenthesis. If sufficient cells, at least 100-200 nuclei...
were always counted. A Cut-off for FISH del(5q); median + 2SD. RA indicates refractory anemia; FAB, French-American-British; WHO, World Health Organization; BM, bone marrow; Hb, hemoglobin; WBC, white blood count; Plt, platelets; E, erythrocyte transfusions; EPO, erythropoietin; P, prednisolon.
† (n =10)
**Table 2. Normal and 5q- clonal LTC-IC activity of 5q- CD34+CD38−Thy-1− and CD34+CD38−Thy-1+ cells**

<table>
<thead>
<tr>
<th>Patient No</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CD34+CD38−Thy-1− cells</td>
<td></td>
<td>CD34+CD38−Thy-1+ cells</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Normal 5q deleted</td>
<td>Normal 5q deleted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Normal mean (n = 5)</td>
<td>0</td>
<td>0</td>
<td>1838</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Six-week LTC-CFC activity of CD34+CD38−Thy-1− (75-750 cells/well) and CD34+CD38−Thy-1+ (750 cells/well) cells from healthy controls and 5q- syndrome patients were evaluated as described in Experimental procedures. After six weeks of co-culture, the number of CFCs produced after an additional two weeks in methylcellulose were counted. Results are the means of 3-6 replicate wells from each patient and cell population. No LTC-CFCs were derived from normal or 5q- CD34+CD38−Thy-1− candidate progenitor cells.
## Table 3. Differentially expressed genes in 5q- versus normal CD34+CD38-Thy-1+ cells

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Gene symbol</th>
<th>Accession Number</th>
<th>Known function</th>
<th>Up/down</th>
<th>Fisher score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interferon induced transmembrane protein 1</td>
<td>IFITM1*</td>
<td>NM_003641</td>
<td>Inhibition of cell proliferation</td>
<td>Up</td>
<td>2.7</td>
</tr>
<tr>
<td>Serine/threonine kinase 19</td>
<td>STK19</td>
<td>NM_032454</td>
<td>Transcriptional regulation</td>
<td>Up</td>
<td>2.5</td>
</tr>
<tr>
<td>Kelch-like ECT2 interacting protein</td>
<td>KLEIP</td>
<td>NM_014458</td>
<td>Cytoskeleton organization</td>
<td>Up</td>
<td>2.0</td>
</tr>
<tr>
<td>Kinesin family member 2C</td>
<td>KIF2C</td>
<td>NM_006845</td>
<td>Chromosome movements during cell division</td>
<td>Up</td>
<td>2.0</td>
</tr>
<tr>
<td>Growth arrest-specific 2</td>
<td>GAS2</td>
<td>NM_005256</td>
<td>p53-dependent apoptosis</td>
<td>Up</td>
<td>2.0</td>
</tr>
<tr>
<td>Cytochrome b-245, beta polypeptide</td>
<td>CYBB</td>
<td>NM_000397</td>
<td>Microbicidal phagocytic oxidative system</td>
<td>Up</td>
<td>1.9</td>
</tr>
<tr>
<td>Solute carrier family 9</td>
<td>SLC9A3R1</td>
<td>NM_004252</td>
<td>Wnt signaling</td>
<td>Up</td>
<td>1.9</td>
</tr>
<tr>
<td>Gp25L2 protein</td>
<td>HSGP25L2G</td>
<td>BC001123</td>
<td>Located at 5q35.3; Unknown function</td>
<td>Up</td>
<td>1.9</td>
</tr>
<tr>
<td>TAF5 RNA polymerase II</td>
<td>TAF5*</td>
<td>NM_006951</td>
<td>Transcriptional regulation</td>
<td>Up</td>
<td>1.8</td>
</tr>
<tr>
<td>Thiosulfate sulfurtransferase (rhodanese)</td>
<td>TST</td>
<td>NM_003312</td>
<td>Mitochondrial enzyme</td>
<td>Up</td>
<td>1.8</td>
</tr>
<tr>
<td>Topoisomerase (DNA) II binding protein</td>
<td>TOPBP1</td>
<td>NM_007027</td>
<td>DNA damage response protein</td>
<td>Up</td>
<td>1.7</td>
</tr>
<tr>
<td>Golgin-67</td>
<td>GOLGIN-67</td>
<td>AF204231</td>
<td>Golgi apparatus protein</td>
<td>Up</td>
<td>1.6</td>
</tr>
<tr>
<td>CHK2 checkpoint homolog (S.pombe)</td>
<td>CHEK2</td>
<td>NM_007194</td>
<td>Putative tumor suppressor</td>
<td>Up</td>
<td>1.6</td>
</tr>
<tr>
<td>Enah/Vasp-like</td>
<td>EVL</td>
<td>NM_016337</td>
<td>Cytoskeleton rearrangement and signal transduction</td>
<td>Up</td>
<td>1.6</td>
</tr>
<tr>
<td>Cell division cycle associated 7</td>
<td>CDCA7</td>
<td>NM_031942</td>
<td>C-myc target gene; lymphoblastoid cell transformation</td>
<td>Up</td>
<td>1.6</td>
</tr>
<tr>
<td>CCAAT/enhancer binding protein (C/EBP), alpha</td>
<td>CEBPA*</td>
<td>NM_004364</td>
<td>TF; neutrophil differentiation; mutated in AML</td>
<td>Up</td>
<td>1.5</td>
</tr>
<tr>
<td>Cell division cycle 25 B</td>
<td>CDC25B</td>
<td>NM_021874</td>
<td>Cell cycle regulator; Oncogenic properties</td>
<td>Up</td>
<td>1.5</td>
</tr>
<tr>
<td>Polymerase (D directed), epsilon 2 (p59 subunit)</td>
<td>POLE2</td>
<td>NM_002962</td>
<td>DNA replication</td>
<td>Up</td>
<td>1.5</td>
</tr>
<tr>
<td>CD99 antigen</td>
<td>CD99</td>
<td>NM_002414</td>
<td>T cell apoptosis; involved in gastric adenocarcinomas</td>
<td>Up</td>
<td>1.4</td>
</tr>
<tr>
<td>Bloom syndrome</td>
<td>BLM</td>
<td>NM_000057</td>
<td>Suppressor of inappropriate recombination.</td>
<td>Up</td>
<td>1.4</td>
</tr>
<tr>
<td>Gene Name</td>
<td>Accession</td>
<td>Function</td>
<td>Fold Change</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defective in sister chromatid cohesion homolog 1</td>
<td>DCC1</td>
<td>Faithful inheritance of chromosomes to daughter cells</td>
<td>Up 1.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D (cytosine-5-)-methyltransferase 3, alpha</td>
<td>DNMT3</td>
<td>De novo methylation</td>
<td>Up 1.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inhibitor of DNA binding 1</td>
<td>ID1</td>
<td>Malignant progression</td>
<td>Up 1.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delta-like homolog</td>
<td>DLK1*</td>
<td>Notch-related differentiation-associated</td>
<td>Up 1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor protein p53 binding protein 1</td>
<td>TP53BP1</td>
<td>Tumor suppressor gene</td>
<td>Up 1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V-myc myelocytoma</td>
<td>MYC</td>
<td>Oncogene</td>
<td>Up 1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B lymphoma Mo-MLV insertion region</td>
<td>BMI1*</td>
<td>Implicated in HSC self-renewal</td>
<td>Up 0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solute carrier family 25A, member 4</td>
<td>SLC25A4</td>
<td>Apoptosis inhibitor</td>
<td>Down 2.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ribosomal protein S21</td>
<td>RPS21</td>
<td>Component of the small 40S ribosomal subunit</td>
<td>Down 2.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arachidonate 12-lipoxygenase, 12R type</td>
<td>ALOX12B</td>
<td>Iron ion binding</td>
<td>Down 1.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amphiregulin schwannoma-derived growth factor</td>
<td>AREG*</td>
<td>Interacts with EGF/TGF-α; inhibits carcinoma</td>
<td>Down 1.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hepatic leukemia factor</td>
<td>HLF*</td>
<td>TF involved in t(17;19) translocation in ALL</td>
<td>Down 1.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAF7 RNA polymerase II</td>
<td>TAF7*</td>
<td>Located at 5q31; Transcriptional regulation</td>
<td>Down 1.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pleckstrin homolog domain</td>
<td>PLEKHA6</td>
<td>Signal transduction</td>
<td>Down 1.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WAS protein family</td>
<td>WASF3</td>
<td>Member of the Wiskott-Aldrich syndrome protein family</td>
<td>Down 1.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myotrophin</td>
<td>MTPN</td>
<td>Regulation of NFκB</td>
<td>Down 1.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycoprotein Ib (platelet), beta-polypeptide</td>
<td>GP1BB</td>
<td>Associated with Bernard-Soulier or giant platelet disorder</td>
<td>Down 1.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target of myb1-like 1 (chicken)</td>
<td>TOM1L1</td>
<td>Signal transduction</td>
<td>Down 1.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-box and leucin-rich repeat protein 5</td>
<td>FBXL5</td>
<td>Proteasome degradation</td>
<td>Down 1.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activated leukocyte cell adhesion molecule</td>
<td>ALCAM*</td>
<td>Associated with tumor progression</td>
<td>Down 1.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zinc finger protein 366</td>
<td>ZNF366</td>
<td>Located at 5q13.2; Transcriptional regulation</td>
<td>Down 1.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forkhead box O1A</td>
<td>FOXO1A</td>
<td>TF; involved in malignant transformation</td>
<td>Down 1.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MADS box transcription enhancer factor 2</td>
<td>MEF2C</td>
<td>Located at 5q14; Transcriptional regulation</td>
<td>Down 1.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**CREBB/EP300 inhibitor 1**

CREBB/EP300 implicated as tumor suppressor gene

**Endoplasmic reticulum chaperone (SIL1)**

SIL1, NM_022464

Located at 5q31; Enhances nucleotide exchange

**CDC42 small effector 2**

CDC42SE2*, AK025620

Located at 5q23.3; Rho GTPase; regulates actin and kinase signaling

**Tyrosine kinase with Ig/EGF homology domains**

TIE, NM_005424

High expression correlates with shorter survival in CML

**Transcription factor 8**

TCF8, NM_030751

Involved in TGF and BMP signaling

**Transforming growth factor beta 1**

TGFB1, NM_000660

Multifunctional; proliferation/differentiation, transformation

**Stem cell growth factor**

SCGF, NM_002975

Growth factor for primitive hematopoietic stem cells

**Suppression of tumorigenicity 13**

ST13, U17714

Candidate tumor suppressor gene

**R binding motif protein 27**

RBM27, AB037732

Located at 5q32; Unknown function

**Tumor necrosis factor, alpha-induced protein 8**

TNFAIP8, NM_014350

Located at 5q23.1; Oncogenic factor in cancer cells

**Bone morphogenetic protein 6**

BMP6, NM_001718

Secreted signaling protein involved in bone formation

---

Selected list of genes up- or down-regulated in 5q- (n = 4) CD34⁺CD38⁻Thy-1⁺ candidate HSCs when compared to normal (n = 5) CD34⁺CD38⁻Thy-1⁺ cells (for complete list see Supplementary Table 3). *Microarray-based expression validated by Q-PCR. TF indicates transcription factor.
Figure legends

Figure 1. Thy-1 expression within CD34^+CD38^- progenitor and CD34^+CD38^- stem cell compartments in normal and 5q- syndrome BM.

CD34 enriched normal (a) and 5q- (b, patient #6) BM cells were stained with MAbs against CD34, CD38 and Thy-1 or irrelevant isotype control Abs (Methods). Left panels show expression profiles for already CD34 enriched cells with the size of the CD34^+CD38^- and CD34^+CD38^- populations given as percentages of total MNC. Shown are also the gates for CD34^+CD38^- and CD34^+CD38^- cells used for sorting and for further investigation of Thy-1 expression (middle panels). Dotted lines represent the negative isotype control and solid lines the specific Thy-1 expression. Note that Thy-1 expression is much higher in CD34^+CD38^- than in CD34^+CD38^- cells for both healthy and 5q- subjects. Sorted CD34^+CD38^-Thy-1^+ candidate HSCs and CD34^+CD38^-Thy-1^- candidate progenitors were analyzed by FISH for the 5q deletion. Normal CD34^+CD38^-Thy-1^+ cells (a) show two green and two red signals while CD34^+CD38^-Thy-1^- cells with 5q deletion (b) show two green and one red signal (right panels). Bars in the FISH pictures represents 10 µm.

Figure 2. Comparison of the global gene expression profiles of normal and 5q- CD34^+CD38^-Thy-1^+ and CD34^+CD38^-Thy-1^- cell populations.

(a) Number of genes versus minimal Fisher score for 5q- CD34^+CD38^-Thy-1^+ candidate HSCs when compared to normal CD34^+CD38^-Thy-1^+ HSCs (solid line), normal CD34^+CD38^-Thy-1^- progenitors (dashed line), and 5q- CD34^+CD38^-Thy-1^- progenitors (dotted line).

(b) Number of differentially expressed genes with Fisher score above 2 for 5q- CD34^+CD38^-Thy-1^+ candidate HSCs when compared to normal CD34^+CD38^-Thy-1^+ HSCs, normal CD34^+CD38^-Thy-1^- progenitors, and 5q- CD34^+CD38^-Thy-1^- progenitors, respectively. Error bars indicate 95% confidence interval [Poisson statistics].

(c) False discovery rate as a function of number of top ranked genes using the Fisher score. Results are shown for 5q- CD34^+CD38^-Thy-1^+ candidate HSCs when compared to normal CD34^+CD38^-Thy-1^+ HSCs (solid line), normal CD34^+CD38^-Thy-1^- progenitors (dashed line), and 5q- CD34^+CD38^-Thy-1^- progenitors (dotted line). The false discovery rate, i.e. number of accepted genes in a permutation test divided by the same number in the correct list, was high when comparing 5q- HSCs with normal HSCs.
but essentially zero when compared to normal progenitors. A high false discovery rate implies that the compared populations do not differ in much more genes than expected by random fluctuations, while a low false discovery rate implies significant differences in gene expression between the compared cell types, reflecting more extensive differences than expected by random fluctuation.

**Figure 3. Quantitative-PCR expression analysis of selected genes in 5q- CD34⁺CD38⁻Thy-1⁺ HSCs as compared to normal CD34⁺CD38⁻Thy-1⁺ HSCs.**

RNA was isolated from highly purified normal and 5q- CD34⁺CD38⁻Thy-1⁺ cells (Methods) and analyzed for quantitative expression of ten genes, selected based on the array analysis. The mean expression of 5 healthy subjects (normal; white bars) compared to the mean (grey bars) and individual (Patients #1, #2, #4, #6, #7, #8, #9, #10 and #11; black bars) expression of each gene for the 5q-syndrome patients normalized against HPRT expression levels.

**Figure 4. Comparison of differentially expressed gene expression in MDS stem cells and progenitors**

(a) Number of genes versus minimal Fisher scores for 5q- versus normal CD34⁺CD38⁻Thy-1⁺ progenitors (dashed line) and 5q- versus normal CD34⁺CD38⁻Thy-1⁺ HSCs (solid line), respectively. (b) Number of differentially expressed genes with Fisher score above two when comparing 5q- and normal CD34⁺CD38⁻Thy-1⁺ stem cells and when comparing 5q- and normal CD34⁺CD38⁻Thy-1⁺ progenitors. Error bars indicate 95% confidence interval [Poisson statistics]. (c) Microarray-based gene expression of ten selected genes in 5q- CD34⁺CD38⁻Thy-1⁺ stem cells and 5q- CD34⁺CD38⁻Thy-1⁺ progenitors relative to the mean expression in their normal counterparts (from five healthy subjects and four 5q- patients). [Error bars show SEM]. (d) Q-PCR analysis of CEBPA expression in 5q- CD34⁺CD38⁻Thy-1⁺ stem cells and 5q-CD34⁺CD38⁻Thy-1⁺ progenitors. Shown are the mean [SEM] differential expression for five healthy subjects and nine 5q- patients and individual expression levels for the investigated 5q- patients (#1, #2, #4, #6, #7, #8, #9, #10, and #11) normalized against HPRT expression levels. Grey staples for HSCs and white staples for progenitors throughout the figure.
Fig. 1

(a) NORMAL

(b) 5q- SYNDROME

CD38

CD34

Thy-1

CD38

CD34

5q- SYNDROME

Thy-1

For personal use only.
Fig. 2

(a) Graph showing the relationship between minimal Fisher score and number of genes.

(b) Bar graph comparing the number of genes (Fisher score >2) between Normal CD34+CD38-Thy-1+ versus Normal CD34+CD38-Thy-1- versus 5q- CD34+CD38-Thy-1-.

(c) Graph showing the false-discovery rate versus number of top-ranked genes.

Legend:
- Solid line: 5q- CD34+CD38-Thy-1+ versus normal CD34+CD38-Thy-1+
- Dashed line: 5q- CD34+CD38-Thy-1+ versus normal CD34+CD38-Thy-1-
- Dotted line: 5q- CD34+CD38-Thy-1+ versus 5q- CD34+CD38-Thy-1-
Fig. 3

Normalized Expression

TAF7

CDC42SE2

CTNNA1

CTNNB1

HLF

IFITM1

AREG

DLK1

BMI1

CEBPA

For personal use only. From www.bloodjournal.org
Fig. 4

a) The graph shows the relationship between the number of genes and the minimal Fisher score.

b) The histogram illustrates the distribution of genes with a Fisher score greater than 2, comparing CD34+CD38+Thy-1+ and CD34+CD38-Thy-1-.

c) The bar chart displays the relative gene expression levels of various genes, including CEBPA, CTNNA1, CTNNB1, BM1, DLK1, HLF, IFITM1, CDC42SE2, TAF7, and AREG.

d) The chart depicts the normalized expression levels of CEBPA across different conditions, with CD34+CD38+Thy-1+ and CD34+CD38-Thy-1- categories.