

University of Groningen

The biology of human hematopoietic stem and progenitor cells in acute myeloid leukemia, aging and autologous transplantation

Woolthuis, Carolien Marthe

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2013

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Woolthuis, C. M. (2013). *The biology of human hematopoietic stem and progenitor cells in acute myeloid leukemia, aging and autologous transplantation*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

CHAPTER 2

Gene expression profiling in the leukemic stem cell-enriched CD34⁺ fraction identifies target genes that predict prognosis in normal karyotype AML

Hendrik J.M. de Jonge^{1*}, Carolien M. Woolthuis^{2*}, Annet Z. Vos², Andre Mulder³, Eva van den Berg⁴, Philip M. Kluin⁵, Karen van der Weide², Eveline S.J.M. de Bont¹, Gerwin Huls², Edo Vellenga^{2#}, Jan Jacob Schuringa^{2#}

¹ Department of Pediatrics, Division of Pediatric Oncology/Hematology,

² Department of Hematology, ³ Department of Clinical Chemistry, ⁴ Department of Genetics,

⁵ Department of Pathology and Medical Biology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands.

* shared first authorship

shared senior authorship

Leukemia 2011

Abstract

In order to identify acute myeloid leukemia (AML) CD34⁺-specific gene expression profiles, mononuclear cells from AML patients (n=46) were sorted into CD34⁺ and CD34⁻ subfractions and genome-wide expression analysis was performed using Illumina BeadChip Arrays. AML CD34⁺ and CD34⁻ gene expression was compared to a large group of normal CD34⁺ bone marrow cells (n=31). Unsupervised hierarchical clustering analysis showed that CD34⁺ AML samples belonged to a distinct cluster compared to normal bone marrow and that in 61% of the cases the AML CD34⁺ transcriptome did not cluster together with the paired CD34⁻ transcriptome. A top 50 of AML CD34⁺-specific genes was selected by comparing the AML CD34⁺ transcriptome with the AML CD34⁻ and CD34⁺ normal bone marrow transcriptomes. Interestingly, for three of these genes (*ANKRD28*, *GNA15* and *UGP2*) a high transcript level was associated with a significant poorer overall survival in two independent cohorts (n=163 and n=218) of normal karyotype AML. Importantly, the prognostic value of the continuous transcript levels of *ANKRD28* (OS HR: 1.32, *P*=0.008), *GNA15* (OS HR: 1.22, *P*=0.033) and *UGP2* (OS HR: 1.86, *P*=0.009) was shown to be independent from the well known risk factors *FLT3*-ITD, *NPM1c*⁺ and *CEBPA* mutation status.

Introduction

Acute myeloid leukemia (AML) is clinically, cytogenetically and molecularly a heterogeneous disease which makes it challenging to classify it properly. Currently, patients diagnosed with AML are stratified into separate risk groups based on morphological, cytogenetic and molecular abnormalities.^{1,2} However, especially in the intermediate risk group that represents the largest AML subgroup (60%), treatment outcome varies considerably. Within the intermediate risk group, subgroups are recognized based on mutations in the nucleophosmin 1 (leading to cytoplasmic dislocalization of NPM1 (NPM1c⁺)) and *fms*-related tyrosine kinase 3 (*FLT3*) genes or based on biallelic mutations in the CEBP alpha gene.³⁻⁵ Identification of novel molecular markers might therefore be helpful for further risk stratification. In recent years, a number of gene expression profiling (GEP) studies has been performed in order to improve the identification of known cytogenetic subgroups and to recognize new clusters of AML patients with distinct gene-expression signatures.⁶⁻¹¹ Most of these AML GEP studies have been performed using the total AML mononuclear cell (MNC) fraction.⁶⁻⁹ Since cell lineage and differentiation stages affect gene expression-based clustering,^{6,7,10} the differentially expressed genes associated with the differentiation stage might obscure more basic gene expression information related to tumor initiation and maintenance. Consequently, profiling of a more homogenous leukemic cell population, instead of the total MNC fraction might enhance the feasibility of GEP in order to identify novel prognostic markers. AML is thought to be initiated and maintained by relatively small numbers of leukemia-initiating cells (LICs) that have an enhanced self-renewal capacity and can engraft in immunodeficient mice.^{12,13} In the vast majority of leukemias LICs have been found to reside in the CD34⁺ compartment.¹²⁻¹⁷ Studies aimed at further enrichment of LICs revealed substantial heterogeneity in cell surface marker expression,¹⁸⁻²³ and it has been observed that LICs can reside both in the CD34⁺/CD38⁻ as well as in the CD34⁺/CD38⁺ fractions.²⁴ In our current study, AML MNC fractions were sorted in CD34⁺ and CD34⁻ subfractions and gene expression was compared to a large group of normal CD34⁺ bone marrow cells. Thus, we were able to identify AML CD34⁺-specific gene expression profiles, which included a number of genes that could significantly predict prognosis in normal karyotype AML independent of already established prognostic factors.

Materials and Methods

Patient material

GEP was performed on blast cells of 46 patients with AML from a single center (University Medical Center Groningen). The diagnosis of AML was based on cytological examination, immunophenotyping and cytogenetic analysis of blood and bone marrow. AML blasts were collected

from peripheral blood (PB) or bone marrow (BM; n=7) after achieving informed consent. For three out of 46 patients cells from relapsing disease were used and one AML was preceded by a myelodysplastic syndrome. The study was approved by the Medical Ethical Committee of the UMCG. MNC were isolated by density gradient centrifugation using lymphoprep (PAA, Cölbe, Germany) and cryopreserved. After thawing, CD34⁺ and CD34⁻ AML cells were selected by MoFlo sorting (Dako Cytomation, Carpinteria, CA, USA) using a CD34 PE-labeled antibody (Clone 8G12, BD Biosciences, San Jose, California, USA). Cytogenetic risk group distinction (favorable, intermediate, and unfavorable) according to current HOVON/SAKK protocols is provided in Table 1.²⁵⁻²⁷ Potential donors for allogeneic bone marrow transplantation and patients who underwent elective total hip replacement served as normal controls (n=31) after providing informed consent. From the normal bone marrow aspirates the MNC fraction was isolated and CD34-enriched fractions were obtained using anti-CD34 magnetic beads (Clone QBEND/10, Miltenyi Biotec, Auburn, CA, USA). We realize that differences in procedures that were used to isolate CD34⁺ cells from NBM and AML samples may have resulted in some changes at the transcriptome level. Nevertheless, we currently do not have indications that different CD34⁺ populations were obtained via these methods since different antibodies were used, or that these potential differences in transcriptomes were main determinants in our clustering analyses.

Analysis of mutations in NPM1 and FLT3

Detection of *NPM1* mutations was performed by immunohistochemical staining on bone marrow biopsies that were fixed in 10% neutral phosphate buffered formalin (3.6% formaldehyde) for at least 12 hours and decalcified in a solution containing 10% (v/v) acetic acid and 10% formalin (v/v; 3.6% formaldehyde) for one or two days. Detection of *NPM1* localization was performed on paraffin embedded 3µm tissue sections by immunohistochemical staining using a Benchmark XT (Ventana Medical Systems S.A., USA). The antigens were retrieved with EDTA buffer (pH 8.5) and endogenous peroxidase was blocked with H₂O₂. Slides were incubated with 1:50 diluted supernatant of the biotinylated anti-*NPM1* antibody (kindly provided by Prof. Falini, Perugia, Italy). Antigens were visualized using the ultraview universal DAB detection kit (Ventana). Exclusive nuclear versus a combined nuclear and cytoplasmic localization of the protein was scored by an experienced hematopathologist. The sensitivity and specificity of the assay have been tested on a series of more than 100 AML also analyzed for mutations in the *NPM1* gene.

Mutational analysis of ITD within the JM domain of the *FLT3* gene was performed with RT-PCR using primers 5'-caatttagtatgaagcc-3' and 5'-caaactctaaatttctct-3', as previously described.²⁸

Gene expression profiling

Total RNA was isolated using the RNeasy mini kit from Qiagen (Venlo, The Netherlands) according to the manufacturer's recommendations. RNA quality was examined using the Agilent 2100 Bioanalyzer

(Agilent Technologies, Waldbronn, Germany). Genome-wide expression analysis was performed on Illumina (Illumina, Inc., San Diego, CA) BeadChip Arrays Sentrix Human-6 (46k probesets). Typically, 0.5-1 µg of mRNA was used in labeling reactions and hybridization with the arrays was performed according to the manufacturer's instructions.

The RMA method in R version 2.4.0 was used to compute probe sets summaries.²⁹ Quantile normalization was applied to log₂-transformed intensities.³⁰ Principal component analysis (PCA) was performed for quality control.³¹⁻³⁴ The MIAME-compliant microarray data are available at <http://www.ncbi.nlm.nih.gov/geo/> under accession number GSExxxx [will be publicly deposited upon publication].

To perform further analyses two publicly available cytogenetically normal AML data sets (GSE16891 [NCBI GEO] ; n = 218)^{6,35} and (GSE12417 [NCBI GEO] ; n = 163)³⁶ were utilized. The selected data sets provided us with clinically-annotated gene expression values from Affymetrix Human Genome U133A or U95 gene chip array (Affymetrix, Santa Clara, CA).

Class comparison

Class comparison was performed using the software package Biometric Research Branch ArrayTools (BRB ArrayTools) version 3.6.0, developed by the Biometric Research Branch of the US National Cancer Institute (<http://linus.nci.nih.gov/BRB-ArrayTools.html>). As the CD34⁺ and CD34⁻ AML samples were paired samples (n=44), a paired t-test was used to identify differentially expressed genes. For the comparison CD34⁺ AML (n=46) versus CD34⁺ NBM (n=31), as well as CD34⁻ AML (n=44) versus CD34⁺ NBM (n=31) a random variance t-test was used. Differential expression was considered significant at $P < 0.0001$. Average linkage hierarchical clustering with the centered correlation distance metric was performed using Cluster 3.0 and TreeView software.³⁷

Statistical analyses

Statistical analyses were performed with SPSS software, release 16.0. Actuarial probabilities of overall survival (OS) (with death due to any cause) as well as event-free survival (EFS, with failure in case of no complete remission [CR1] or relapse or death) were estimated according to the Kaplan-Meier method. Overall group differences were evaluated using a Mann-Whitney U test, Chi-square test or Fisher exact in 2x2 setting. Correlations were calculated with the Spearman rank correlation coefficient (rho). The association between transcript levels of the top 50 CD34⁺ AML specific genes and OS and EFS was tested in univariate Cox models. Multivariate cox regression analysis was applied to determine the association of *ANKRD28*, *GNA15* and *UGP2* and OS/EFS with adjustment for known disease-related risk factors such as *FLT3*-ITD, *NPM1c⁺* and *CEBPA* mutations. Of note, when multiple probe sets representing the same gene were present on the Affymetrix arrays, probes were averaged. All tests were 2 tailed, and a $P < .05$ was considered statistically significant.

Results

Samples were available from 46 AML patients, of which 44 were paired CD34⁺ and CD34⁻ samples. Patient characteristics are provided in Table 1. With regard to the risk group 7% of the patients belonged to the good risk group, 63% to the intermediate risk group and 28% to the poor risk group. Mean percentage of sorted CD34⁺ cells within the AML mononuclear cell fractions was 28.9% (ranging 1-90%).

Table 1 Patient characteristics (n=46)

Age – yrs		
Median	55	
Range	20-80	
White blood cell count (x10 ⁹ /L)		
Median	45.8	
Range	2.2-229.3	
Blast percentage (%)		
Median	64%	
Range	12-98%	
French-American-British classification – no. (%)		
M0	3	7%
M1	9	20%
M2	12	26%
M4	3	7%
M5	17	37%
M6	1	2%
M7	1	2%
Cytogenetic characteristics* – no. (%)		
t(8;21)	4	9%
inv(16)	2	4%
t(v;11)(v;q23)	4	9%
del(5q)	4	9%
normal karyotype	24	52%
complex cytogenetic abn. (≥3)	10	22%
other	5	11%
Risk group according to HOVON/SAKK – no.† (%)		
good	3	7%
intermediate	29	63%
poor	13	28%
unknown	1	2%
FLT3/NPM1 – no. (%)		
FLT3 wt/NPM1 wt	26	57%
FLT3-ITD/NPM1 wt	10	22%
FLT3 wt/NPMc ⁺	1	2%
FLT3-ITD/NPMc ⁺	9	20%

* Patients may be counted more than once owing to the coexistence of more than one cytogenetic abnormality. † Risk group distinction was made based on current HOVON/SAKK protocols with favorable risk: t(15;17), t(8;21) and WBC ≤ 20 x 10⁹/l at diagnosis, or inv(16)/t(16;16) and no unfavorable cytogenetic abn.; unfavorable risk: inv(3) or t(3;3), t(6;9), t(v;11)(v;q23) other than t(9;11), -5 or del(5q), -7, abn(17p), complex karyotype (three or more abnormalities in the absence of a WHO designated recurring chromosome abnormality); intermediate risk: all chromosome abnormalities not classified as favorable or unfavorable. Abbreviations: wt = wild type.

Transcriptome differences among AML CD34⁺ and CD34⁻ cell populations versus normal bone marrow CD34⁺ cells

Unsupervised hierarchical clustering analysis using all ~48,000 probe sets of CD34⁺ AML, CD34⁻ AML as well as CD34⁺ normal bone marrow samples showed that CD34⁺ normal bone marrow samples belonged to a distinct cluster, indicating that transcriptome differences between AML and normal bone marrow samples are relatively large, irrespective of the AML CD34 population status (Fig 1A). Within the AML group not all paired CD34⁺ and CD34⁻ samples clustered together. In 61% of the cases (27/44 AMLs) the CD34⁺ transcriptome did not cluster together with the paired CD34⁻ transcriptome, while in 39% of the cases (17/44 AMLs) the CD34⁺ and CD34⁻ transcriptomes did cluster together (Fig 1A). Thus, these data indicate that in the majority of AML cases the leukemic stem cell-enriched CD34⁺ gene expression profile is quite distinct from the leukemic CD34⁻ compartment. Between these two groups, no differences were observed regarding CD34% ($P = .21$, Mann-Whitney U), frequency of NPMc⁺ ($P = .28$, Fisher Exact) and *FLT3*-ITD ($P = .99$, Fisher Exact), age at diagnosis ($P = .38$, Mann-Whitney U) and cytogenetic risk group ($P = .40$, Chi-square). The difference in clustering was also not related to a difference in contamination of lymphoid cells. FACS analysis of the AML MNC fraction revealed no significant differences in the mean percentages of CD3⁺ (3.9 vs. 4.3%, $P = .54$) and CD19⁺ (5.6 vs. 4.6%, $P = .81$, Mann-Whitney U) cells between AML MNC fractions of which the CD34⁺ and CD34⁻ cell populations did or did not cluster together. We also analyzed the correlation between clustering of the paired CD34⁺ and CD34⁻ AML transcriptomes and the tissue source (either PB or BM). No significant correlation between these two ($P = 0.80$) was found. To further investigate a possible effect of tissue source on the clustering of the samples, we performed micro-array analyses of paired PB and BM samples from five additional AML patient samples. For this analysis the PB and BM AML samples were derived from the same patients. An unsupervised clustering analysis of these samples shows that the paired samples from one patient cluster together (Supplemental Figure 1), indicating that gene expression patterns are more different between AML samples than between PB and BM within one patient.

A significant difference in blast percentages of the AML samples of which the paired CD34⁺ and CD34⁻ transcriptomes clustered together (median percentage of blasts 80% (12-98%)) versus those of which the paired CD34⁺ and CD34⁻ transcriptomes did not cluster together (median percentage of blasts 54% (13-96%)) was found ($P = 0.045$). However importantly, when the AML CD34⁺ samples were divided in two groups separated based on the median blast percentage, no statistically significant differences in gene expression were found between samples with a high pre-sorting blast percentage and those with a low pre-sorting blast percentage.

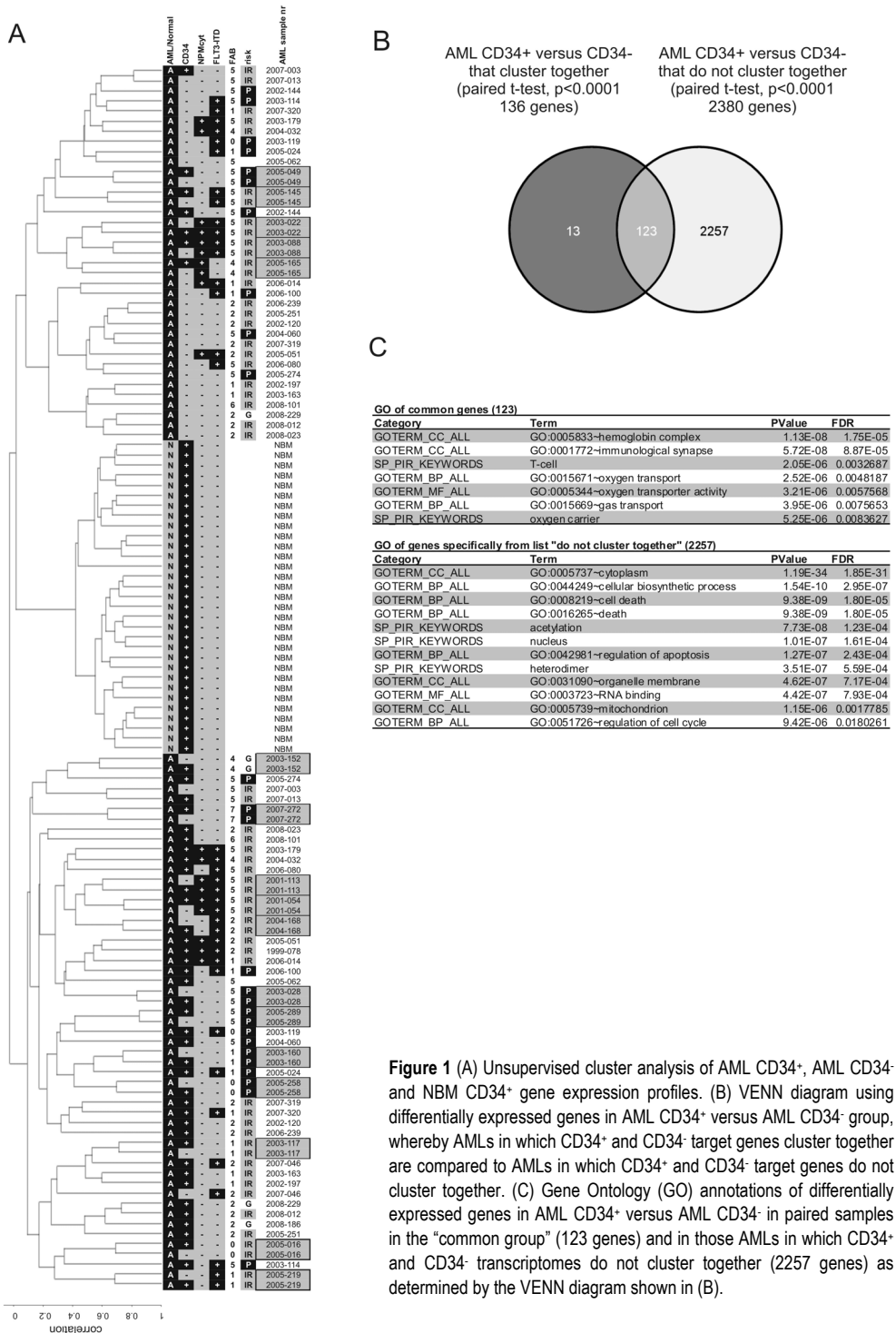


Figure 1 (A) Unsupervised cluster analysis of AML CD34+, AML CD34- and NBM CD34+ gene expression profiles. (B) VENN diagram using differentially expressed genes in AML CD34+ versus AML CD34- group, whereby AMLs in which CD34+ and CD34- target genes cluster together are compared to AMLs in which CD34+ and CD34- target genes do not cluster together. (C) Gene Ontology (GO) annotations of differentially expressed genes in AML CD34+ versus AML CD34- in paired samples in the "common group" (123 genes) and in those AMLs in which CD34+ and CD34- transcriptomes do not cluster together (2257 genes) as determined by the VENN diagram shown in (B).

Differentially expressed probe sets were identified of AML samples of which the CD34⁺ and CD34⁻ cell populations clustered together (n=27, resulting in 136 genes) versus AML samples of which the CD34⁺ and CD34⁻ cell populations did not cluster together (n=17, resulting in 2380 genes), and these differentially expressed gene sets were compared in a VENN diagram (Fig 1B and Supplemental Tables 1 and 2). Regardless of whether or not CD34⁺ and CD34⁻ transcriptomes clustered together, GO analysis revealed that common differences in gene expression between CD34⁺ and CD34⁻ groups in all AML cases (123 genes) were particularly enriched for genes that associated with T-cells and erythropoiesis (Figure 1C). As expected, CD34 was the highest differentially expressed gene in this subset (Supplementary Tables 1 and 2). Indeed, this indicates that genes that associate with a more committed phenotype particularly specify differences between CD34⁺ and CD34⁻ compartments, but this is the case in all AML samples and not just in those samples where CD34⁺ and CD34⁻ transcriptomes did not cluster together. Moreover, gene expression differences between AML CD34⁺ and CD34⁻ transcriptomes associated with differentiation programs are found in both samples with a high pre-sorting blast percentage as well as those with a low pre-sorting blast percentage (Supplemental Figure 2). This strongly suggests that the identified differences between CD34⁺ and CD34⁻ AML cells associated with differentiation are not just a reflection of a difference in pre-sorting blast percentages of the AML samples.

Identification of CD34⁺ AML specific genes

Gene expression profiles were also compared between CD34⁺ AML cells versus CD34⁺ normal bone marrow cells. This analysis revealed 3809 differentially expressed unique genes (Figure 2A, 2180 up and 1629 down in CD34⁺ AML, Supplementary Table 3). Furthermore, 3162 unique genes were found to be differentially expressed between CD34⁻ AMLs and CD34⁺ normal bone marrow samples (1717 up and 1445 down in CD34⁻ AML, Supplementary Table 4). 2132 genes were overlapping between both analyses, and thus differentially expressed between AML and normal bone marrow irrespective of CD34 status (hereafter referred to as common AML specific genes). 1677 genes were specifically expressed in the AML CD34⁺ group versus normal CD34⁺, of which 1013 were significantly higher expressed in CD34⁺ AML cells compared to normal bone marrow cells (Figure 2A). Figure 2B shows examples of GO-ontologies representing the differentially expressed genes between AML CD34⁺ versus normal CD34⁺ cells, AML CD34⁻ versus normal CD34⁺ cells, as well as the common AML specific genes. A hierarchical clustering analysis using only the top 50 of AML CD34⁺ specific genes is shown in Figure 2C. Of note, this top 50 list of genes is adjusted by selecting only those genes that were also present on the gene expression arrays of the validation cohorts.

By comparing the AML CD34⁺ specific UP signature (1013 genes) with the list of genes differentially expressed between AML CD34⁺ and CD34⁻ transcriptomes in all AML samples (123 genes) an overlap of 12 genes (*ZMYM3*, *GUCY1A3*, *C15ORF17*, *TARBP1*, *SRBD1*, *LOC149134*, *TFPI*, *SUPT3H*, *MAP2K5*, *PAQR7*, *LOC727935*, *ZBTB8*, *C14ORF159*) was found.

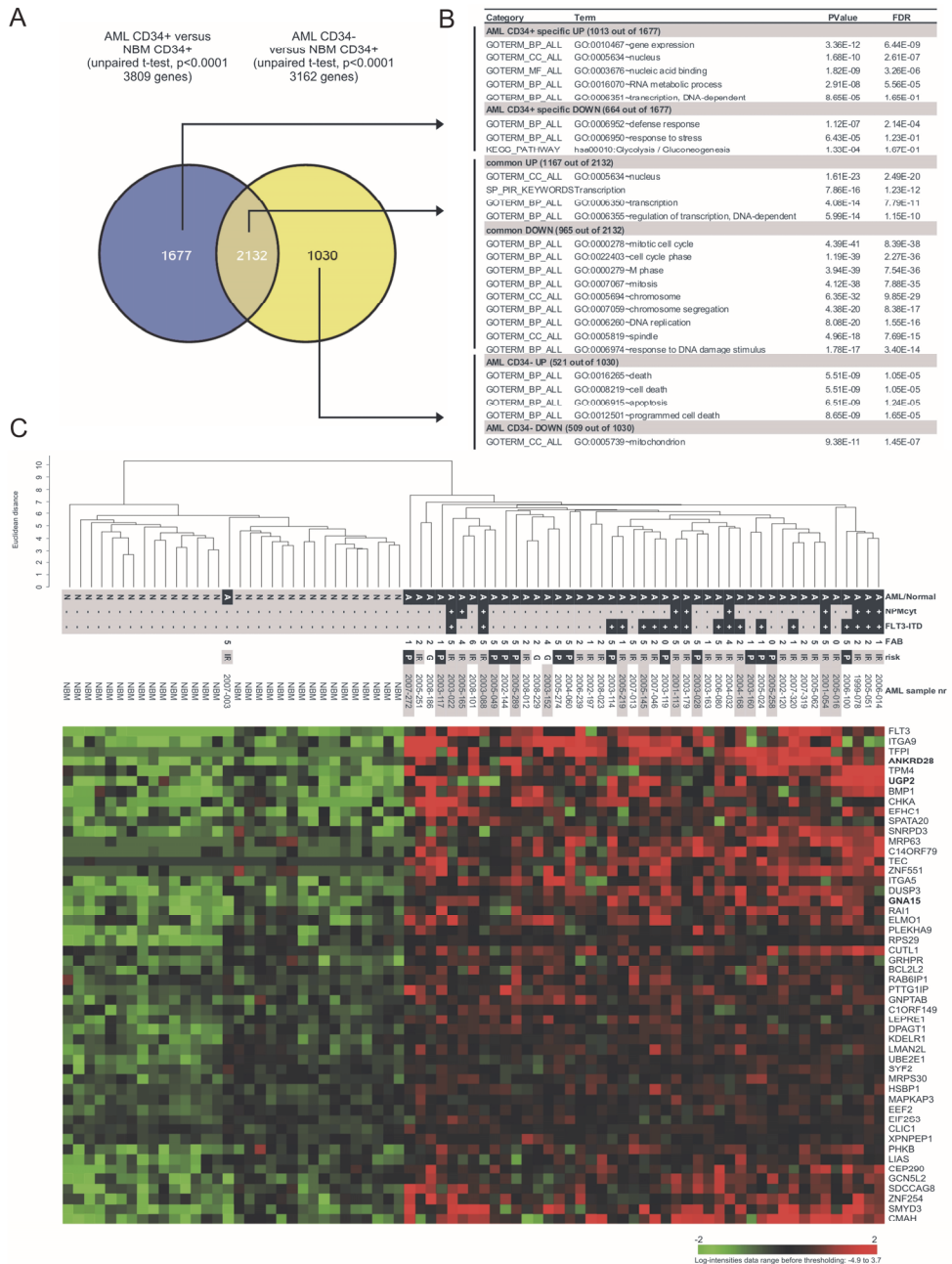


Figure 2 (A) Gene expression profiles were compared between AML CD34+ ($n=46$) versus normal bone marrow CD34+ ($n=31$) as well as between AML CD34- ($n=46$) versus normal bone marrow ($n=31$). A VENN diagram with these two gene sets is shown in order to identify AML CD34+ specific genes (1677), AML CD34- specific genes (1030), and common genes that are differentially expressed between normal BM CD34+ cells versus AML (2132). (B) Gene Ontology (GO) annotations of differentially expressed genes as determined in (A). (C) Supervised cluster analysis (Euclidian, average linkage) using the top 50 AML CD34+ specific genes as shown in Supplemental Table 5.

Prognostic value of top 50 AML CD34⁺ specific genes in normal karyotype AML

Next, we wondered whether the set of 50 CD34⁺ specific AML genes had prognostic significance. Therefore, univariate cox regression analyses were performed between the continuous transcript levels of these 50 CD34⁺ specific genes and OS in a large series of *de novo* normal karyotype AMLs³⁶ (n = 163) (Supplemental Table 5 summarizes the results for all 50 genes). These analyses revealed that 6 genes out of these 50 had predictive significance at a P-value \leq .01 whereby higher expression levels correlated with poor survival rates.

Next, these findings were validated in an independent cohort of 218 normal karyotype AMLs⁶ (Supplemental Table 5). Interestingly, higher transcript levels of three of the 50 CD34⁺ AML specific genes (i.e. ankyrin repeat domain 28 (*ANKRD28*), guanine nucleotide binding protein, alpha 15 (*GNA15*) and UDP-glucose pyrophosphorylase 2 (*UGP2*)) were associated with a significant poorer OS in both cohorts of normal karyotype AML at the significance of $P < .01$ (Supplemental Table 5). For the cohort of 218 normal karyotype AMLs also event free survival (EFS) data were available for further analyses. A significant association between the continuous transcript levels of *ANKRD28*, *GNA15* and *UGP2* with unfavourable EFS was evident (*ANKRD28*: HR: 1.30, 95% CI: 1.07-1.56, $P = .007$; *GNA15*: HR: 1.23, 95% CI: 1.05-1.45, $P = .012$ and *UGP2*: HR: 2.04, 95% CI: 1.34-3.11, $P = .001$). Subsequently, these three genes were summed and divided in tertiles (low, intermediate and high expression). As expected, higher transcript levels of these three genes were strongly associated with poorer OS and EFS in the cohort of 218 normal karyotype AMLs (Figure 3A and B, $P = .007$ and $P = .006$, respectively). Also in the independent cohort of 163 normal karyotype AMLs, Kaplan-Meier curves showed that patients with high expression levels of these three genes had a significantly worse OS compared to patients with low expression levels (Figure 3C, $P < .001$). Of note, we also analyzed the prognostic relevance of the sum of expression of the top 50 AML CD34⁺ specific genes. Univariate cox regression analyses revealed that higher transcript levels of the sum of these 50 genes were associated with poorer outcome (Metzeler dataset, OS: HR: 1.03, 95% CI 1.01-1.04, $P = 0.003$; Valk dataset: OS: HR: 1.02, 95% CI 1.00-1.03, $P = 0.018$ and EFS: HR: 1.02, 95% CI 1.00-1.03, $P = 0.022$). Comparable results were observed for analyses of the sum of the top 100 and top 200 AML CD34⁺ specific genes and survival (data not shown). As a control we analyzed the prognostic relevance of the sum of the top 50 AML CD34⁻ genes. Interestingly, this revealed no significant correlation with survival in both cohorts of NK AML. Also the signature of genes differentially expressed between AML CD34⁺ and CD34⁻ irrespective of clustering of the CD34⁺ and CD34⁻ transcriptomes (123 genes) was not associated with clinical outcome.

Prognostic value of *ANKRD28*, *GNA15* and *UGP2* and the sum expression level in the context of known disease-related risk factors

Well known disease-related risk factors such as *FLT3*-ITD, *NPM1c*⁺ and *CEBPA* mutation status and white blood cell count (WBC) were considered in a multivariate analysis in the cohort of 218 normal

karyotype AMLs for both OS (*ANKRD28*: HR: 1.32, 95% CI: 1.07-1.62, $P = .008$; *GNA15*: HR: 1.22, 95% CI: 1.02-1.47, $P = .033$; *UGP2*: HR: 1.86, 95% CI: 1.17-2.97, $P = .009$ and sum expression: HR: 1.20, 95% CI 1.08-1.33, $P = 0.001$) as well as EFS (*ANKRD28*: HR: 1.31, 95% CI: 1.09-1.59, $P = .005$; *GNA15*: HR: 1.21, 95% CI: 1.02-1.43, $P = .030$; *UGP2*: HR: 1.73, 95% CI: 1.11-2.68, $P = .015$ and sum expression: HR: 1.17, 95% CI 1.07-1.29, $P = 0.001$) (details in Table 2). These data indicate that the continuous transcript levels of *ANKRD28*, *GNA15* and *UGP2* when analyzed as individual genes as well as the sum of these three genes are independent risk indicators for both OS and EFS.

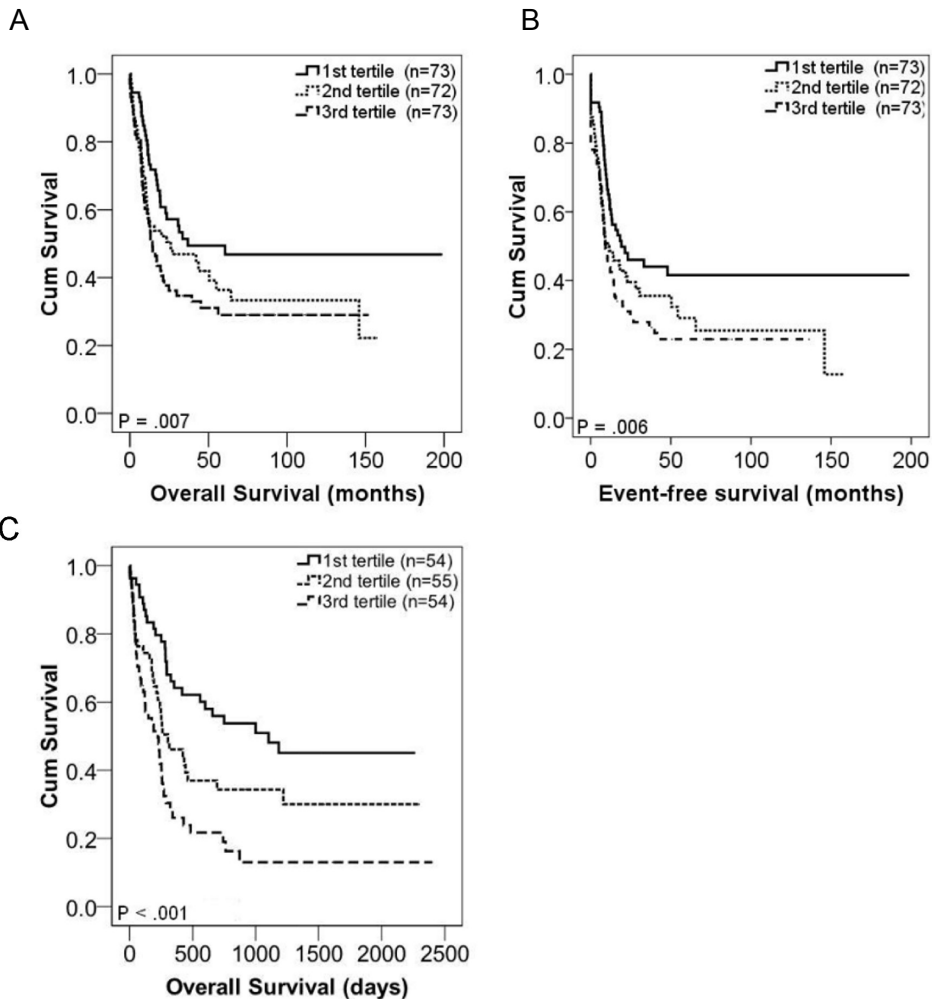


Figure 3 Kaplan-Meier plots show the overall survival (OS) (A) and event-free survival (EFS) (B) in 218 adult normal karyotype AML patient subgroups. The expression levels of *ANKRD28*, *GNA15* and *UGP2* were summed and divided in tertiles (low, intermediate and high expression). P-values are given for the comparison first versus third tertile. (C) Kaplan-Meier plot shows the OS in an independent cohort of 163 adult normal karyotype AML patients. The expression levels of *ANKRD28*, *GNA15* and *UGP2* were summed and divided in tertiles (low, intermediate and high expression). P-value is given for the comparison first versus third tertile.

Table 2 Multivariate cox regression analyses

Variable	Overall survival		Event-free survival	
	HR (95% CI)	P	HR (95% CI)	P
<i>ANKRD28</i>	1.34 (1.09-1.65)	0.006	1.32 (1.09-1.61)	0.004
<i>FLT3</i> -ITD #	1.47 (0.99-2.17)	0.054	1.45 (1.00-2.10)	0.049
<i>NPM1</i> §	0.58 (0.39-0.86)	0.006	0.59 (0.41-0.86)	0.005
<i>CEBPA</i> ^	0.58 (0.32-1.05)	0.071	0.59 (0.34-1.03)	0.064
WBC	1.00 (1.00-1.00)	0.025	1.00 (1.00-1.00)	0.143
<i>GNA15</i>	1.25 (1.03-1.52)	0.025	1.20 (1.01-1.44)	0.042
<i>FLT3</i> -ITD #	1.41 (0.94-2.13)	0.100	1.41 (0.95-2.10)	0.085
<i>NPM1</i> §	0.67 (0.46-0.99)	0.042	0.68 (0.47-0.98)	0.040
<i>CEBPA</i> ^	0.55 (0.30-1.01)	0.049	0.58 (0.33-1.01)	0.056
WBC	1.00 (1.00-1.01)	0.008	1.00 (1.00-1.00)	0.063
<i>UGP2</i>	1.85 (1.16-2.95)	0.010	1.71 (1.10-2.65)	0.018
<i>FLT3</i> -ITD #	1.47 (0.98-2.18)	0.060	1.44 (0.99-2.11)	0.058
<i>NPM1</i> §	0.69 (0.47-1.02)	0.061	0.69 (0.48-1.00)	0.050
<i>CEBPA</i> ^	0.64 (0.36-1.16)	0.142	0.66 (0.38-1.14)	0.139
WBC	1.00 (1.00-1.00)	0.046	1.00 (1.00-1.00)	0.230
SUM EXPRESSION	1.20 (1.08-1.33)	0.001	1.17 (1.07-1.29)	0.001
<i>FLT3</i> -ITD #	1.29 (0.85-1.94)	0.229	1.30 (0.88-1.91)	0.188
<i>NPM1</i> §	0.62 (0.42-0.91)	0.015	0.64 (0.44-0.92)	0.015
<i>CEBPA</i> ^	0.52 (0.29-0.95)	0.034	0.55 (0.31-0.96)	0.035
WBC	1.00 (1.00-1.01)	0.008	1.00 (1.00-1.00)	0.066

HR indicates hazard ratio; CI, confidence interval, *FLT3*; fms-related tyrosine kinase 3, ITD, internal tandem duplication; *NPM1*, nucleophosmin 1 and *CEBPA*, CCAAT/enhancer binding protein α . SUM EXPRESSION indicates the sum of the expression of *ANKRD28*, *GNA15* and *UGP2*.

FLT3-ITD versus no *FLT3*-ITD.

§ *NPM1* mutation versus no *NPM1* mutation.

^ *CEBPA* mutation versus no *CEBPA* mutation.

Discussion

In recent years, major advances have been achieved in predicting the outcome of newly diagnosed AML patients. However, there is still need for more powerful and independent prognostic factors that can guide treatment decisions, especially for the large subgroup of patients presenting with normal karyotype AML. In the present study, GEP was performed on purified CD34⁺ AML cells and the results were compared with normal CD34⁺ cells. This allowed the identification of a top 50 list of genes which were differentially expressed in AML CD34⁺ versus normal CD34⁺ cells and versus AML CD34⁻ cells. Interestingly, for three of these genes (*ANKRD28*, *GNA15* and *UGP2*) a high transcript level was associated with a significant poorer OS in normal karyotype AML patients, as validated in two independent cohorts of normal karyotype AML patients. Importantly, the prognostic value of the continuous transcript levels of *ANKRD28*, *GNA15* and *UGP2* was shown to be independent from the well known risk factors *FLT3*-ITD, *NPM1c⁺* and *CEBPA* mutation status.

Genome-wide gene expression profiling techniques have been applied to derive prognostic signatures for AML that would identify subsets of patients with certain outcomes. In normal karyotype AML patients, Bullinger et al. distinguished two different prognostically relevant clusters using 133 probe sets.⁷ More recently, an additional study in normal karyotype AML identified a gene signature of 86 probe sets correlating significantly with OS in 163 patients. This was validated and confirmed in an independent cohort of 79 normal karyotype AML patients.³⁶ A limitation in applying these results to clinical practice is the number of genes that have to be studied. By using the more purified AML CD34⁺ population and a large cohort of normal bone marrow samples we were able to identify a small gene set with strong prognostic significance. A single PCR measurement of these three genes can be easily incorporated in routine diagnostic measurements.

To place the results of our study in the perspective of previous findings, we compared our list of top 50 AML CD34⁺ specific genes with previously identified prognostically relevant AML gene signatures. Importantly to note, our three identified target genes are not included in the previously described prognostically relevant sets of 133 genes⁷ and 86 genes³⁶. Even when considering all top 50 AML CD34⁺ specific genes, only one gene was found to overlap with one of the previously identified prognostic signatures. *FLT3* was identified previously by Bullinger et al.⁷ and is the first gene in our list of AML CD34⁺ specific genes, suggesting that upregulation of *FLT3* might be important in the biology of AML. Although almost no overlap was seen, the different identified signatures might nevertheless reflect the same underlying biology. Underestimation of the overlap in gene expression patterns can not be excluded since different studies used different array platforms and filtered datasets were used for the comparisons.

Very recently, also Gentles et al. defined a leukemic stem cell (LSC) gene expression signature and found this signature to be strongly associated with clinical outcome in AML.³⁸ The LSC signature identified by Gentles et al. was shown to be highly expressed not only in leukemic, but also in normal

hematopoietic stem cells. One could hypothesize that particularly those genes that are differentially expressed between leukemic and normal stem cells are of interest in the perspective of the biology of leukemia and also for prognosis and therapy. Indeed, while our top 50 AML CD34⁺ specific gene signature shows prognostic relevance, the signature of genes differentially expressed between AML CD34⁺ and CD34⁻ irrespective of clustering of the CD34⁺ and CD34⁻ transcriptomes (123 genes) is not associated with clinical outcome. These data strongly suggest the importance of comparing leukemic CD34⁺ transcriptomes with its normal counterpart. The importance of sorting the CD34⁺ AML population was reflected by the observation that survival could not be predicted by the top 50 of AML CD34⁻ specific genes.

GO analysis of the differentially expressed genes between AML CD34⁺ and CD34⁻ revealed biological processes involved in differentiation and apoptosis. Since the gene expression differences associated with differentiation do not seem to correlate with the contamination of lymphoid cells or pre-sorting blast percentages, we believe that these differences might reflect underlying biology of CD34⁺ and CD34⁻ AML cells and are not just determined by contaminating non-leukemic cells. This is in line with the idea that although leukemic CD34⁻ cells appear as immature blasts, still a hierarchy of maturation might be present in leukemia.³⁹ Some upregulated GO biological processes in CD34⁺ AML versus normal bone marrow are involved in metabolic processes, suggesting higher metabolic activity in leukemia. Future studies should be aimed at a further characterisation of these pathways in leukemias, and whether targeting these might provide alternative means to eradicate leukemic stem cells.

There is only limited data on the three target genes we identified. The protein encoded by *ANKRD28* (also known as PITK; Phosphatase Interactor Targeting K protein) was shown to selectively bind and dephosphorylate the transcriptional regulator heterogeneous nuclear ribonucleoprotein K by targeting protein phosphatase-1. Interestingly, a gene fusion of *ANKRD28* on 3p25 to *NUP98* on 11p15 was described in a patient with high risk myelodysplasia.⁴⁰ In the latter study, NIH/3T3 cells transfected with wild type *ANKRD28* showed suppressed colony formation compared to control cells whereas the *ANKRD28-NUP98* fusion induced malignant transformation.⁴⁰ The expression of G15 α -subunit (formerly the human isoform was named G16) encoded by *GNA15* seems to be restricted to immature hematopoietic and epithelial cells (reviewed in ⁴¹). Although poorly understood, G15 α -subunit was shown to regulate cell differentiation and apoptosis through modulation of MAPKs⁴² and transcription factors such as NF- κ B⁴³ and is capable of activating STAT3⁴⁴. Interestingly, the expression of the third target gene *UGP2* was recently shown to be upregulated by hypoxia in hepatocytes.⁴⁵ *UGP2* was put forward as a potential target of Hypoxia Inducible Factors (HIF), which in turn are required for hematopoiesis^{46,47} and hypothesized to be relevant in hematopoietic stem cell (HSC) maintenance in hypoxic osteoblastic bone marrow niches. Further studies are needed to elucidate specific roles of *UGP2* as well as *ANKRD28* and *GNA15* in AML.

In conclusion, our study suggests that risk classification of normal karyotype AML might be further improved by using a selective set of genes which are highly expressed in the leukemic stem cell-enriched CD34⁺ cell fraction.

Acknowledgements

This study was supported by a grant from the “Innovatie Fonds” UMCG, The Netherlands. We would like to acknowledge Prof.dr. B. Löwenberg and Dr. P.J.M. Valk (Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands) for providing clinical data of the NK AML gene expression set and for critical reading of the manuscript.

References

1. Vardiman JW, Thiele J, Arber DA et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* 2009;114:937-951.
2. Byrd JC, Mrozek K, Dodge RK et al. Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461). *Blood* 2002;100:4325-4336.
3. Falini B, Mecucci C, Tiacci E et al. Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N.Engl.J.Med.* 2005;352:254-266.
4. Nakao M, Yokota S, Iwai T et al. Internal tandem duplication of the *flt3* gene found in acute myeloid leukemia. *Leukemia* 1996;10:1911-1918.
5. Wouters BJ, Lowenberg B, Erpelinck-Verschueren CA et al. Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. *Blood* 2009;113:3088-3091.
6. Valk PJ, Verhaak RG, Beijnen MA et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. *N.Engl.J.Med.* 2004;350:1617-1628.
7. Bullinger L, Dohner K, Bair E et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N.Engl.J.Med.* 2004;350:1605-1616.
8. Alcalay M, Tiacci E, Bergomas R et al. Acute myeloid leukemia bearing cytoplasmic nucleophosmin (NPMc+ AML) shows a distinct gene expression profile characterized by up-regulation of genes involved in stem-cell maintenance. *Blood* 2005;106:899-902.
9. Radmacher MD, Marcucci G, Ruppert AS et al. Independent confirmation of a prognostic gene-expression signature in adult acute myeloid leukemia with a normal karyotype: a Cancer and Leukemia Group B study. *Blood* 2006;108:1677-1683.
10. Wouters BJ, Lowenberg B, Delwel R. A decade of genome-wide gene expression profiling in acute myeloid leukemia: flashback and prospects. *Blood* 2009;113:291-298.
11. Wouters BJ, Lowenberg B, Erpelinck-Verschueren CA et al. Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. *Blood* 2009;113:3088-3091.
12. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat.Med.* 1997;3:730-737.

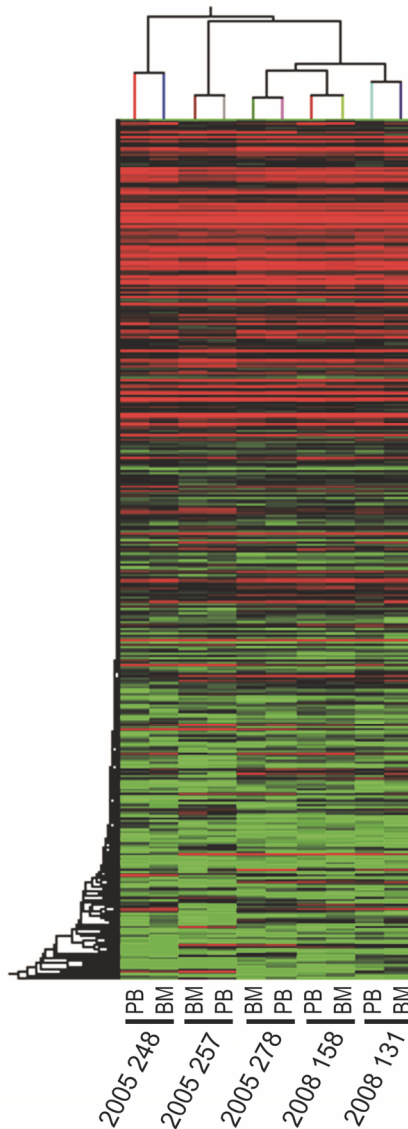
13. Lapidot T, Sirard C, Vormoor J et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* 1994;367:645-648.
14. Warner JK, Wang JC, Hope KJ, Jin L, Dick JE. Concepts of human leukemic development. *Oncogene* 2004;23:7164-7177.
15. Wang JC, Dick JE. Cancer stem cells: lessons from leukemia. *Trends Cell Biol.* 2005;15:494-501.
16. Schuringa JJ, Schepers H. Ex vivo assays to study self-renewal and long-term expansion of genetically modified primary human acute myeloid leukemia stem cells. *Methods Mol.Biol.* 2009;538:287-300.
17. van Goslign D, Schepers H, Rizo A et al. Establishing long-term cultures with self-renewing acute myeloid leukemia stem/progenitor cells. *Exp.Hematol.* 2007;35:1538-1549.
18. Blair A, Hogge DE, Ailles LE, Lansdorp PM, Sutherland HJ. Lack of expression of Thy-1 (CD90) on acute myeloid leukemia cells with long-term proliferative ability in vitro and in vivo. *Blood* 1997;89:3104-3112.
19. Blair A, Hogge DE, Sutherland HJ. Most acute myeloid leukemia progenitor cells with long-term proliferative ability in vitro and in vivo have the phenotype CD34(+)/CD71(-)/HLA-DR-. *Blood* 1998;92:4325-4335.
20. Blair A, Sutherland HJ. Primitive acute myeloid leukemia cells with long-term proliferative ability in vitro and in vivo lack surface expression of c-kit (CD117). *Exp.Hematol.* 2000;28:660-671.
21. Hosen N, Park CY, Tatsumi N et al. CD96 is a leukemic stem cell-specific marker in human acute myeloid leukemia. *Proc.Natl.Acad.Sci.U.S.A* 2007;104:11008-11013.
22. Jordan CT, Upchurch D, Szilvassy SJ et al. The interleukin-3 receptor alpha chain is a unique marker for human acute myelogenous leukemia stem cells. *Leukemia* 2000;14:1777-1784.
23. Taussig DC, Pearce DJ, Simpson C et al. Hematopoietic stem cells express multiple myeloid markers: implications for the origin and targeted therapy of acute myeloid leukemia. *Blood* 2005;106:4086-4092.
24. Taussig DC, Miraki-Moud F, Anjos-Afonso F et al. Anti-CD38 antibody-mediated clearance of human repopulating cells masks the heterogeneity of leukemia-initiating cells. *Blood* 2008;112:568-575.
25. Cornelissen JJ, van Putten WL, Verdonck LF et al. Results of a HOVON/SAKK donor versus no-donor analysis of myeloablative HLA-identical sibling stem cell transplantation in first remission acute myeloid leukemia in young and middle-aged adults: benefits for whom? *Blood* 2007;109:3658-3666.

26. van der Holt B, Breems DA, Berna Beverloo H et al. Various distinctive cytogenetic abnormalities in patients with acute myeloid leukaemia aged 60 years and older express adverse prognostic value: results from a prospective clinical trial. *Br.J.Haematol.* 2007;136:96-105.
27. Lowenberg B, Pabst T, Vellenga E et al. Cytarabine dose for acute myeloid leukemia. *N.Engl.J.Med.* 2011;364:1027-1036.
28. Yokota S, Kiyoi H, Nakao M et al. Internal tandem duplication of the FLT3 gene is preferentially seen in acute myeloid leukemia and myelodysplastic syndrome among various hematological malignancies. A study on a large series of patients and cell lines. *Leukemia* 1997;11:1605-1609.
29. Irizarry RA, Bolstad BM, Collin F et al. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* 2003;31:e15.
30. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics.* 2003;19:185-193.
31. Alter O, Brown PO, Botstein D. Singular value decomposition for genome-wide expression data processing and modeling. *Proc.Natl.Acad.Sci.U.S.A* 2000;97:10101-10106.
32. Crijns AP, Gerbens F, Plantinga AE et al. A biological question and a balanced (orthogonal) design: the ingredients to efficiently analyze two-color microarrays with Confirmatory Factor Analysis. *BMC.Genomics* 2006;7:232.
33. Crijns AP, Fehrmann RS, de Jong S et al. Survival-related profile, pathways, and transcription factors in ovarian cancer. *PLoS.Med.* 2009;6:e24.
34. Sherlock G. Analysis of large-scale gene expression data. *Brief.Bioinform.* 2001;2:350-362.
35. de Jonge HJ, de Bont ES, Valk PJ et al. AML at older age: age-related gene expression profiles reveal a paradoxical down-regulation of p16INK4A mRNA with prognostic significance. *Blood* 2009;114:2869-2877.
36. Metzeler KH, Hummel M, Bloomfield CD et al. An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia. *Blood* 2008;112:4193-4201.
37. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc.Natl.Acad.Sci.U.S.A* 1998;95:14863-14868.
38. Gentles AJ, Plevritis SK, Majeti R, Alizadeh AA. Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia. *JAMA* 2010;304:2706-2715.

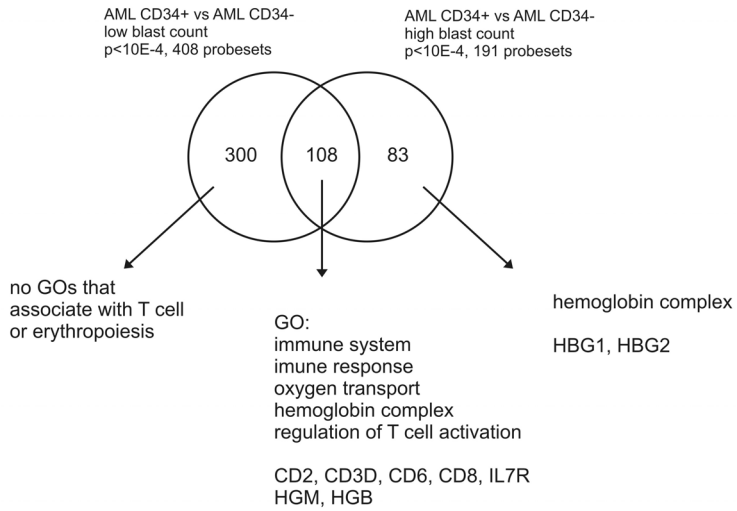
39. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat.Med.* 1997;3:730-737.
40. Ishikawa M, Yagasaki F, Okamura D et al. A novel gene, ANKRD28 on 3p25, is fused with NUP98 on 11p15 in a cryptic 3-way translocation of t(3;5;11)(p25;q35;p15) in an adult patient with myelodysplastic syndrome/acute myelogenous leukemia. *Int.J.Hematol.* 2007;86:238-245.
41. Giannone F, Malpeli G, Lisi V et al. The puzzling uniqueness of the heterotrimeric G15 protein and its potential beyond hematopoiesis. *J.Mol.Endocrinol.* 2010;44:259-269.
42. Heasley LE, Storey B, Fanger GR et al. GTPase-deficient G alpha 16 and G alpha q induce PC12 cell differentiation and persistent activation of cJun NH2-terminal kinases. *Mol.Cell Biol.* 1996;16:648-656.
43. Yang M, Sang H, Rahman A et al. G alpha 16 couples chemoattractant receptors to NF-kappa B activation. *J.Immunol.* 2001;166:6885-6892.
44. Lo RK, Cheung H, Wong YH. Constitutively active Galpha16 stimulates STAT3 via a c-Src/JAK- and ERK-dependent mechanism. *J.Biol.Chem.* 2003;278:52154-52165.
45. Pescador N, Villar D, Cifuentes D et al. Hypoxia promotes glycogen accumulation through hypoxia inducible factor (HIF)-mediated induction of glycogen synthase 1. *PLoS.One.* 2010;5:e9644.
46. Adelman DM, Maltepe E, Simon MC. Multilineage embryonic hematopoiesis requires hypoxic ARNT activity. *Genes Dev.* 1999;13:2478-2483.
47. Scortegagna M, Morris MA, Oktay Y, Bennett M, Garcia JA. The HIF family member EPAS1/HIF-2alpha is required for normal hematopoiesis in mice. *Blood* 2003;102:1634-1640.

Supplemental tables can be found at the Leukemia website
(<http://www.nature.com/leu/journal/v25/n12/supinfo/leu2011172s1.html>)

Supplemental figures



Supplemental Figure 1 Unsupervised clustering analysis (Pearson uncentered with average linkage) of paired peripheral blood (PB) and bone marrow (BM) samples from five AML patients. A paired t-test comparing gene expression in AML PB versus BM revealed no significant differences.



Supplemental Figure 2 VENN diagram showing differentially expressed genes between AML CD34⁺ and CD34⁻ in samples with a low percentage of pre-sorting blasts (left) and a high percentage of pre-sorting blasts (right). Results of GO analyses are indicated.



