Association of a murine leukaemia stem cell gene signature based on nucleostemin promoter activity with prognosis of acute myeloid leukaemia in patients

著者	モハメド アリ エル サイド アリ
著者別表示	Mohamed Ali El Sayed Ali
journal or publication title	博士論文本文Full
学位授与番号	13301甲第4103号
学位名	博士（医学）
学位授与年月日	2014-09-26
URL	http://hdl.handle.net/2297/41994

doi: <https://doi.org/10.1016/j.bbrc.2014.06.066>

Association of a murine leukaemia stem cell gene signature based on nucleostemin promoter activity with prognosis of acute myeloid leukaemia in patients

Mohamed A.E. Ali^a, Kazuhito Naka^b, Akiyo Yoshida^c, Kyoko Fuse^a, Atsuo Kasada^a, Takayuki Hoshii^a, Yuko Tadokoro^a, Masaya Ueno^a, Kumiko Ohta^a, Masahiko Kobayashi^a, Chiaki Takahashi^c, Atsushi Hirao^{a*}

^aDivision of Molecular Genetics, ^bExploratory Project on Cancer Stem Cells, and

^cDivision of Oncology and Molecular Biology, Cancer Research Institute, Kanazawa University, Kakuma-machi, Kanazawa, Ishikawa, 920-1192, Japan

*Corresponding author.

E-mail address: ahirao@staff.kanazawa-u.ac.jp (A. Hirao).

Phone: +81-76-264-6755

Abstract

Acute myeloid leukaemia (AML) is a heterogeneous neoplastic disorder in which a subset of cells function as leukaemia-initiating cells (LICs). In this study, we prospectively evaluated the leukaemia-initiating capacity of AML cells fractionated according to the expression of a nucleolar GTP binding protein, nucleostemin (NS). To monitor NS expression in living AML cells, we generated a mouse AML model in which green fluorescent protein (GFP) is expressed under the control of a region of the NS promoter (NS-GFP). In AML cells, NS-GFP levels were correlated with endogenous NS mRNA. AML cells with the highest expression of NS-GFP were very immature blast-like cells, efficiently formed leukaemia colonies *in vitro*, and exhibited the highest leukaemia-initiating capacity *in vivo*. Gene expression profiling analysis revealed that cell cycle regulators and nucleotide metabolism-related genes were highly enriched in a gene set associated with leukaemia-initiating capacity that we termed the 'leukaemia stem cell gene signature'. This gene signature stratified human AML patients into distinct clusters that reflected prognosis, demonstrating that the mouse leukaemia stem cell gene signature is significantly associated with the malignant properties of human AML. Further analyses of gene regulation in leukaemia stem cells could provide novel insights into diagnostic and therapeutic approaches to AML.

Keywords:

Acute myeloid leukaemia, Leukaemia stem cells, Nucleostemin

1. Introduction

Acute myeloid leukaemia (AML) is a heterogeneous neoplastic disorder characterized by substantial cellular heterogeneity. Only a rare subset of cells is assumed to have the ability to self-renew and to initiate and sustain the disease [1]. Understanding the functional regulatory machinery of these leukaemia initiating cells (LICs) is critically important for designing an efficient therapeutic approach, but these cells have not yet been characterized in detail.

Nucleostemin (NS, also known as GNL3) encodes a GTP-binding protein that is mainly located in rRNA-free sites of the nucleolus [2]. NS contributes to ribosomal biogenesis by interacting with nucleolar proteins involved in pre-rRNA processing [3]. Moreover, NS is essential for the maintenance of nucleolar architecture and the integrity of nucleolar RNA-protein complexes [4]. NS also interacts with telomeric repeat-binding factor 1 (TRF1) to facilitate its degradation, which enhances the maintenance of telomere length [5]. NS and GNL3L form a complex with the telomerase catalytic subunit, human telomerase reverse transcriptase (hTERT) [6]. NS was originally discovered during research on genes that are highly expressed in neural stem cells, embryonic stem (ES) cells, and developing organs during embryogenesis [7]. NS functions as a reprogramming factor for induced pluripotent stem (iPS) cells [8], indicating that NS is essential for the maintenance of the undifferentiated status of ES/iPS cells. In neural stem cells, NS contributes to genomic stability [9], and NS deficiency in neural stem cells causes replication-dependent DNA damage that is associated with defective self-renewal. In haematopoietic stem cells, NS knockdown impairs the long-term reconstitution capacity and induces apoptosis as a result of accumulated DNA damage [10].

We previously generated transgenic mice in which green fluorescent protein (GFP) is expressed under the control of a particular region of the NS promoter (NS-GFP). In this system, we successfully identified the stem cell population among neonatal testicular cells, liver cells, and foetal brain tissues [11,12,13] as well as tumour-initiating cells, conceptually termed ‘cancer stem cells,’ in brain tumours and germ cell tumours [13,14]. Consistent with our studies, NS-enriched mammary tumour cells are highly tumorigenic [15]. Although the mechanisms are still unclear, there are particular programs that control NS expression and functions that are also commonly involved in maintaining stem cell properties in normal and malignant tissues. Recently, a study with newly diagnosed AML patients reported that high expression of NS is closely correlated with the percentage of immature blast cells and with haematopoietic stem/progenitor cell surface markers and that high NS transcript levels are found in AML patients with poor prognosis [16]; this suggests that NS expression may be an indicator of immature properties of AML cells. In this study, we prospectively evaluated the leukaemia-initiating capacity of subpopulations of AML cells with varying NS expression by using NS-GFP transgenic mice. We found that a leukaemia stem cell gene signature containing genes differentially expressed in cells with high and low NS promoter activity could also be used to stratify human AML patients into distinct clusters that reflected prognosis.

2. Materials and methods

2.1. Generation of an AML model

NS-GFP transgenic mice were generated as described previously [11]. All procedures were performed in accordance with the animal care guidelines of Kanazawa University, Kanazawa, Japan. A murine AML model was generated as previously reported [17]. Briefly, c-Kit⁺Sca-1⁺Lineage⁻ cells isolated from bone marrow (BM) of NS-GFP transgenic mice were infected with a retrovirus carrying *Hoxa9* and *Meis1a* (MSCV-Hoxa9-IRES-Meis1a) and transplanted into lethally irradiated syngeneic mice along with 5×10^5 normal BM mononuclear cells (rescue cells).

2.2. Leukaemia initiating capacity *in vitro* and *in vivo*.

The leukaemia-initiating capacity of fractionated NS-GFP cells was evaluated with a colony formation assay (MethoCult GF M3434, Stem Cell Technologies Inc., Vancouver, Canada) and by transplantation into lethally irradiated recipient mice, as previously described [18].

2.3. Cell cycle analysis

The recipient mice were sacrificed 2 hours after intraperitoneal injection of 5-bromo-2'-deoxyuridine (BrdU, 100 mg/kg, Sigma-Aldrich., Missouri, USA). The fractionated cells were stained with FITC-conjugated anti-BrdU antibody (BD Biosciences., New Jersey, USA) and propidium iodide (PI, 20 µg/ml, Sigma-Aldrich).

2.4. Quantitative RT-PCR analysis

RNA samples were purified from fractionated leukaemia cells (1.5×10^4) using an

RNeasy kit (QIAGEN, Tokyo, Japan) and reverse-transcribed using an Advantage RT-for-PCR kit (Clontech, Takara Bio Inc., Shiga, Japan). PCR for NS was performed using a Dice PCR Thermal Cycler (Takara Bio Inc.) as previously reported [11].

2.5. Preprocessing of microarray data

Gene expression profiling of mouse AML cells was performed as described in the Supplementary Methods. Briefly, total RNA was isolated from (a) c-Kit⁻NS-GFP^{low}, (b) c-Kit⁻NS-GFP^{middle}, (c) c-Kit⁺NS-GFP^{middle}, and (d) c-Kit⁺NS-GFP^{high} cells, followed by cDNA synthesis. cDNA microarray analysis was performed with a GeneChip Mouse 430 2.0 (Affymetrix Inc., High Wycombe, UK). The complete microarray data set is available at the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/, accession number GSE 58032). Unsupervised hierarchical clustering was performed on the normalized and log-transformed data of 17,419 probes. Similarity was measured by Euclidean distance metric, and the average linkage was used to define the linking distance between clusters. To select probes whose expression changed progressively among the 4 subpopulations, we extracted 3,494 probes that showed gradual increases or decreases in expression as AML differentiation increased. Among the 3,494 probes, 500 probes that exhibited in the largest ratios between c-Kit⁻NS-GFP^{low} and c-Kit⁺NS-GFP^{high} were extracted. The 500 probes corresponded to 382 mouse gene symbols, forming the 'leukaemia stem cell signature' shown in Supplementary Table 1. Pathway analysis using the 382 mouse gene symbols was performed by using DAVID v6.7 (<http://david.abcc.ncifcrf.gov/>) with Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). The 382 mouse gene symbols corresponded to 338 human gene symbols, as identified at Mouse Genome Informatics

(<http://www.informatics.jax.org/>).

2.6. Analysis of human AML patient data

We used data from 179 AML patient samples published by The Cancer Genome Atlas Research Network [19] (<http://cancergenome.nih.gov/>). For patient clustering, we used 318 human genes from AML patients that are included in the database and were among the 338 human genes in the leukaemia stem cell signature. Clustering of AML patients was performed by using the R ‘cluster’ package, PAM (Partitioning Around Medoids; <http://stat.ethz.ch/R-manual/R-patched/library/cluster/html/pam.html>).

2.7. Statistical analyses

For analysis of mouse AML samples, statistical differences among multiple groups were evaluated by a one-way ANOVA and Tukey as a post-hoc test. Statistical differences in Kaplan-Meier survival curves were determined using the log-rank test. For analysis of human AML data, age of patients, initial white blood cell (WBC) number of peripheral blood, and percent of blasts among mononuclear BM cells were compared among the clusters with Student’s *t*-test or a one-way ANOVA/Tukey. Patients’ sex, French-American-British (FAB) classification, and molecular risk classification were compared with Fisher’s exact test and Bonferroni as a post-hoc test. Overall survival was plotted using the Kaplan–Meier estimator and compared with the log-rank test. The statistical analyses were performed using R software, and a two-sided significance level of $\alpha < 0.05$ was used.

3. Results

3.1. Fractionation of living AML cells based on NS expression level

We first investigated whether our NS-GFP reporter system could be used to monitor NS mRNA in AML cells *in vivo*. To do so, we infected c-Kit⁺Sca-1⁺Lineage⁻ BM mononuclear cells from NS-GFP transgenic mice with a retrovirus expressing both the *Hoxa9* and *Meis1a* genes. This procedure generates a well-characterized AML once the transduced cells are transplanted into lethally irradiated syngeneic hosts [20]. As expected, all recipients of NS-GFP transgenic cells transduced with *Hoxa9* and *Meis1a* developed leukaemia. Morphological analysis showed that the recipient BM was filled with leukaemia cells, including many blast-like cells and some more mature myeloid-like cells such as neutrophils and monocytes (data not shown), which is categorized as AML with maturation under the WHO classification (FAB Classification M2) [20]. We fractionated the approximately 95% of BM cells that were positive for NS-GFP into 3 subpopulations depending on GFP fluorescence intensity: NS-GFP⁺⁺⁺ (approximately top 5%), NS-GFP⁺⁺ (next 45%), and NS-GFP⁺ (next lower 45%); the remaining 5% were NS-GFP⁻ (**Fig. 1A**). GFP fluorescence intensity in all 4 fractions closely paralleled the endogenous NS mRNA expression as measured by RT-PCR (**Fig. 1B**).

3.2. Correlation of maturation stages of leukaemia cells with NS-GFP level

To characterize the fractionated AML cells, we analysed several surface markers for haematopoietic differentiation (**Fig. 1C**). As expected, all 4 fractions expressed the myeloid lineage marker Mac-1 but not markers for the T, B, or erythroid lineages. Sca-1, which is expressed in haematopoietic stem cells but not myeloid progenitor cells, was

not expressed in any of the fractionated cells. Most NS-GFP⁺⁺⁺ leukaemia cells expressed c-Kit, consistent with the positive correlation between NS and c-KIT [16]. The expression of c-Kit decreased with decreasing GFP intensity, so that the NS-GFP⁺⁺ and NS-GFP⁺ cell populations consisted of both c-Kit⁺ and c-Kit⁻ cells (**Fig. 2A**), and NS-GFP⁻ cells did not express c-Kit. Therefore, we designated these populations as (a) c-Kit⁻NS-GFP^{low}, (b) c-Kit⁻NS-GFP^{middle}, (c) c-Kit⁺NS-GFP^{middle}, and (d) c-Kit⁺NS-GFP^{high}, reflecting the combined patterns of NS-GFP and c-Kit expression.

The morphology of leukaemia cells differed remarkably among the subpopulations (**Fig. 2B**). The c-Kit⁺NS-GFP^{high} cells appeared to be very immature blast-like cells with large nuclei, whereas c-Kit⁻NS-GFP^{low} cells appeared to be mature myeloid cells with lobular nuclei. The NS-GFP^{middle} cells (both c-Kit⁺ and c-Kit⁻) had intermediate morphological features. Consistent with these findings, the c-Kit⁺NS-GFP^{high} cells were larger than cells in the other subpopulations as evaluated by flow cytometry (**Fig. 2C**). We next examined the gene expression profiles of each of the 4 leukaemic cell fractions to see if there was a relationship between GFP fluorescence intensity and pattern of gene expression. The gene expression profile of c-Kit⁺NS-GFP^{high} cells was closer to that of c-Kit⁺NS-GFP^{middle} cells than to that of c-Kit⁻NS-GFP^{middle} and c-Kit⁻NS-GFP^{low} cells, whereas the gene expression pattern of c-Kit⁻NS-GFP^{low} was closest to that of c-Kit⁻NS-GFP^{middle} (**Fig. 2D**). These progressive changes in morphology, cell surface marker profile, and gene expression pattern suggested that c-Kit⁺NS-GFP^{high} leukaemia cells differentiate into c-Kit⁺NS-GFP^{middle} and c-Kit⁻NS-GFP^{middle} cells, which subsequently differentiate into c-Kit⁻NS-GFP^{low} cells.

3.3. Characterization of undifferentiated leukaemia cells

We evaluated the cell cycle status of the fractionated NS-GFP leukaemia cells *in vivo* by measuring BrdU incorporation and DNA content (**Fig. 3A**). The proportion of BrdU-positive cells was significantly higher in c-Kit⁺NS-GFP^{high} cells more than the other subpopulations (**Fig. 3B**), indicating that these immature blast-like cells were actively cycling. To characterise the gene signature of undifferentiated leukaemia cells, we selected 3,494 probes whose expression gradually increased or decreased along with the progressive changes among the 4 subpopulations. Among these 3,494 probes, we further selected 500 probes (382 genes) that exhibited the largest ratios between differentiated (c-Kit⁻NS-GFP^{low}) and undifferentiated (c-Kit⁺NS-GFP^{high}) cells (**Supplementary Table 1**). Consistent with the data shown in Fig. 1B, the NS gene was included in the gene set. Pathway analysis with the 382 genes showed that cell cycle regulators and nucleotide metabolism-related genes were highly enriched (**Supplementary Table 2**). These data are consistent with the dramatic differences in cell cycle status between the cell fractions (**Fig. 3B**).

3.4. Enrichment of LICs in cells expressing high levels of NS-GFP *in vitro* and *in vivo*

To investigate properties of the fractionated cells, we performed *in vitro* colony-forming assays with the 4 types of fractionated AML cells shown in Fig. 1A (NS-GFP⁺⁺⁺, NS-GFP⁺⁺, NS-GFP⁺, and NS-GFP⁻). We found that NS-GFP⁺⁺⁺ cells exhibited the highest colony-forming ability among the subpopulations. Both NS-GFP⁺⁺⁺ and NS-GFP⁺⁺ leukaemia cells readily formed colonies, whereas NS-GFP⁻ did not (**Fig. 3C**), indicating that NS expression is correlated with colony-forming ability.

Next, we evaluated leukaemia-initiating capacity *in vivo* with a transplantation assay. When we transplanted 100 cells of each of the 4 fractionated leukaemia cell populations into separate groups of mice, recipient mice in all groups developed leukaemia (data not shown), indicating that the frequency of LICs was very high. Therefore, to distinguish between the cell populations, we transplanted 10 cells of each. In this condition, we found that only the NS-GFP⁺⁺⁺ cells generated a secondary leukaemia (**Fig. 3D**). Thus, LICs are enriched in cells with highest level of NS-GFP.

3.5. Association of a murine leukaemia stem cell gene signature with prognosis of human AML patients

Finally, we investigated whether NS is associated with the properties of human AML cells. To do so, we first evaluated NS gene expression in human AML patients by using the public database developed by the Cancer Genome Atlas (TCGA) Research Network [19]. Unexpectedly, we did not find a correlation between human NS expression and FAB classification, molecular risk, or prognosis (**Supplementary Fig. 1**). We therefore evaluated the relationship between the leukaemia stem cell gene signature shown above and characteristics of AML patients. We selected 318 human genes corresponding to the 500 mouse probes described above (**Supplementary Table 1**). Using this human gene set, we performed a clustering algorithm related to the k-medoids algorithm (Partitioning Around Medoids; PAM) for 179 AML patient samples. Because the clustering analysis showed that the average silhouette width was higher for $k = 2$ or 3 than for $k = 4$ or 5 , we first stratified AML into two clusters (Cluster 1, $n = 34$; Cluster 2, $n = 145$) (**Fig. 4A**). Patients in the two clusters showed significant differences in age, number of WBCs, FAB classification, and molecular risks (**Supplementary Table 3A**).

Cluster 1 consisted of younger patients who had lower WBC numbers than patients in Cluster 2. All patients with M3 FAB classification were included in Cluster 1, whereas all patients whose leukaemia was classified as M0, M6, or M7, which are relatively poor prognostic types, were included in Cluster 2 (**Fig. 4B**). Molecular risk classification showed that patients in Cluster 1 exhibited more good prognostic factors and fewer intermediate and poor prognostic factors than those in Cluster 2 (**Fig. 4C and Supplementary Table 3A**). Consistent with these findings, the overall survival rate was significantly higher for Cluster 1 patients than for Cluster 2 patients (**Fig. 4D**). When we stratified AML patients into three clusters, we found that one cluster mainly consisted of patients whose leukaemia was classified as M3 and that there were significant differences in prognosis between the clusters (**Supplementary Fig. 2 and Supplementary Table 3B**). Thus, the gene signature of leukaemia stem cells in a mouse model is correlated with the malignant status of human AML.

4. Discussion

We successfully identified LICs by monitoring NS expression in a mouse AML model. In this system, endogenous NS correlated well with the undifferentiated status of leukaemia cells *in vivo*. In normal haematopoiesis, NS is more highly expressed in myeloid progenitors and haematopoietic stem cells than in differentiated haematopoietic cells [10]. Because a recent study indicated that AML stem cells are immunophenotypically similar to progenitors, including lymphoid-primed multipotent progenitors and granulocyte-macrophage progenitors [21], the higher expression of NS in myeloid progenitor cells may be conserved in leukaemia stem cells. Our data in the AML mouse model suggest that NS plays a critical role in maintaining undifferentiated status in AML cells, as has been shown in ES/iPS cells [8].

There are critical differences in the characteristics of AML between mouse and human. We found that NS gene expression itself was not correlated with the prognostic characteristics of AML in patients in the TCGA database, which was inconsistent with a previously reported result [16]. Although mouse NS was included in the leukaemia stem cell gene signature, NS expression levels were not significantly different among AML patient clusters in our study (data not shown), presumably due to heterogeneity among patients in the regulation of human NS expression. In addition, LICs from patient samples and the experimental mice also exhibited critical differences. For example, although human LICs are rare and cycle slowly [22], c-Kit⁺NS-GFP^{high} leukaemia cells were actively cycling in our mouse model. Therefore, the gene signature identified in this experiment may not be directly associated with the endogenous pattern of NS expression. Even though NS itself may not function as a potent driver of malignant

status in AML patients, our data suggest that a network of genes associated with NS promoter activity affects the malignant status of human and mouse AML cells.

In clinical investigations, high NS expression is associated with poor prognosis in breast cancer [23], oesophageal cancer [24], and gastric and liver cancer [25]. Therefore, the core network of the leukaemia stem cell signature may be conserved in stem cells in solid tumours. In this study, we identified a gene set associated with AML that included cell cycle and nucleotide metabolism–related genes. These genes may coordinate stem cell properties with response to DNA damage and telomere stability to control self-renewal activity in normal and malignant cells, including solid tumours. Further analysis of NS-related genes may contribute to the understanding of the factors that regulate a variety of cancers.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We thank Dr. Guy Sauvageau for providing the MSCV-Hoxa9-IRES-Meis1a plasmid. This study was supported by a Grant-in-Aid for Scientific Research (25130705, 24240119) from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

- [1] S.J. Horton, B.J. Huntly, Recent advances in acute myeloid leukemia stem cell biology, *Haematologica* 97 (2012) 966-974.
- [2] J.C. Politz, I. Polena, I. Trask, D.P. Bazett-Jones, T. Pederson, A nonribosomal landscape in the nucleolus revealed by the stem cell protein nucleostemin, *Mol. Biol. Cell* 16 (2005) 3401-3410.
- [3] L. Romanova, A. Grand, L.Y. Zhang, S. Rayner, N. Katoku-Kikyo, S. Kellner, N. Kikyo, Critical Role of Nucleostemin in Pre-rRNA Processing, *J. Biol. Chem.* 284 (2009) 4968-4977.
- [4] L. Romanova, S. Kellner, N. Katoku-Kikyo, N. Kikyo, Novel Role of Nucleostemin in the Maintenance of Nucleolar Architecture and Integrity of Small Nucleolar Ribonucleoproteins and the Telomerase Complex, *J. Biol. Chem.* 284 (2009) 26685-26694.
- [5] L.J. Meng, J.K. Hsu, Q.B. Zhu, T. Lin, R.Y.L. Tsai, Nucleostemin inhibits TRF1 dimerization and shortens its dynamic association with the telomere, *J. Cell Sci.* 124 (2011) 3706-3714.
- [6] N. Okamoto, M. Yasukawa, C. Nguyen, V. Kasim, Y. Maida, R. Possemato, T. Shibata, K.L. Ligon, K. Fukami, W.C. Hahn, K. Masutomi, Maintenance of tumor initiating cells of defined genetic composition by nucleostemin, *Proc. Natl. Acad. Sci. U. S. A.* 108 (2011) 20388-20393.
- [7] R.Y.L. Tsai, R.D.G. McKay, A nucleolar mechanism controlling cell proliferation in stem cells and cancer cells, *Gene Dev.* 16 (2002) 2991-3003.
- [8] J. Qu, J.M. Bishop, Nucleostemin maintains self-renewal of embryonic stem cells and promotes reprogramming of somatic cells to pluripotency, *J. Cell Biol.* 197

(2012) 731-745.

- [9] L. Meng, T. Lin, G. Peng, J.K. Hsu, S. Lee, S.Y. Lin, R.Y. Tsai, Nucleostemin deletion reveals an essential mechanism that maintains the genomic stability of stem and progenitor cells, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 11415-11420.
- [10] M. Yamashita, E. Nitta, G. Nagamatsu, Y.M. Ikushima, K. Hosokawa, F. Arai, T. Suda, Nucleostemin is indispensable for the maintenance and genetic stability of hematopoietic stem cells, *Biochem. Biophys. Res. Commun.* 441 (2013) 196-201.
- [11] M. Ohmura, K. Naka, T. Hoshii, T. Muraguchi, H. Shugo, A. Tamase, N. Uema, T. Ooshio, F. Arai, K. Takubo, G. Nagamatsu, I. Hamaguchi, M. Takagi, M. Ishihara, K. Sakurada, H. Miyaji, T. Suda, A. Hirao, Identification of stem cells during prepubertal spermatogenesis via monitoring of nucleostemin promoter activity, *Stem Cells* 26 (2008) 3237-3246.
- [12] H. Shugo, T. Ooshio, M. Naito, K. Naka, T. Hoshii, Y. Tadokoro, T. Muraguchi, A. Tamase, N. Uema, T. Yamashita, Y. Nakamoto, T. Suda, S. Kaneko, A. Hirao, Nucleostemin in injury-induced liver regeneration, *Stem Cells Dev.* 21 (2012) 3044-3054.
- [13] A. Tamase, T. Muraguchi, K. Naka, S. Tanaka, M. Kinoshita, T. Hoshii, M. Ohmura, H. Shugo, T. Ooshio, M. Nakada, K. Sawamoto, M. Onodera, K. Matsumoto, M. Oshima, M. Asano, H. Saya, H. Okano, T. Suda, J. Hamada, A. Hirao, Identification of tumor-initiating cells in a highly aggressive brain tumor using promoter activity of nucleostemin, *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 17163-17168.

- [14] N. Uema, T. Ooshio, K. Harada, M. Naito, K. Naka, T. Hoshii, Y. Tadokoro, K. Ohta, M.A. Ali, M. Katano, T. Soga, Y. Nakanuma, A. Okuda, A. Hirao, Abundant nucleostemin expression supports the undifferentiated properties of germ cell tumors, *Am. J. Pathol.* 183 (2013) 592-603.
- [15] T. Lin, L. Meng, Y. Li, R.Y. Tsai, Tumor-initiating function of nucleostemin-enriched mammary tumor cells, *Cancer Res.* 70 (2010) 9444-9452.
- [16] Y. You, X.Q. Li, J.N. Zheng, Y.H. Wu, Y.L. He, W. Du, P. Zou, M. Zhang, Transcript level of nucleostemin in newly diagnosed acute myeloid leukemia patients, *Leukemia Res.* 37 (2013) 1636-1641.
- [17] J. Lessard, G. Sauvageau, Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells, *Nature* 423 (2003) 255-260.
- [18] T. Hoshii, Y. Tadokoro, K. Naka, T. Ooshio, T. Muraguchi, N. Sugiyama, T. Soga, K. Araki, K. Yamamura, A. Hirao, mTORC1 is essential for leukemia propagation but not stem cell self-renewal, *J. Clin. Invest.* 122 (2012) 2114-2129.
- [19] Cancer Genome Atlas Research, Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia, *N. Engl. J. Med.* 368 (2013) 2059-2074.
- [20] E. Kroon, J. Kros, U. Thorsteinsdottir, S. Baban, A.M. Buchberg, G. Sauvageau, Hoxa9 transforms primary bone marrow cells through specific collaboration with Meis1a but not Pbx1b, *EMBO J.* 17 (1998) 3714-3725.
- [21] N. Goardon, E. Marchi, A. Atzberger, L. Quek, A. Schuh, S. Soneji, P. Woll, A. Mead, K.A. Alford, R. Rout, S. Chaudhury, A. Gilkes, S. Knapper, K. Beldjord, S. Begum, S. Rose, N. Geddes, M. Griffiths, G. Standen, A. Sternberg, J.

- Cavenagh, H. Hunter, D. Bowen, S. Killick, L. Robinson, A. Price, E. Macintyre, P. Virgo, A. Burnett, C. Craddock, T. Enver, S.E. Jacobsen, C. Porcher, P. Vyas, Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia, *Cancer Cell* 19 (2011) 138-152.
- [22] F. Ishikawa, S. Yoshida, Y. Saito, A. Hijikata, H. Kitamura, S. Tanaka, R. Nakamura, T. Tanaka, H. Tomiyama, N. Saito, M. Fukata, T. Miyamoto, B. Lyons, K. Ohshima, N. Uchida, S. Taniguchi, O. Ohara, K. Akashi, M. Harada, L.D. Shultz, Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region, *Nat. Biotechnol.* 25 (2007) 1315-1321.
- [23] T. Kobayashi, K. Masutomi, K. Tamura, T. Moriya, T. Yamasaki, Y. Fujiwara, S. Takahashi, J. Yamamoto, H. Tsuda, Nucleostemin expression in invasive breast cancer, *BMC Cancer* 14 (2014) 215.
- [24] G. Zhang, Q. Zhang, Q. Zhang, L. Yin, S. Li, K. Cheng, Y. Zhang, H. Xu, W. Wu, Expression of nucleostemin, epidermal growth factor and epidermal growth factor receptor in human esophageal squamous cell carcinoma tissues, *J. Cancer Res. Clin. Oncol.* 136 (2010) 587-594.
- [25] S.J. Liu, Z.W. Cai, Y.J. Liu, M.Y. Dong, L.Q. Sun, G.F. Hu, Y.Y. Wei, W.D. Lao, Role of nucleostemin in growth regulation of gastric cancer, liver cancer and other malignancies, *World J. Gastroentero.* 10 (2004) 1246-1249.

Figure legends

Fig. 1. Characterization of fractionated AML cells by expression levels of NS-GFP.

(A) GFP intensity of leukaemia cells derived from NS-GFP transgenic mice. BM cells of recipient mice with AML were analysed by flow cytometry and divided into 4 subfractions according to GFP fluorescence intensity. (B) NS mRNA expression levels. Data are presented as the mean $\pm$ SD of the ratios of NS to β -actin mRNA copy numbers in fractionated cells ($n = 3$). $*p < 0.05$; $**p < 0.01$; $***p < 0.001$ (C) Expression of surface markers, including CD4, CD8, Mac-1, Gr-1, B220, Ter119, Sca-1 and c-Kit on fractionated cells.

Fig. 2. Enrichment of immature blastic leukaemia cells in the NS-high cell population.

(A) Representative data of NS-GFP and c-Kit expression on BM cells in leukaemic mice. AML cells were fractionated into 4 subpopulations: (a) c-Kit⁻NS-GFP^{low}, (b) c-Kit⁻NS-GFP^{middle}, (c) c-Kit⁺NS-GFP^{middle}, (d) c-Kit⁺NS-GFP^{high}. (B) Morphology of fractionated leukaemia cells. The fractionated AML cells were stained with May-Grünwald-Giemsa staining. (C) Representative data from forward-scattered (FSC) and side-scattered (SSC) light showing cell size of fractionated AML cells. (D) The gene expression profiling of fractionated AML cells. Unsupervised hierarchical clustering was performed on the normalized and log-transformed data of 17,419 probes, similarity was measured by Euclidean distance metric, and the average linkage was used to define the linking distance between clusters. Red and green reflect high and low expression levels, respectively, as indicated in the scale bar.

Fig. 3. Correlation of NS expression with leukaemia-initiating capacity. (A and B)

Cell cycle analysis with BrdU incorporation and DNA content. The leukaemic mice were sacrificed 2 hours after BrdU injection, followed by staining of fractionated AML cells, including (a) c-Kit⁻NS-GFP^{low}, (b) c-Kit⁻NS-GFP^{middle}, (c) c-Kit⁺NS-GFP^{middle}, and (d) c-Kit⁺NS-GFP^{high} with an anti-BrdU antibody and 7-AAD. **(A)** Representative data are shown. **(B)** The data are presented as the mean $\pm$ SD of the percentage of BrdU⁺ cells among all cells in the fraction (n = 3). *** p < 0.001 **(C)** Colony-forming capacity of fractionated leukaemia cells. The fractionated AML cells shown in Fig. 1A were cultured in methylcellulose media. The data shown are the mean $\pm$ SD of the number of colonies (n = 3). ** p < 0.01; *** p < 0.001 **(D)** Survival of recipient mice with fractionated AML cells. Ten leukaemia cells from each fraction shown in Fig. 1A were transplanted along with 5×10^5 bone marrow competitor cells (rescue cells) into separate lethally irradiated mice. Significant differences between NS-GFP⁺⁺⁺ and other fractionated AML cells were observed.

Fig. 4. Correlation of prognosis of AML patients with the leukaemia stem cell signature. **(A)** Clustering of 179 AML patients by a clustering algorithm (PAM) using the leukaemia stem cell gene signature (318 human genes). Clustering data (k = 2) with silhouette width are shown. Average silhouette width is 0.63. **(B and C)** Numbers of patients with FAB classification (B) or molecular risk classification (C) are shown. (N.C.) not classified. (N.D.) no data. **(D)** Overall survival of the clustered AML patients. Cluster 1 patients had a significantly higher survival rate than Cluster 2 patients.

Figure 1

Figure 2

Figure 3

Figure 4

Supplementary Methods

Preprocessing of microarray data

Leukaemia cells were sorted into 4 fractions: (a) c-Kit⁻NS-GFP^{low}, (b) c-Kit⁻NS-GFP^{middle}, (c) c-Kit⁺NS-GFP^{middle}, and (d) c-Kit⁺NS-GFP^{high}. Total RNA from 2×10^4 cells in each fraction was isolated by using TRIzol (Invitrogen). Microarray experiments were carried out using the RNA samples (One RNA sample per fraction). Two-cycle cDNA synthesis, in vitro transcription, and labelling were performed according to the methods in the GeneChip Expression Analysis Technical Manual (Affymetrix Inc., High Wycombe, UK), and the cDNA microarray analysis was performed with a GeneChip Mouse 430 2.0 (Affymetrix Inc., High Wycombe, UK). The complete microarray data set is available at the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/, accession number GSE 58032). Background correction and probe summarization were performed by using Microarray Suite version 5.0 (MAS 5.0), then 17,419 of 45,101 probes were extracted as "present" in all 4 samples. The data were normalized with R software. In global normalization, the trimmed mean target intensity of each array was arbitrarily set to 500. The expression data were log₂ transformed.

Unsupervised hierarchical clustering was performed on the normalized and log-transformed data of 17,419 probes to obtain the result shown in Fig. 2D. Similarity was measured by Euclidean distance metric and the average linkage was used to define the linking distance between clusters. To select genes associated with leukaemia stem cells, we first extracted 5,474 of 17,419 probes showing $|\log_2(\text{c-Kit}^+\text{NS-GFP}^{\text{high}}/\text{c-Kit}^-\text{NS-GFP}^{\text{low}})| \geq 1$, then further selected 3,494 probes that exhibited a gradual increase or decrease in expression that paralleled the progressive

changes among the 4 subpopulations by matching the following criteria (1+2+3 or 4+5+6): (1) $c\text{-Kit}^+\text{NS-GFP}^{\text{high}} \geq c\text{-Kit}^+\text{NS-GFP}^{\text{middle}}$, (2) $c\text{-Kit}^+\text{NS-GFP}^{\text{middle}} \geq c\text{-Kit}^-\text{NS-GFP}^{\text{middle}}$, (3) $c\text{-Kit}^-\text{NS-GFP}^{\text{middle}} \geq c\text{-Kit}^-\text{NS-GFP}^{\text{low}}$, (4) $c\text{-Kit}^+\text{NS-GFP}^{\text{high}} \leq c\text{-Kit}^+\text{NS-GFP}^{\text{middle}}$, (5) $c\text{-Kit}^+\text{NS-GFP}^{\text{middle}} \leq c\text{-Kit}^-\text{NS-GFP}^{\text{middle}}$, (6) $c\text{-Kit}^-\text{NS-GFP}^{\text{middle}} \leq c\text{-Kit}^-\text{NS-GFP}^{\text{low}}$. Among 3,494 probes, the 500 probes that exhibited the largest ratios between $c\text{-Kit}^-\text{NS-GFP}^{\text{low}}$ and $c\text{-Kit}^+\text{NS-GFP}^{\text{high}}$ were extracted. The 500 probes were converted to 382 mouse gene symbols by matching GRCm38.p2 of Ensembl database with the R ‘Bioconductor’ package, biomaRt (<http://www.bioconductor.org/packages/release/bioc/html/biomaRt.html>) and are shown as the ‘leukaemia stem cell signature’ in Supplementary Table 1.

Supplementary Fig. 1. Evaluation of human NS gene expression in human AML patients in the TCGA Research Network database. (A) Overall survival of patients with high and low NS expression. We used RNAseq gene-level RPKM (Reads per kilobase of exon per million mapped reads) data to evaluate the NS (GNL3) expression level in AML samples. Patients were divided into two groups based on NS levels: NS-High ($\geq$ median) and NS-Low ($<$ median). (B) NS expression level among AML samples classified by FAB. Horizontal bars indicate the median, and boxes represent the interquartile range (IQR). Upper whiskers represent the lower of either the maximum or (75th percentile + $IQR \times 1.5$). Lower whiskers represent the higher of either the minimum or (25th percentile - $IQR \times 1.5$). Circles represent outliers. M0, n = 16; M1, n = 44; M2, n = 40; M3, n = 16; M4, n = 35; M5, n = 21; M6, n = 2; M7, n = 31; not classified (N.C.), n = 2. There are no significant differences among the groups.

Supplementary Fig. 2. Correlation of prognosis of AML patients with the leukaemia stem cell signature. (A) Clustering of 179 AML patients by a clustering algorithm (PAM) using the leukaemia stem cell gene signature (318 human genes). Clustering data ($k = 3$) with average silhouette width are shown. Average silhouette width is 0.54. (B and C) Numbers of patients with FAB classification (B) or molecular risk classification (C) are shown. (N.C.) not classified. (N.D.) no data. (D) Overall survival of the clustered AML patients. Cluster 3 patients had a significantly higher survival rate than Cluster 2 patients.

Supplementary Table 1. Leukaemia stem cell signature.

M.mu	H.sa	M.mu	H.sa	M.mu	H.sa
Abca1	ABCA1	Cdc7	CDC7	Foxred2	FOXRED2
Abca13	ABCA13	Cdca2	CDCA2	Frmd4b	FRMD4B
Abca9	ABCA9	Cdca8	CDCA8	Fyb	FYB
Acss2	ACSS2	Ceacam1	CEACAM1	Gatm	GATM
Acvr1	ACVR1	Ceacam2	CEACAM1	Gbp3	GBP4
Add3	ADD3	Cenpe	CENPE	Gfi1	GFI1
Adrbk2	ADRBK2	Cenph	CENPH	Gnl3	GNL3
Agpat4	AGPAT4	Cenpi	CENPI	Gns	GNS
Ahnak	AHNAK	Cenpn	CENPN	Golm1	GOLM1
Ahr	AHR	Cenpp	CENPP	Gpnmb	GPNUMB
Akna	AKNA	Cenpq	CENPQ	Gpr137b	GPR137B
Ampd3	AMPD3	Cep55	CEP55	Gpr146	GPR146
Antxr2	ANTXR2	Chchd10	CHCHD10	Gpr18	GPR18
Anxa6	ANXA6	Chek1	CHEK1	Gsn	GSN
Apitd1	APITD1	Cks1b	CKS1B	Hip1	HIP1
Apobr	APOBR	Clec10a	CLEC10A	Hist1h2bq	HIST1H2BH
Apoc2	APOC2	Cmss1	CMSS1	Hist1h2br	HIST1H2BH
Arsb	ARSB	Crispld2	CRISPLD2	Hist1h4i	HIST4H4
Asns	ASNS	Csf1r	CSF1R	Hrsp12	HRSP12
Atp13a2	ATP13A2	Ctsb	CTSB	Idi1	IDI1
Atp1a3	ATP1A3	Ctse	CTSE	Ifi203	IFI16
Atp2b1	ATP2B1	Cxcl2	CXCL3	Ifitm3	IFITM3
Atp7a	ATP7A	Cxcr3	CXCR3	Igfbp4	IGFBP4
Aurka	AURKA	Cybb	CYBB	Ikbkb	IKBKB
Aurkb	AURKB	Cyp27a1	CYP27A1	Il17ra	IL17RA
Bach1	BACH1	Cyp4f18	CYP4F2	Il4ra	IL4R
Bank1	BANK1	Cyth4	CYTH4	Inpp5f	INPP5F
Bcas2	BCAS2	Cytip	CYTIP	Irf5	IRF5
Bcl3	BCL3	Ddi2	DDI2	Itfg3	ITFG3
Bcl6	BCL6	Ddit4	DDIT4	Itgb2	ITGB2
Birc5	BIRC5	Depdc1a	DEPDC1	Itgb5	ITGB5
Bmf	BMF	Dhfr	DHFR	Itgb7	ITGB7
Brca1	BRCA1	Dnajb14	DNAJB14	Junb	JUNB
Bub1	BUB1	Dopey2	DOPEY2	2810417H13Rik	KIAA0101
4632428N05Rik	C10orf54	Dpp4	DPP4	5031439G07Rik	KIAA0930
2310047M10Rik	C17orf59	Dtymk	DTYMK	C330027C09Rik	KIAA1524
BC055324	C1orf112	Dusp22	DUSP22	Kif18a	KIF18A
BC028528	C1orf54	Dusp3	DUSP3	Kif2c	KIF2C
3110043O21Rik	C9orf72	Dut	DUT	Kif4	KIF4A
6330416G13Rik	C9orf91	Ect2	ECT2	Klf3	KLF3
Camk1d	CAMK1D	Egln3	EGLN3	Kmo	KMO
Capg	CAPG	Elane	ELANE	L1cam	L1CAM
Capn2	CAPN2	Elk3	ELK3	Lgals3	LGALS3
Card6	CARD6	Emilin2	EMILIN2	Gm14548	LILRB3
Casp6	CASP6	Epsti1	EPSTI1	Gm15931	LILRB3
Cblb	CBLB	Errfi1	ERRFI1	Lilra6	LILRB3
Cbx5	CBX5	Evi2a	EVI2A	Pira2	LILRB3
Ccdc34	CCDC34	Exo1	EXO1	Pirb	LILRB3
Ccnd3	CCND3	Fabp5	FABP5	Gp49a	LILRB4
Ccne2	CCNE2	Fam129a	FAM129A	Lilrb4	LILRB4
Ccr2	CCR2	Fam214b	FAM214B	Lpin1	LPIN1
Cd14	CD14	Fam69a	FAM69A	Lpp	LPP
Cd244	CD244	Fbxl5	FBXL5	Lrp1	LRP1
Cd2ap	CD2AP	Fcgrt	FCGRT	Lrr1	LRR1
AF251705	CD300C	Fgr	FGR	Lsp1	LSP1
Cd300lf	CD300LF	Figl1	FIGL1	Map1lc3a	MAP1LC3A
Cd97	CD97	Filip1l	FILIP1L	Map4k4	MAP4K4
Cdc14b	CDC14B	Fnip2	FNIP2	Mapk1	MAPK1

M.mu	H.sa
Matn2	MATN2
Mef2a	MEF2A
Melk	MELK
Mfsd6	MFSD6
Mgam	MGAM
Mis18bp1	MIS18BP1
Mitf	MITF
Mmp8	MMP8
Mns1	MNS1
Mpo	MPO
Ms4a3	MS4A3
Mthfd1	MTHFD1
Mvp	MVP
Myadm	MYADM
Mybl2	MYBL2
Ncapg	NCAPG
Ncapg2	NCAPG2
Ncaph	NCAPH
Nckap1l	NCKAP1L
Ncl	NCL
Ncoa3	NCOA3
Ndc1	NDC1
Ndrp1	NDRG1
Nek2	NEK2
Nfam1	NFAM1
Nfil3	NFIL3
Nfkbiz	NFKBIZ
Nop58	NOP58
Npc2	NPC2
Nrip1	NRIP1
Nrp1	NRP1
Nsl1	NSL1
Nt5c3b	NT5C3B
Nudcd2	NUDCD2
Nudt5	NUDT5
Nup43	NUP43
Odc1	ODC1
Olfm1	OLFM1
Osbpl3	OSBPL3
P2rx7	P2RX7
Pag1	PAG1
Pbk	PBK
Peli1	PELI1
Pfas	PFAS
Pfkfb4	PFKFB4
Phgdh	PHGDH
Pid1	PID1
Pik3r5	PIK3R5
Pik3r6	PIK3R6
Pip4k2a	PIP4K2A
Pitpnm1	PITPNM1
Pld4	PLD4
Plec	PLEC
Plek	PLEK
Plekhg2	PLEKHG2
Plekhm3	PLEKHM3
Plk4	PLK4
Plxnb2	PLXNB2

M.mu	H.sa
Ppat	PPAT
Prex1	PREX1
Prim1	PRIM1
Prkag2	PRKAG2
Psmc3ip	PSMC3IP
Ptgrn	PTGFRN
Ptplad2	PTPLAD2
Ptpn1	PTPN1
Ptpro	PTPRO
Rab19	RAB19
Rab43	RAB43
5430435G22Rik	RAB7B
Rad51	RAD51
Rad51ap1	RAD51AP1
Rad54l	RAD54L
Rap1gap2	RAP1GAP2
Rbl1	RBL1
Rfc4	RFC4
Rgcc	RGCC
Rhoq	RHOQ
Rnase4	RNASE4
Rnf128	RNF128
Rrm1	RRM1
Rrm2	RRM2
S100a9	S100A9
Samd9l	SAMD9L
Samhd1	SAMHD1
Sema4a	SEMA4A
Sema4d	SEMA4D
Serf1	SERF1A
Serpib1a	SERPINB1
Sfxn5	SFXN5
Sgms1	SGMS1
Sgol2	SGOL2
Sh2b3	SH2B3
Sh3bp5	SH3BP5
Shcbp1	SHCBP1
Shisa5	SHISA5
Ska1	SKA1
Sla	SLA
Slc11a1	SLC11A1
Slc12a4	SLC12A4
Slc12a9	SLC12A9
Slc16a1	SLC16A1
Slc28a2	SLC28A2
Slc2a6	SLC2A6
Slc43a2	SLC43A2
Slc44a1	SLC44A1
Slc7a7	SLC7A7
Slfn8	SLFN13
Slfn9	SLFN13
Smc2	SMC2
Smim5	SMIM5
Snrpd1	SNRPD1
Spc24	SPC24
Spdl1	SPDL1
Sqstm1	SQSTM1
Ssh1	SSH1

M.mu	H.sa
Ssh2	SSH2
St6gal1	ST6GAL1
Stk10	STK10
Stx3	STX3
Suc1g2	SUCLG2
Syce2	SYCE2
Tbc1d1	TBC1D1
Tcf25	TCF25
Tep1	TEP1
Tfpi	TFPI
Tgfbr2	TGFBR2
Timm8a1	TIMM8A
Tk1	TK1
Tle1	TLE1
Tlr2	TLR2
Tmcc1	TMCC1
Tmem107	TMEM107
Tmem71	TMEM71
Tmem86a	TMEM86A
Tnfrsf1b	TNFRSF1B
Top2a	TOP2A
Trp53inp1	TP53INP1
Tpd52	TPD52
Tpx2	TPX2
Traip	TRAIIP
Trim37	TRIM37
Trip13	TRIP13
Trip6	TRIP6
Trpm2	TRPM2
Trpv2	TRPV2
Tspan17	TSPAN17
Ttk	TTK
Ttll3	TTLL3
Tube1	TUBE1
Tulp4	TULP4
Ube2h	UBE2H
Unc119	UNC119
Unc119b	UNC119B
Usp1	USP1
Utrn	UTRN
Wdr12	WDR12
Wee1	WEE1
Xdh	XDH
Ypel3	YPEL3
Zbtb4	ZBTB4
Zfhx3	ZFHX3
Zwilch	ZWILCH
Zyx	ZYX
Ccl6	N/A
Cd33	N/A
Clec4a3	N/A
Fabp5l2	N/A
Fnbp1l	N/A
Gm10693	N/A
Gm14085	N/A
Gm14328	N/A
Gm14928	N/A
Gm15163	N/A

M.mu	H.sa
Gm15922	N/A
Gm15925	N/A
Gm15930	N/A
Gm17709	N/A
Gm24428	N/A
Gm25122	N/A
Gm4070	N/A
Gm4879	N/A
Gm5385	N/A
Gm6100	N/A
Gm6166	N/A
Gm7901	N/A
Gm9115	N/A
Gm9252	N/A
Gm9797	N/A
Gvin1	N/A
Hmgn5	N/A
Ifitm6	N/A
Jhdm1d	N/A
Mlf1ip	N/A
Ms4a4b	N/A
Ms4a4c	N/A
Mt2	N/A
Mx1	N/A
Ppfia4	N/A
Pvr	N/A
Slfn2	N/A
Snhg5	N/A
Snord11	N/A
Snord89	N/A
Tlr13	N/A
Tpd52-ps	N/A
Wfdc17	N/A
Zfp658	N/A

The 382 mouse (M.mu) genes that showed the largest ratios between c-Kit^{NS}-GFP^{low} and c-Kit⁺NS-GFP^{high} and the 338 corresponding human (H.sa) gene symbols are shown.

Supplementary Table 2. Pathway analysis of the 382 mouse genes performed by DAVID v6.7 with GO and KEGG

< GO term : Biological Process >

Term	Count	p-value	Bonferroni	Benjamini	FDR
GO:0000279~M phase	28	4.69E-12	8.00E-09	8.00E-09	7.91E-09
GO:0022402~cell cycle process	32	1.65E-11	2.82E-08	1.41E-08	2.79E-08
GO:0007049~cell cycle	40	2.30E-11	3.93E-08	1.31E-08	3.89E-08
GO:0022403~cell cycle phase	29	2.67E-11	4.56E-08	1.14E-08	4.51E-08
GO:0051301~cell division	26	1.37E-10	2.34E-07	4.67E-08	2.31E-07
GO:0000280~nuclear division	20	3.57E-09	6.09E-06	1.02E-06	6.02E-06
GO:0007067~mitosis	20	3.57E-09	6.09E-06	1.02E-06	6.02E-06
GO:0000087~M phase of mitotic cell cycle	20	5.06E-09	8.63E-06	1.23E-06	8.53E-06
GO:0048285~organelle fission	20	6.53E-09	1.11E-05	1.39E-06	1.10E-05
GO:0000278~mitotic cell cycle	22	7.93E-09	1.35E-05	1.50E-06	1.34E-05
GO:0007059~chromosome segregation	10	3.09E-06	0.005258	5.27E-04	0.005212
GO:0006955~immune response	26	3.52E-06	0.005987	5.46E-04	0.005937
GO:0001775~cell activation	16	7.32E-05	0.117444	0.010357	0.123429
GO:0045321~leukocyte activation	15	7.66E-05	0.122500	0.010002	0.129103
GO:0009611~response to wounding	19	1.15E-04	0.177786	0.013885	0.193332
GO:0034654~nucleobase, nucleoside, nucleotide and nucleic acid biosynthetic process	13	1.59E-04	0.238365	0.017989	0.268817
GO:0034404~nucleobase, nucleoside and nucleotide biosynthetic process	13	1.59E-04	0.238365	0.017989	0.268817

< GO term : Molecular Function >

Term	Count	p-value	Bonferroni	Benjamini	FDR
GO:0017076~purine nucleotide binding	69	3.96E-10	1.67E-07	1.67E-07	5.57E-07
GO:0001882~nucleoside binding	61	6.75E-10	2.85E-07	1.43E-07	9.49E-07
GO:0032555~purine ribonucleotide binding	66	1.36E-09	5.73E-07	1.91E-07	1.91E-06
GO:0032553~ribonucleotide binding	66	1.36E-09	5.73E-07	1.91E-07	1.91E-06
GO:0030554~adenyl nucleotide binding	59	2.95E-09	1.24E-06	3.11E-07	4.14E-06
GO:0001883~purine nucleoside binding	59	4.04E-09	1.71E-06	3.41E-07	5.68E-06
GO:0032559~adenyl ribonucleotide binding	56	9.61E-09	4.06E-06	6.76E-07	1.35E-05
GO:0005524~ATP binding	55	1.72E-08	7.25E-06	1.04E-06	2.42E-05
GO:0000166~nucleotide binding	72	1.80E-08	7.59E-06	9.48E-07	2.53E-05
GO:0004674~protein serine/threonine kinase activity	21	3.66E-05	0.015344	0.001717	0.051494
GO:0019900~kinase binding	9	9.96E-05	0.041166	0.004195	0.139929
GO:0004672~protein kinase activity	24	1.64E-04	0.066901	0.006275	0.230383
GO:0019901~protein kinase binding	8	2.03E-04	0.082273	0.007129	0.285575

< GO term : Cellular Component >

Term	Count	p-value	Bonferroni	Benjamini	FDR
GO:0000775~chromosome, centromeric region	20	8.19E-14	1.99E-11	1.99E-11	1.06E-10
GO:0005694~chromosome	32	1.14E-12	2.77E-10	1.39E-10	1.48E-09
GO:0044427~chromosomal part	28	1.65E-11	4.00E-09	1.33E-09	2.13E-08
GO:0000793~condensed chromosome	17	6.75E-11	1.64E-08	4.10E-09	8.73E-08
GO:0000779~condensed chromosome, centromeric region	11	9.60E-08	2.33E-05	4.66E-06	0.000124
GO:0000776~kinetochore	11	1.33E-07	3.22E-05	5.37E-06	0.000172
GO:0000777~condensed chromosome kinetochore	10	3.62E-07	8.79E-05	1.26E-05	0.000468
GO:0043228~non-membrane-bounded organelle	61	6.29E-06	0.001527	0.000191	0.008132
GO:0043232~intracellular non-membrane-bounded	61	6.29E-06	0.001527	0.000191	0.008132
GO:0045121~membrane raft	8	0.000814	0.179536	0.021747	1.047653
GO:0000794~condensed nuclear chromosome	6	0.000875	0.191542	0.021038	1.125259

The significance of enrichment was evaluated by a modified Fisher's exact test. The pathways that showed significant enrichment (corrected $p < 0.05$) are shown. The Bonferroni correction was used to correct for multiple comparisons and to control the family-wise error rate. The false discovery rate (FDR) was controlled with the Benjamini-Hochberg procedure. Counts indicate the number of genes corresponding to individual terms.

Supplementary Table 3. Clustering of 179 AML patient samples by the k-medoids algorithm (PAM) using the human gene set.

A		Cluster 1 (n = 34)	Cluster 2 (n = 145)	<i>p</i> -value
Age (years old)	median (range)	50 (23-76)	60 (18-88)	0.0038
Sex	Male / Female	17 / 17	77 / 68	0.8491
WBC ($\times 10^3$ /ul)	median (range)	11.25 (0.4-98.8)	19.6 (0.6-297.4)	0.0495
FAB	M0	0	16	2.18E-13
	M1	5	39	
	M2	10	30	
	M3	16	0	
	M4	3	32	
	M5	0	21	
	M6	0	2	
	M7	0	3	
	N.C.	0	2	
Blast (%) in BM	mean (range)	71.3 (40.0-100.0)	68.5 (30.0-100.0)	0.4539
Molecular Risk	Good	24	10	2.50E-14
	Intermediate	5	88	
	Poor	4	45	
	N.D.	1	2	

B		Cluster 1 (n = 24)	Cluster 2 (n = 143)	Cluster 3 (n = 12)	<i>p</i> -value	post hoc <i>p</i> (corr) < 0.05
Age (years old)	median (range)	50 (21-76)	60 (18-88)	46 (29-68)	0.005	Cluster 2 - Cluster 1
Sex	Male / Female	13 / 11	76 / 67	5 / 7	0.753	
WBC ($\times 10^3$ /ul)	median (range)	25.0 (0.5-98.8)	19.6 (0.6-297.4)	1.6 (0.4-29.0)	0.034	Cluster 3 - Cluster 2
FAB	M0	0	16	0	4.24E-14	Cluster 2 - Cluster 1 Cluster 3 - Cluster 1 Cluster 3 - Cluster 2
	M1	6	38	0		
	M2	10	29	1		
	M3	5	0	11		
	M4	3	32	0		
	M5	0	21	0		
	M6	0	2	0		
	M7	0	3	0		
	N.C.	0	2	0		
Blast (%) in BM	mean (range)	67.75 (40.0-91.0)	68.35 (30.0-100.0)	79.58 (40.0-100.0)	0.145	
Molecular Risk	Good	13	10	11	2.51E-13	Cluster 2 - Cluster 1 Cluster 3 - Cluster 2
	Intermediate	7	86	0		
	Poor	1	2	0		
	N.D.	3	45	1		

Characterization of patients classified into two clusters ($k = 2$) (**A**) or three clusters ($k = 3$) (**B**) are shown. Differences in age of patients, number of WBC in peripheral blood, and percent of blast cells in BM at diagnosis were evaluated with Student's *t*-test (A), or a one-way ANOVA and Tukey post hoc test (B). Differences in the distribution of patients' sex, FAB classification, and molecular risk classification were evaluated by Fisher's exact test in (A) and (B). The Bonferroni test was used as a post hoc test in (B). Cluster pairs with significant differences are shown in the table. Molecular risk was classified as described previously [1,2,3].

- [1] K. Mrozek, G. Marcucci, D. Nicolet, K.S. Maharry, H. Becker, S.P. Whitman, K.H. Metzeler, S. Schwind, Y.Z. Wu, J. Kohlschmidt, M.J. Pettenati, N.A. Heerema, A.W. Block, S.R. Patil, M.R. Baer, J.E. Kolitz, J.O. Moore, A.J. Carroll, R.M. Stone, R.A. Larson, C.D. Bloomfield, Prognostic significance of the European LeukemiaNet standardized system for reporting cytogenetic and molecular alterations in adults with acute myeloid leukemia, *J. Clin. Oncol.* 30 (2012) 4515-4523.
- [2] H. Dohner, E.H. Estey, S. Amadori, F.R. Appelbaum, T. Buchner, A.K. Burnett, H. Dombret, P. Fenaux, D. Grimwade, R.A. Larson, F. Lo-Coco, T. Naoe, D. Niederwieser, G.J. Ossenkoppele, M.A. Sanz, J. Sierra, M.S. Tallman, B. Lowenberg, C.D. Bloomfield, L. European, Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet, *Blood* 115 (2010) 453-474.
- [3] J.P. Patel, M. Gonen, M.E. Figueroa, H. Fernandez, Z. Sun, J. Racevskis, P. Van Vlierberghe, I. Dalgalev, S. Thomas, O. Aminova, K. Huberman, J. Cheng, A. Viale, N.D. Socci, A. Heguy, A. Cherry, G. Vance, R.R. Higgins, R.P. Ketterling, R.E. Gallagher, M. Litzow, M.R. van den Brink, H.M. Lazarus, J.M. Rowe, S. Luger, A. Ferrando, E. Paietta, M.S. Tallman, A. Melnick, O. Abdel-Wahab, R.L. Levine, Prognostic relevance of integrated genetic profiling in acute myeloid leukemia, *N. Engl. J. Med.* 366 (2012) 1079-1089.