CEBPA–double-mutated acute myeloid leukemia displays a unique phenotypic profile: a reliable screening method and insight into biological features

Francesco Mannelli,[1,2] Vanessa Ponziani,[1,2] Sara Bencini,[1,2] Maria Ida Bonetti,[1,2] Matteo Benelli,[3]* Ilaria Cutini,[1,2] Giacomo Gianfaldoni,[1,2] Barbara Scappini,[1,2] Fabiana Pancani,[1,2] Matteo Piccini,[1,2] Tommaso Rondelli,[4] Roberto Caporale,[4] Anna Maria Grazia Gelli,[4] Benedetta Peruzzi,[4] Marco Chiarini,[5] Erika Borlenghi,[6] Orietta Spinelli,[7] Damiano Giupponi,[7] Pamela Zanghì,[7] Renato Bassan,[8] Alessandro Rambaldi,[7] Giuseppe Rossi[6] and Alberto Bosi[1,2]

[1]Unità Funzionale di Ematologia, Università degli Studi, AOU Careggi, Firenze; [2]Istituto Toscano Tumori, Firenze; [3]SOD Diagnostica Genetica, AOU Careggi, Firenze; [4]SOD Laboratorio Centrale, Settore Citometria Clinica, AOU Careggi, Firenze; [5]Centro di Ricerca Emato-Oncologica AIL (CREA), Spedali Civili, Brescia; [6]Divisione di Ematologia, Spedali Civili, Brescia; [7]Unità Strutturale Complessa di Ematologia, Ospedali Riuniti, Bergamo and [8]Divisione di Ematologia, Ospedale dell'Angelo & Ospedale SS. Giovanni e Paolo, Mestre-Venezia, Italy

** Current address: Centro di Biologia Integrativa, Università di Trento, Trento, Italy*

 doi:10.3324/haematol.2016.151910

Received: July 5, 2016.

Accepted: October 28, 2016.

Pre-published: November 10, 2016.

Correspondence: francesco.mannelli@unifi.it

Supplemental data

CEBPA–double mutated acute myeloid leukemia displays a unique phenotypic profile: a reliable screening method and insight into biological features

Francesco Mannelli[1,2], Vanessa Ponziani[1,2], Sara Bencini[1,2], Maria Ida Bonetti[1,2], Matteo Benelli[3,*], Ilaria Cutini[1,2], Giacomo Gianfaldoni[1,2], Barbara Scappini[1,2], Fabiana Pancani[1,2], Matteo Piccini[1,2], Tommaso Rondelli[4], Roberto Caporale[4], Anna Maria Grazia Gelli[4], Benedetta Peruzzi[4], Marco Chiarini[5], Erika Borlenghi[6], Orietta Spinelli[7], Damiano Giupponi[7], Pamela Zanghì[7], Renato Bassan[8], Alessandro Rambaldi[7], Giuseppe Rossi[6], and Alberto Bosi[1,2]

1) Unità Funzionale di Ematologia, Università degli Studi, AOU Careggi, Firenze, Italy;

2) Istituto Toscano Tumori, Firenze, Italy;

3) SOD Diagnostica Genetica, AOU Careggi, Firenze, Italy

4) SOD Laboratorio Centrale, Settore Citometria Clinica, AOU Careggi, Firenze, Italy;

5) Centro di Ricerca Emato-Oncologica AIL (CREA), Spedali Civili, Brescia, Italy;

6) Divisione di Ematologia, Spedali Civili, Brescia, Italy;

7) Unità Strutturale Complessa di Ematologia, Ospedali Riuniti, Bergamo, Italy;

8) Divisione di Ematologia, Ospedale dell'Angelo & Ospedale SS. Giovanni e Paolo, Mestre-Venezia, Italy

* Current address: Centro di Biologia Integrativa, Università di Trento, Trento, Italy

Corresponding author:

Francesco Mannelli, MD

UF di Ematologia, Dipartimento di Medicina Sperimentale e Clinica, Università di Firenze, and Istituto Toscano Tumori, Largo Brambilla 3, 50134 Firenze, Italy

e-mail: francesco.mannelli@unifi.it; Telephone and Fax: +39.055.7946742.

TABLE OF CONTENTS

Supplemental Materials and Methods

Treatment protocols

Protocol-1: since 2006 to March 2007, patients received induction according to standard ICE course (Cytarabine 100 mg/sqm bid on days 1-7; Idarubicin 12 mg/sqm on days 1-3; Etoposide 100 mg/sqm on days 1-5). High-dose Cytarabine (3000 mg/sqm bid days 1, 3, 5) was used as first consolidation in patients aged < 61 years attaining complete remission (CR) after ICE. Patients with persistent disease (*i.e.* > 5% BM blasts at hematopoietic recovery) after first course received a salvage regimen (FLA-Ida). In an intention-to-treat approach, patients aged < 55 years with high-risk karyotype, FLT3-ITD or adverse clinical features (secondary AML, CR after second course, hyperleukocytosis) were assigned to undergo allogeneic stem cell transplantation (SCT) from matched related or unrelated donor. Patients with intermediate cytogenetic risk in the absence of FLT3-ITD and adverse clinical features were allocated to allogeneic SCT if a related donor was available. Autologous SCT was offered to patients aged < 61 y with low-risk cytogenetics, intermediate-risk cytogenetics without sibling donor and high-risk disease not eligible to allogeneic SCT. Peripheral blood (PB) stem cells for autologous SCT were collected after a DIA course (Cytarabine 500 mg/sqm bid on days 1-6; Daunorubicin 50 mg/sqm on days 4-6). Patients who failed mobilization received two additional courses with high-dose cytarabine.

Protocol-2: since April 2007 to 2013, patients were enrolled in Northern Italy Leukemia Group (NILG) AML 02-06 protocol (Eudract code: 2006-003817-42). The protocol provided a randomization at induction between a standard ICE induction *versus* an experimental intensified one. Patients aged > 65 y were treated according to standard arm. Upon CR achievement, patients received standard doses cytarabine consolidation and were divided into standard and high risk cases (SR, HR): SR: favorable or intermediate risk cytogenetics (according to SWOG criteria) without any adverse clinical factor (secondary AML, FLT3-ITD, CR after cycle 2, persistence of pre-existing cytogenetic abnormality despite morphological CR; total WBC count >50 $x10^9$/L); HR: all non-SR cases. HR patients were assigned to undergo allogeneic SCT. Provided sufficient CD34+ cells were previously collected (>$2x10^6$/kg) upon recovery from high doses cytarabine, SR patients and HR patients excluded from allo-SCT and aged 65 years or less were randomized between autologous SCT and high doses consolidation therapy (R2). HR/SR patients unable to be randomized in R2 because of inadequate blood stem cell yield received intermediate-dose consolidation. Patients randomized to experimental arm were excluded from outcome analysis.

Karyotype:

Complex karyotype was defined by the presence of ≥ three cytogenetic aberrations according to European Leukemia Net (Döhner H *et al*, Blood. 2010;115:453).

Molecular Genetics:

NPM1 mutations: NPM1+ AML was defined by immunohistochemical criteria and/or by mutational analysis, as specified below. Total RNA was extracted from Ficoll–Hypaque isolated mononuclear cells. RNA was reverse-transcribed using random examers. NPM1 exon 12 was amplified as previously described with minor modification (Noguera *et al*, Leukemia 2005;19:1479–82). Briefly, 2 µl of cDNA were amplified in a total volume of 25 µl of the reaction mixture containing 1x PCR

buffer, 0.2mM dNTP, 2mM MgCl2, 1.25 U Hot Start Taq polymerase, and 10 pmol of each primer (HEX NPM-Rev6 5'- ACCATTTCCATGTCTGAGCACC-3', NPM-F25'-ATCAATTATGTGAAGAATTGCTTAC-3'). Pre-heating of the mixture at 94°C for 5 min was followed by 30 cycles of 30 s at 94°C, 45 s at 57°C, and 45 s at 72°C. A final extension of 10 min was carried out at 72°C on a Gene Amp PCR System 2400 (Perkin Elmer, Emeryville, CA, USA). NPM1 amplified product was separated with a capillary electrophoresis-based system (ABI PRISM 310 genetic analyzer, Applied Biosystems).

FLT3-ITD mutation: Approximately 100 ng of genomic DNA was amplified to produce a 328 bp fragment from wild-type alleles. Amplification was performed for 30 cycles (30 seconds at 95°C, 30 seconds at 60°C, 30 seconds at 72°C) followed by 40 minutes at 72°C, in a reaction mix containing 1x buffer, 2.0 mM MgCl2, 0.2mM dNTPs and 10 pmol each primer (11F: FAM 5'-GCAATTTAGGTATGAAAGCCAGC-3' and 12R: 5'-CTTTCAGCATTTTGACGGCAACC-3') and 1.25U Hot Start Taq polymerase in a total volume of 25 uL. FLT-3 amplified product was separated with a capillary electrophoresis-based system (ABI PRISM 310 genetic analyzer, Applied Biosystems).

CEBPA mutations: Mutations of CEBPA gene were detected by genomic DNA PCR and direct sequencing. The primer sets are those designed by Pabst *et al* (Nat Genet. 2001;27:263). Briefly, three overlapping primer pairs were used to amplify the entire coding region of human CEBPA: CEBPA AF-TCGCCATGCCGGGAGAACTCTAAC, CEBPA AR-AGCTGCTTGGCTTCATCCTCCT (548bp); CEBPA BF-CCGCTGGTGATCAAGCAGGA, CEBPA BR-CCGGTACTCGTTGCTGTTCT (390bp); CEBPA CF-CAAGGCCAAGAAGTCGGTGGACA, CEBPA CR-CACGGTCTGGGCAAGCCTCGAGAT (356bp). PCR reactions were run in a final volume of 50 µL containing genomic DNA (300 ng), KCl (50 mmol/L), Tris-HCl (20 mmol/L, pH 8.4), MgCl2 (2.5 mmol/L), 5 volume % DMSO, primers (2 mmol/L of each), nucleotides (0.1 mmol/L of each), and Taq DNA polymerase (1U). PCR conditions were: 94°C for 45 seconds, 62°C for 45 seconds and 72°C for 45 seconds for 45 cycles, with a final step for 10 minutes at 72°C. PCR products were sequenced using BigDye Terminator Cycle Sequencing Kit o v1.1 kit (Applied Biosystems) on ABI 3730 Genetic Analyzer (Applied Biosystems).

Flow cytometry:

2 x 10^6 cells of a fresh BM suspension were stained for surface markers using a stain-lyse-and-then-wash procedure; intracellular nuclear and cytoplasmic staining were performed after cell fixation and permeabilization, using Fix and Perm kit (Invitrogen, Carlsbad, CA, USA).

Data acquisition was performed using a FACSCalibur flow cytometer (BD Biosciences) and Cell Quest Pro software (BD) or FACSCanto II flow cytometer (BD) and FACSDiva software (BD). A total of 250,000 cells per tube was acquired.

FACSCalibur: Samples were acquired by FACSCalibur flow cytometer and Cell Quest Pro software (BD Biosciences). The following combinations of monoclonal antibodies (MoAb) in 4-color staining (fluorescein isothiocyanate/phycoerythrin/peridinin chlorophyll protein/allophycocyanin) were used:

HLA-DR/CD117/CD45/CD34; HLA-DR/CD123/CD45/CD34; CD11b/CD13/CD45/CD34; CD16/CD33/CD45/CD34; CD15/CD16/CD45/CD34; CD71/CD105/CD45/CD34; CD2/CD56/CD45/CD34; CD65/NG2/CD45/CD34; CD11b/CD64/CD45/CD34; CD36/CD64/CD45/CD34; CD14/CD64/CD45/CD34; nTdT/cyMPO/CD45/CD34; CD19/cyCD79a/CD45/CD34; cyCD3/CD7/CD45/CD34.

All MoAb were purchased from Becton Dickinson (BD,San Jose, CA, USA), excluding nTdT, cyMPO (Dako, Glostrup, Denmark), CD11b, CD65 (Immunotech, Marseille, France) and CD105 (e-Bioscience, San Diego, CA, USA).

Monoclonal antibodies' clones were: HLA-DR: G46-6; CD117: YB5.B8; CD45: HI30; CD34: 581; CD123: 7G3; CD11b: M1/70; CD13: WM15; CD22: S-HCL-1; CD33: HIM3-4; CD15: HI98; CD16: 3G8; CD71: M-A712; CD105: SN6; CD2: RPA-2.10; CD56: MY31; CD65: 88H7; NG2: 7.1; CD36: CB38; CD64: 10.1; CD14: M5E2; nTdT: HT-6; cyMPO: MPO-7; CD19: 4G7; cyCD79a: HM-57; cyCD3: HIT3a; CD7: M-T701.

Instrument setup, calibration and quality control were performed in order to assure measures' stability. Specifically, consistency of fluorescence intensity was monitored at least weekly by running fluorochrome-conjugated beads (Calibrite beads, BD). Fluorescence photomoltiplier voltages were adjusted until the mean channel values for the unlabeled beads corresponded to predetermined target values. Overtime stability of bead mean fluorescence intensity profile was checked by Levey-Jennings diagrams; changes of up to 15% of the mean target MFI were considered acceptable. The mixed-bead suspension was used to determine the appropriate compensation settings.

FACSCanto II: In order to verify the capability of the classifier to predict for CEBPA mutation status also with a different instrument, a subgroup of AML samples (n=94) from main cohort were acquired also by FACSCanto II flow cytometer (BD) and FACSDiva software (BD), along with 20 control bone marrows. Sample processing was the same as above.

The following combinations of monoclonal antibodies (MoAb) in 7-color staining (fluorescein isothiocyanate / phycoerythrin / peridinin chlorophyll protein 5.5 / phycoerythrin-Cy7 / allophycocyanin / allophycocyanin-H7 / pacific blue / pacific orange) were used:

CD15/CD64/CD34/CD117/CD7/CD45/HLA-DR
CD65/CD64/CD34/CD117/CD14/CD45/HLA-DR

MoAb were purchased from: CD15, CD64, CD34, CD14: Becton Dickinson, San Jose, CA, USA; HLA-DR: Biolegend, San Diego, CA, USA; CD45: Invitrogen, Waltham, MA, USA; CD7: eBioscience, San Diego, CA, USA; CD117: Beckman Coulter, Brea, CA, USA; CD65: Immunotech, Marseille, France

Monoclonal antibodies' clones were: HLA-DR: L243; CD117: 104D2D1; CD45: HI30; CD34: 8G12; CD15: HI98; CD65: 88H7; CD64: 10.1; CD14: M0P9; CD7: 124-1D1.

Analysis: For data analysis, Infinicyt (Cytognos SL, Salamanca, Spain) software was used. Some major BM cell compartments were identified on the basis of forward (FSC) and sideward (SSC)

light scatter and their reactivity for CD45 (Supplemental Data Figure 1). These subsets were: (i) blasts; (ii) maturing neutrophil compartment, selected on the basis of CD45+dim/CD34- with high SSC; (iii) monocytic compartment (CD45+hi/CD34- with intermediate SSC signal); (iv) mature erythroid compartment (CD45-/CD34- with low SSC). Within neutrophil compartment, 4 stages of maturation were identified depending on reactivity for CD13 and CD11b (Supplemental Data Figure S1B). Cell compartments were considered not assessable when not detectable as at least 0.01% of total BM cells.

Setup: Instrument setup, calibration and quality control were performed in order to assure measures stability. Specifically, consistency of fluorescence intensity was monitored at least weekly by running fluorochrome-conjugated beads (CS&T beads, BD). Fluorescence photomoltiplier voltages were adjusted until the mean channel values for the unlabeled beads corresponded to predetermined target values. Overtime stability of bead mean fluorescence intensity profile was checked by Levey-Jennings diagrams; changes of up to 15% of the mean target MFI were considered acceptable.

Statistical analysis:

Outcome: The definition of CR used established criteria (Cheson *et al*, JCO 2003;21:4642). Survival was calculated by Kaplan-Meier curves and the log rank test. Disease-free survival (DFS) was measured from CR date to relapse or last follow up; Overall survival (OS) from the diagnosis to the last observation or death. In order to rule out an impact by allogeneic SCT, we censored patients receiving allogeneic SCT at the date of transplant in a further analysis.

Ward hierarchical clustering: The data were stored in a database (Microsoft Excel; Microsoft, Seattle, WA, USA) and normalized. In order to normalize each patient data, we calculated the mean value of each parameter across all normal samples. Normalized data were obtained dividing the value of each individual by the calculated mean. Finally, all the data were $\log_2$-trasformed and the resulting $\log_2$ ratios were used for hierarchical clustering study.

Euclidean distance-based classifier: To predict CEBPA-dm status, we developed a specific euclidean distance based classifier on a selected group of phenotypic parameters. For each phenotypic parameter, we computed the mean of the normalized data across the 16 CEBPA-dm samples of the discovery dataset. We then defined the maximum euclidean distance between CEBPA-dm samples and this set of averages (named CEBPA-dm reference vector) as classification threshold. The samples in the validation dataset showing a distance between their normalized phenotypic data and the CEBPA-dm reference vector less than or equal to the threshold were classified as “highly probable” CEBPA-dm.

Supplemental Figures and Legends

Figure S1. Identification of cell compartments at AML diagnosis. Bivariate dot plots illustrates how the different cells compartments were identified on the basis of light scatter characteristics and reactivity for CD45 PerCP (A). Within neutrophil compartment, 4 stages of maturation were identified depending on reactivity for CD13 and CD11b (B). Dot plots were created by using Infinicyt software.

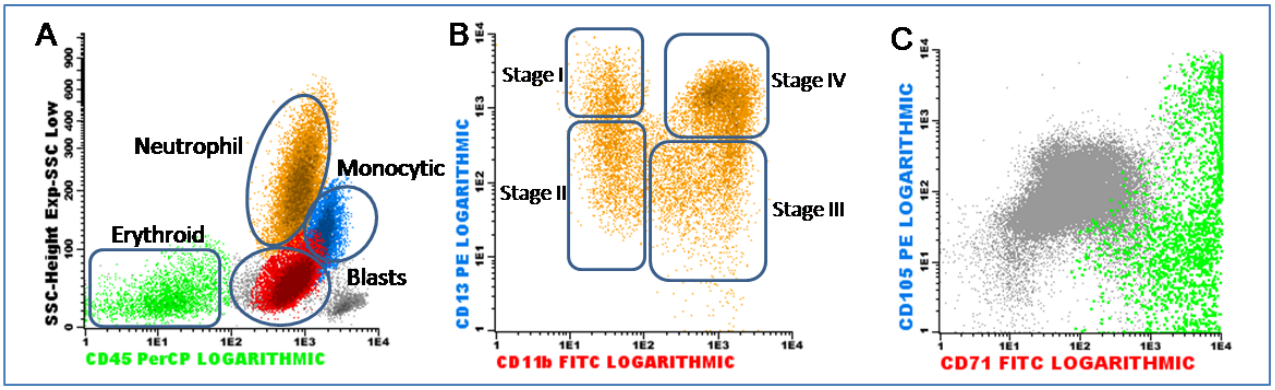
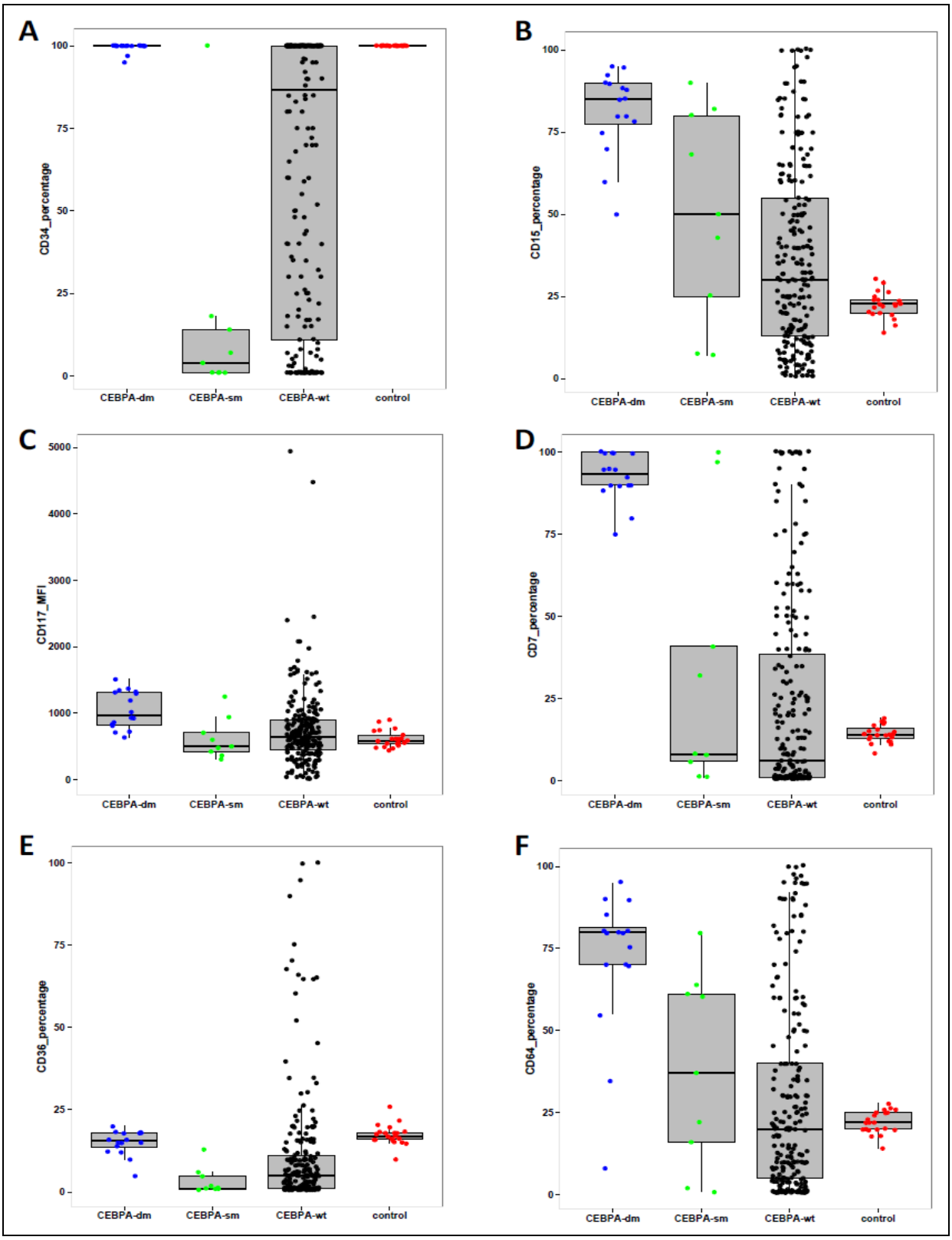
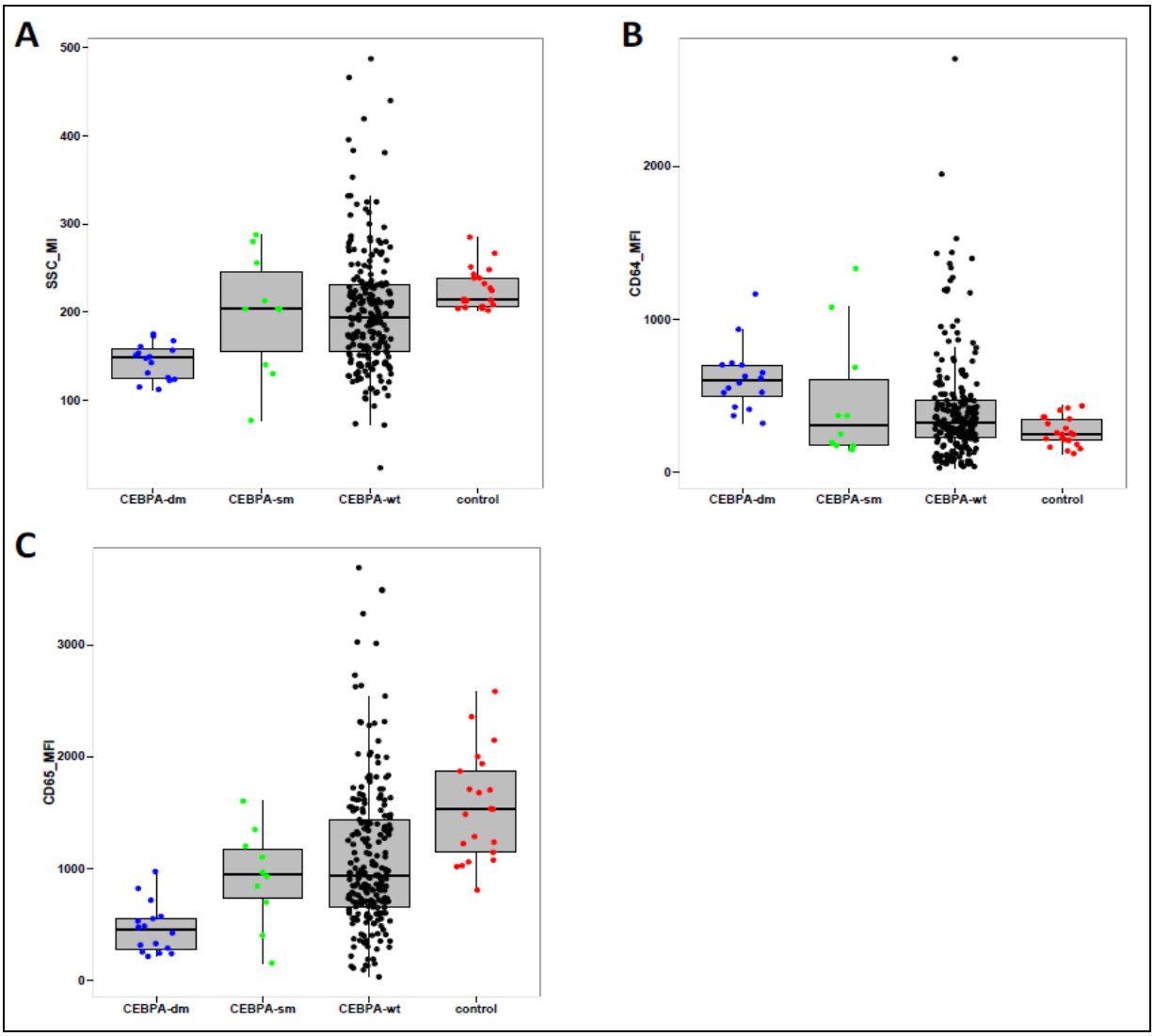


Figure S2. Phenotypic profile of blasts according to CEBPA status. Box plots illustrate the distribution of values in CEBPA-dm, -sm, -wt and controls for some core parameters: (A) CD34%; (B) CD15%; (C) CD117 MFI; (D) CD7%; (E) CD36%; (F) CD64%. Box plots were generated by R software. Boxes represent the interquartile range containing 50% of the cases; the horizontal line marks the median; dots are single cases.



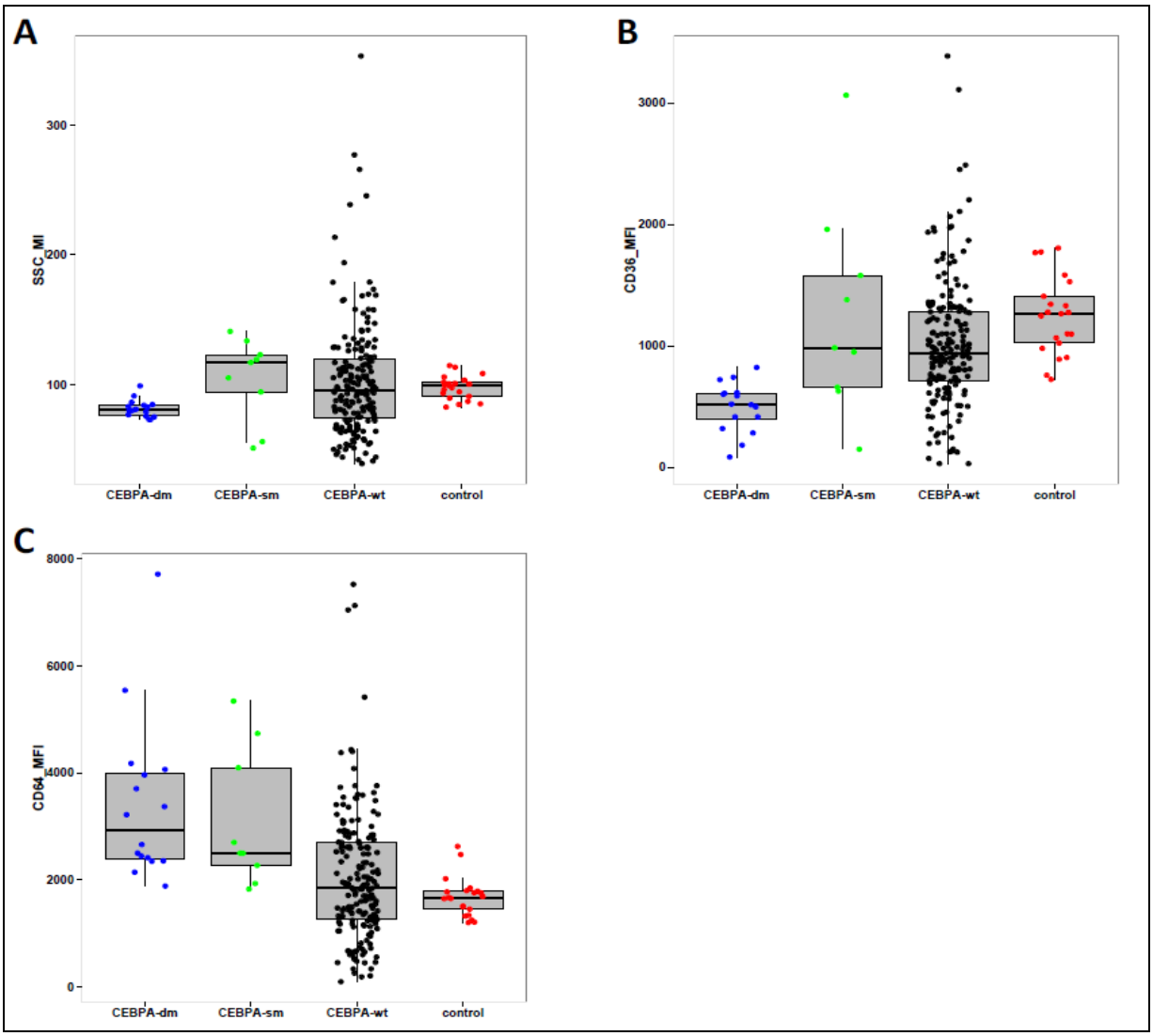
CEBPA-dm: double mutated; -sm: single mutated; -wt: wild type; MFI: Mean Fluorescence Intensity.

Figure S3. Phenotypic profile of neutrophil compartment according to CEBPA status. Box plots illustrate the distribution of values in CEBPA-dm, -sm, -wt and controls for some core parameters: (A) SSC MI; (B) CD64 MFI; (C) CD65 MFI. Box plots were generated by R software. Boxes represent the interquartile range containing 50% of the cases; the horizontal line marks the median; dots are single cases.



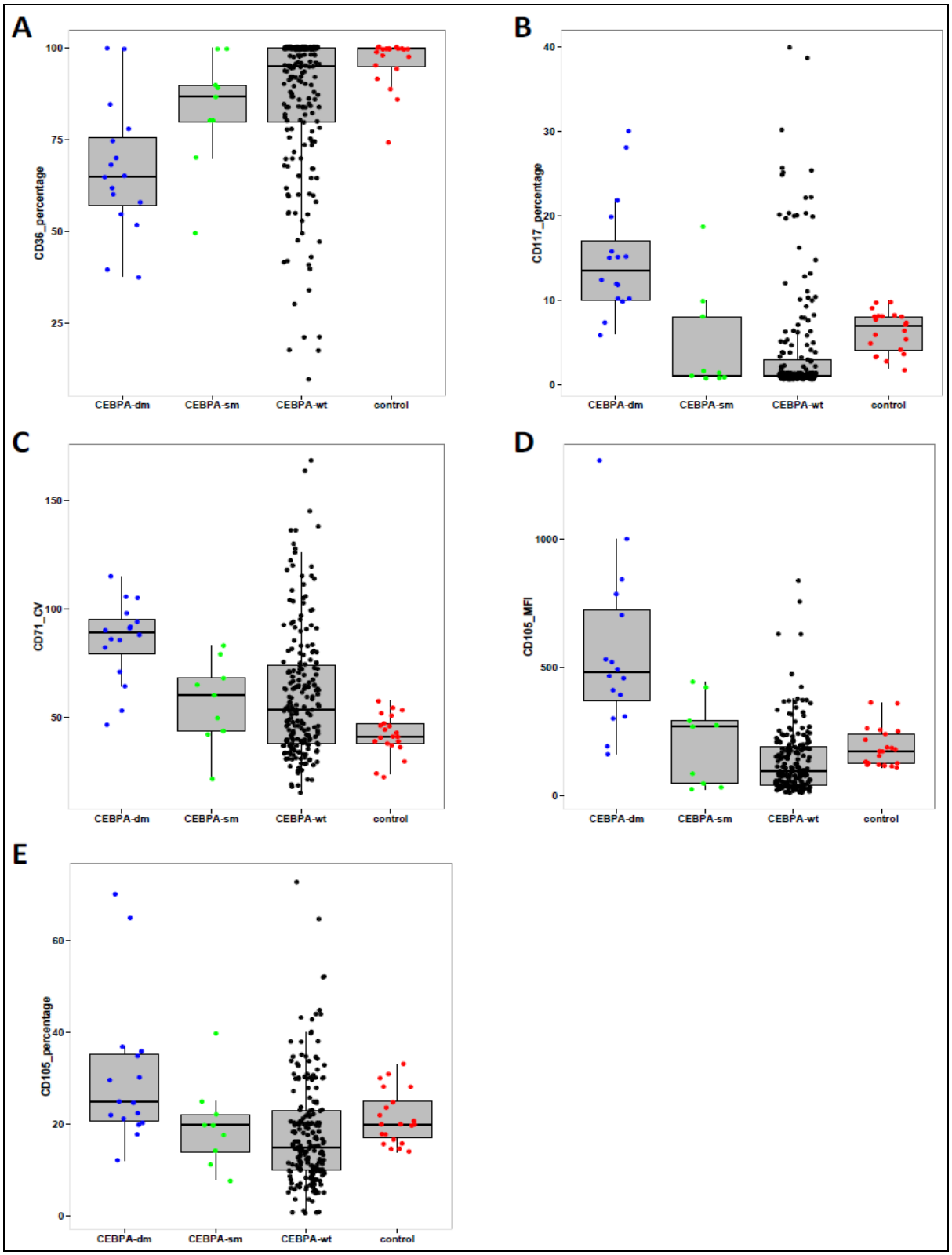
CEBPA-dm: double mutated; -sm: single mutated; -wt: wild type; SSC: Side Scatter; MI: Mean Intensity; MFI: Mean Fluorescence Intensity.

Figure S4. Phenotypic profile of monocytic compartment according to CEBPA status. Box plots illustrate the distribution of values in CEBPA-dm, -sm, -wt and controls for some core parameters: (A) SSC MI; (B) CD36 MFI; (C) CD64 MFI. Box plots were generated by R software. Boxes represent the interquartile range containing 50% of the cases; the horizontal line marks the median; dots are single cases.



CEBPA-dm: double mutated; -sm: single mutated; -wt: wild type; SSC: Side Scatter; MI: Mean Intensity; MFI: Mean Fluorescence Intensity.

Figure S5. Phenotypic profile of erythroid compartment according to CEBPA status. Box plots illustrate the distribution of values in CEBPA-dm, -sm, -wt and controls for some core parameters: (A) CD36%; (B) CD117%; (C) CD71 CV; (D) CD105 MFI; (E) CD105%. Box plots were generated by R software. Boxes represent the interquartile range containing 50% of the cases, the horizontal line marks the median; dots are single cases.



CEBPA-dm: double mutated; -sm: single mutated; -wt: wild type; CV: Coefficient of Variation; MFI: Mean Fluorescence Intensity.

Figure S6. CEBPA mutational analysis on sorted cell fractions in one CEBPA-single mutant patient. Cell compartments are shown on the left, with core phenotypic parameters for blasts (A), erythroid (B) and T lymphocytic (C) compartments. Neutrophil and monocytic cells were not sorted due to scarce cellularity. In the corresponding plots, ungated cells are in grey whereas the relevant cell population is highlighted by color: red for blasts, green for erythroid and orange for T-lymphocytes. The relative data from CEBPA mutational analysis are reported on the right, together with mutation type.

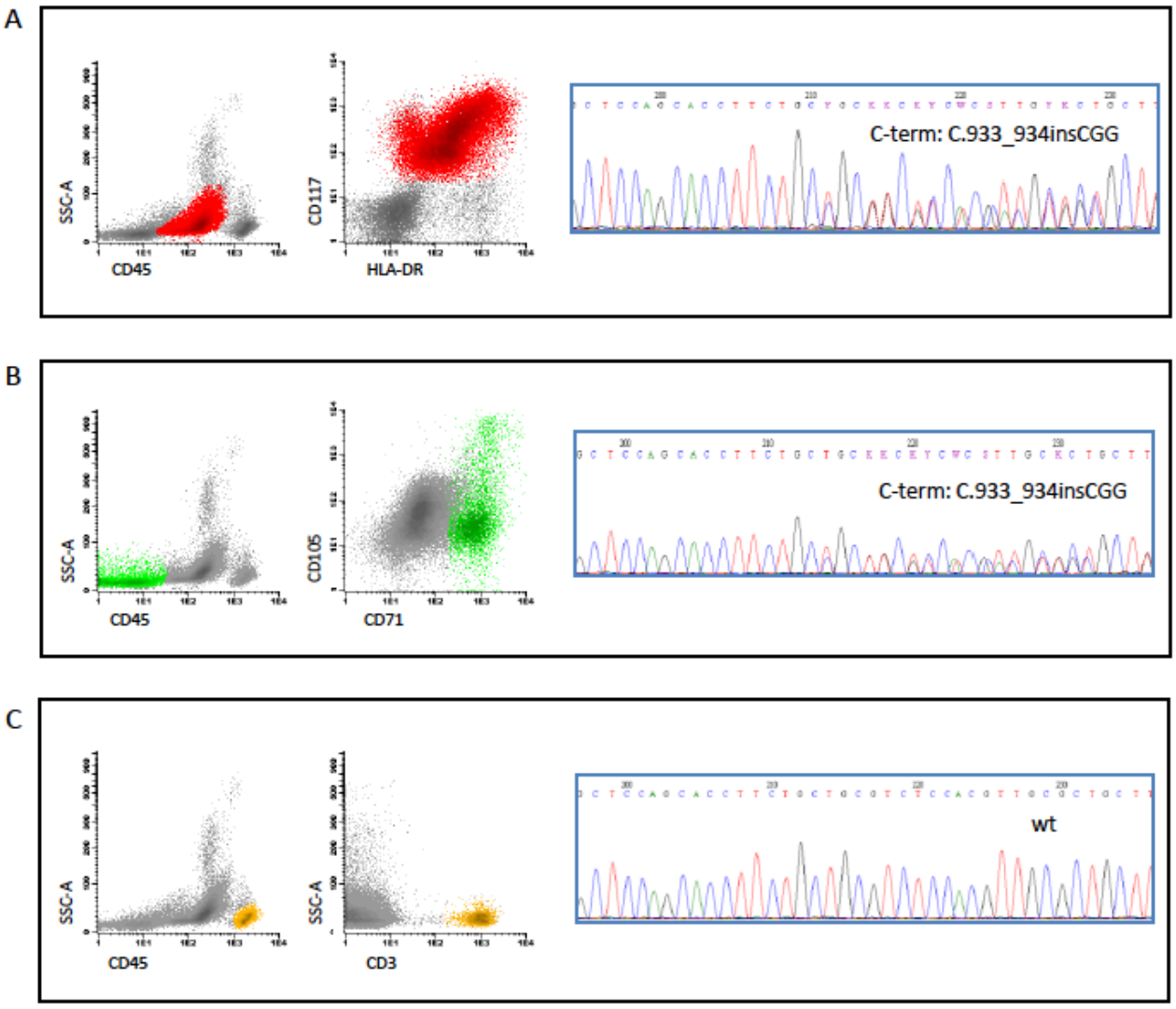


Figure S7. Cluster analysis according to genotypic groups within intermediate risk karyotype category. Cluster analysis of controls (n=21) and AML cases with intermediate risk karyotype (n=150) based on the phenotypic parameters of all bone marrow (BM) cell compartments at diagnosis. Columns represent individual BM samples; rows represent the normalized log_2 ratios of each parameter analyzed in a given cell compartment divided by the mean value obtained for that parameter in all control samples. The value of each parameter is represented in a color code according to control values: blue represents expression greater than the mean, red expression lower than the mean, white when not available; color intensity represents the magnitude of the deviation from the mean. Cluster analysis was carried out by R software.

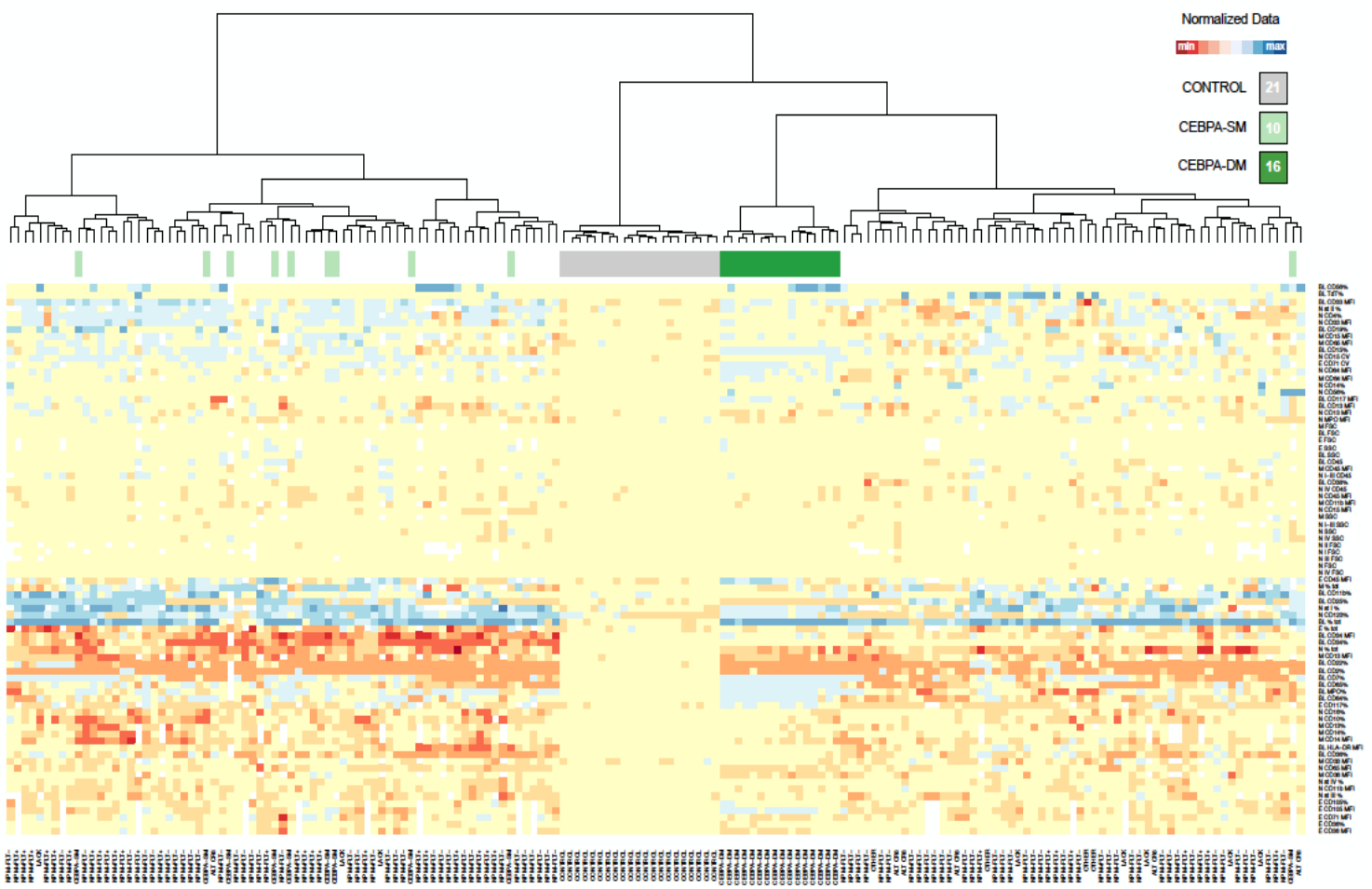


Figure S8. Cluster analysis according to genotypic groups within CD7-positive AML cases. Cluster analysis of controls (n=21) and AML cases (n=109) featured by positivity for CD7, as defined by EGIL (i.e. > 20% of total blasts), based on the phenotypic parameters of all bone marrow (BM) cell compartments at diagnosis. Columns represent individual BM samples; rows represent the normalized log_2 ratios of each parameter analyzed in a given cell compartment divided by the mean value obtained for that parameter in all control samples. The value of each parameter is represented in a color code according to control values: blue represents expression greater than the mean, red expression lower than the mean, white when not available; color intensity represents the magnitude of the deviation from the mean. Cluster analysis was carried out by R software.

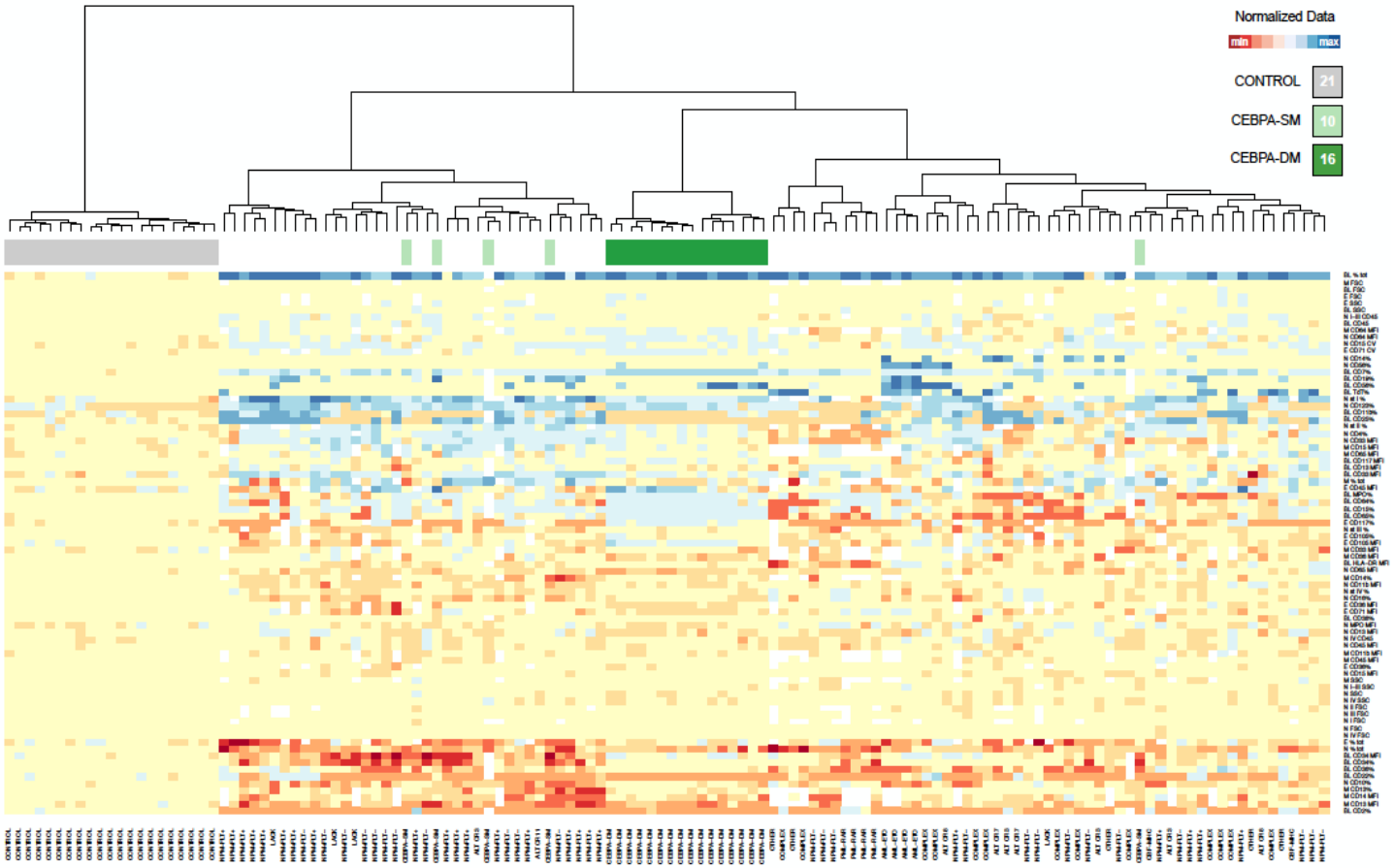


Figure S9. Outcome analysis according to genotypic groups within CD7-positive AML cases. An outcome analysis was carried out for 90 out of 109 CD7+ AML patients that were intensively treated. Kaplan Meier curves are stratified upon CEBPA status: overall CD7+ (black), CEBPA-wt/sm (red), CEBPA-dm (green) with p values representing the comparison between the latter two subgroups. (A) Disease-free survival (DFS); (B) Overall Survival (OS).

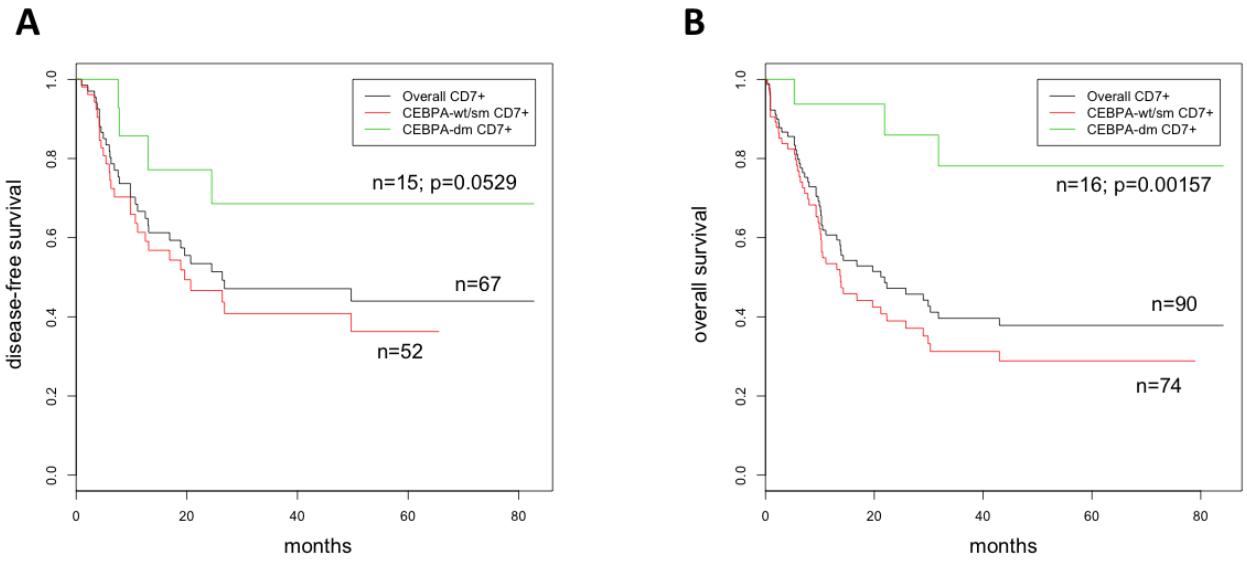


Figure S10. Principal component analysis of CEBPA-dm cases *versus* other genotypes. The multi-dimensional analysis of phenotypic profile of parameters included in the classifier was able to distinguish CEBPA-double mutant cases from other genotypic groups: AML bearing (A) AML1-ETO, (B) CBF-MYH11, (C) NPM1 mutations, (D) CEBPA-single mutant, (E) complex karyotype. Bi-plots are generated by the combination of first 2 principal components (PC), featured by the highest values of variance. Ellipses represent graphically the area of 95% confidence interval of the distribution for PCs. Samples outside ellipse are outliers. PCA analysis was carried out by R software.

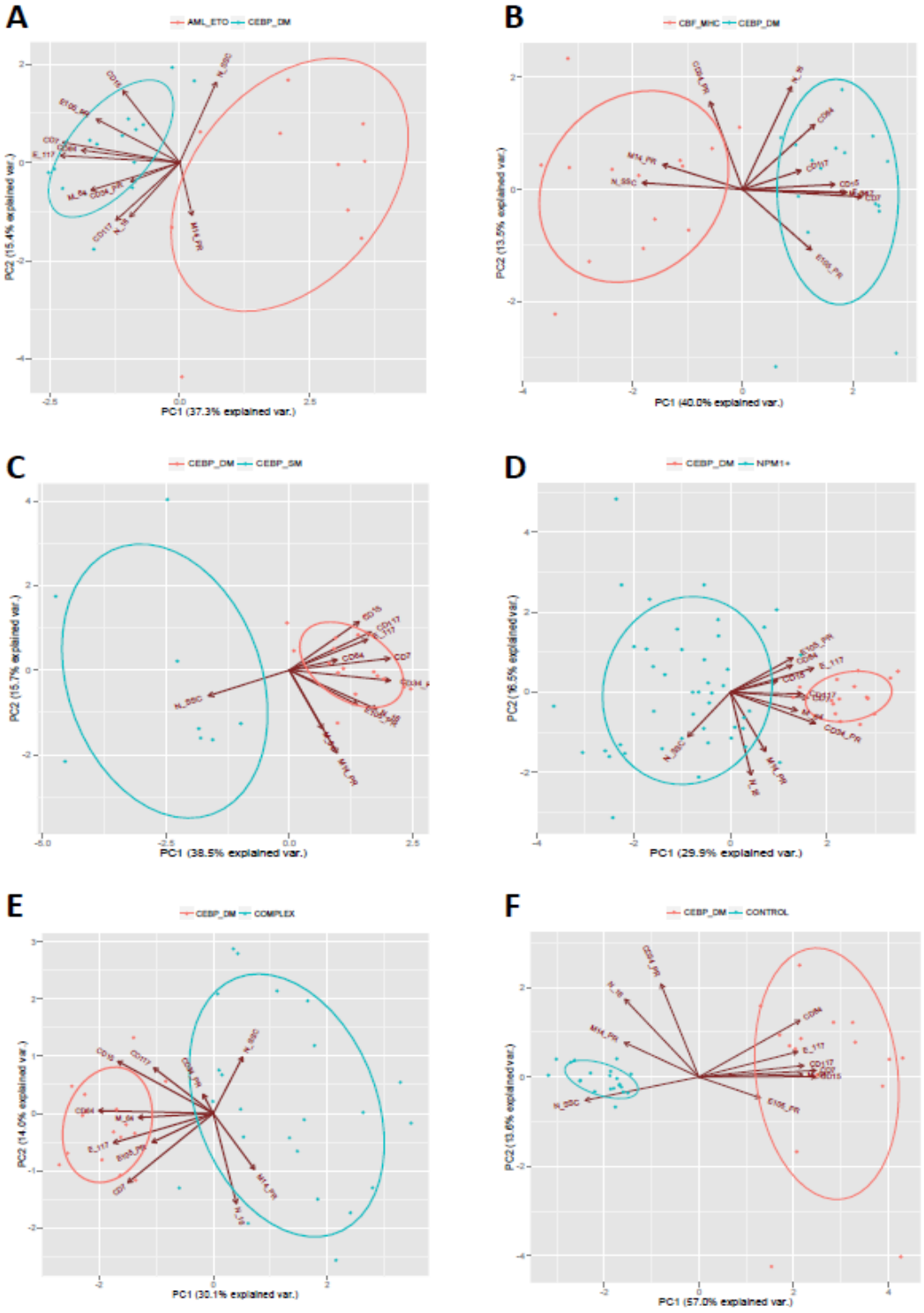


Figure S11. Cluster analysis according to genotypic groups based upon a selected group of phenotypic parameters. Cluster analysis of controls (n=21) and AML cases (n=251), based on the selected group of phenotypic parameters included in the classifier. Columns represent individual BM samples; rows represent the normalized log_2 ratios of each parameter analyzed in a given cell compartment divided by the mean value obtained for that parameter in all control samples. The value of each parameter is represented in a color code according to control values: blue represents expression greater than the mean, red expression lower than the mean, white when not available; color intensity represents the magnitude of the deviation from the mean. Cluster analysis was carried out by R software.

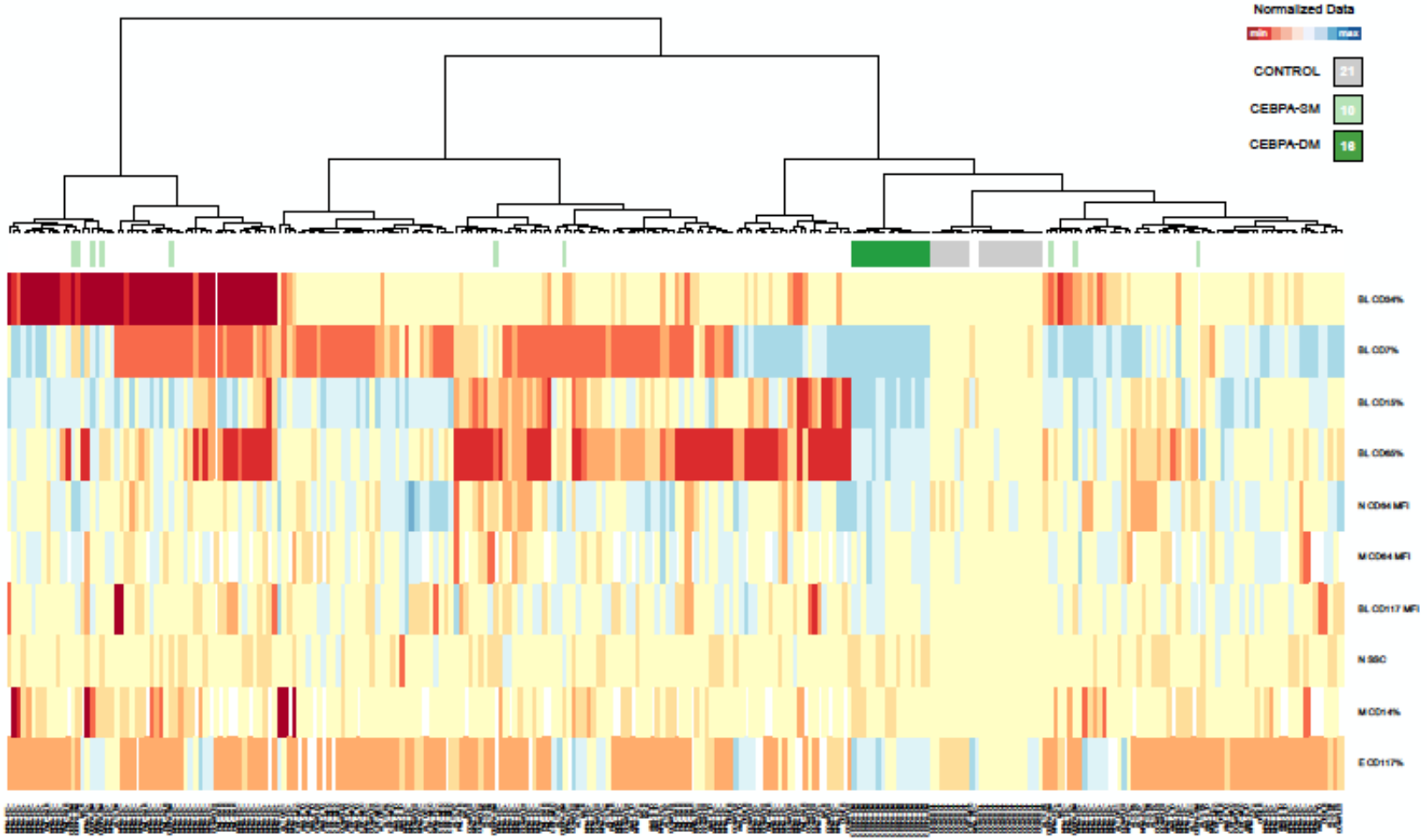


Figure S12. Cluster analysis according to genotypic groups based upon a selected group of phenotypic parameters. Cluster analysis of controls (n=20) and AML cases (n=94) acquired by a different flow cytometer (FACSCanto II), based on the selected group of phenotypic parameters included in the classifier. Columns represent individual BM samples; rows represent the normalized log_2 ratios of each parameter analyzed in a given cell compartment divided by the mean value obtained for that parameter in all control samples. The value of each parameter is represented in a color code according to control values: blue represents expression greater than the mean, red expression lower than the mean, white when not available; color intensity represents the magnitude of the deviation from the mean. Cluster analysis was carried out by R software.

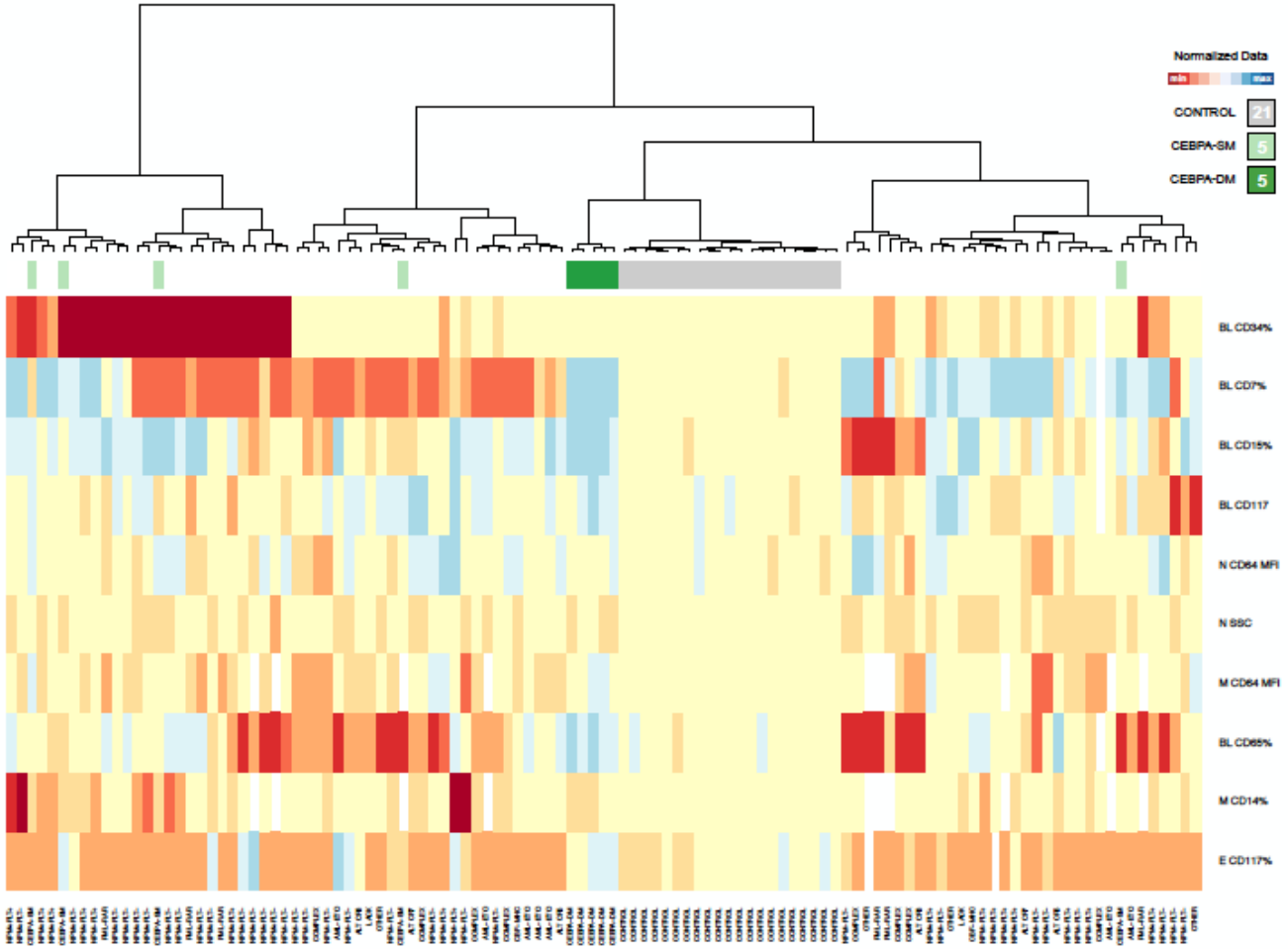
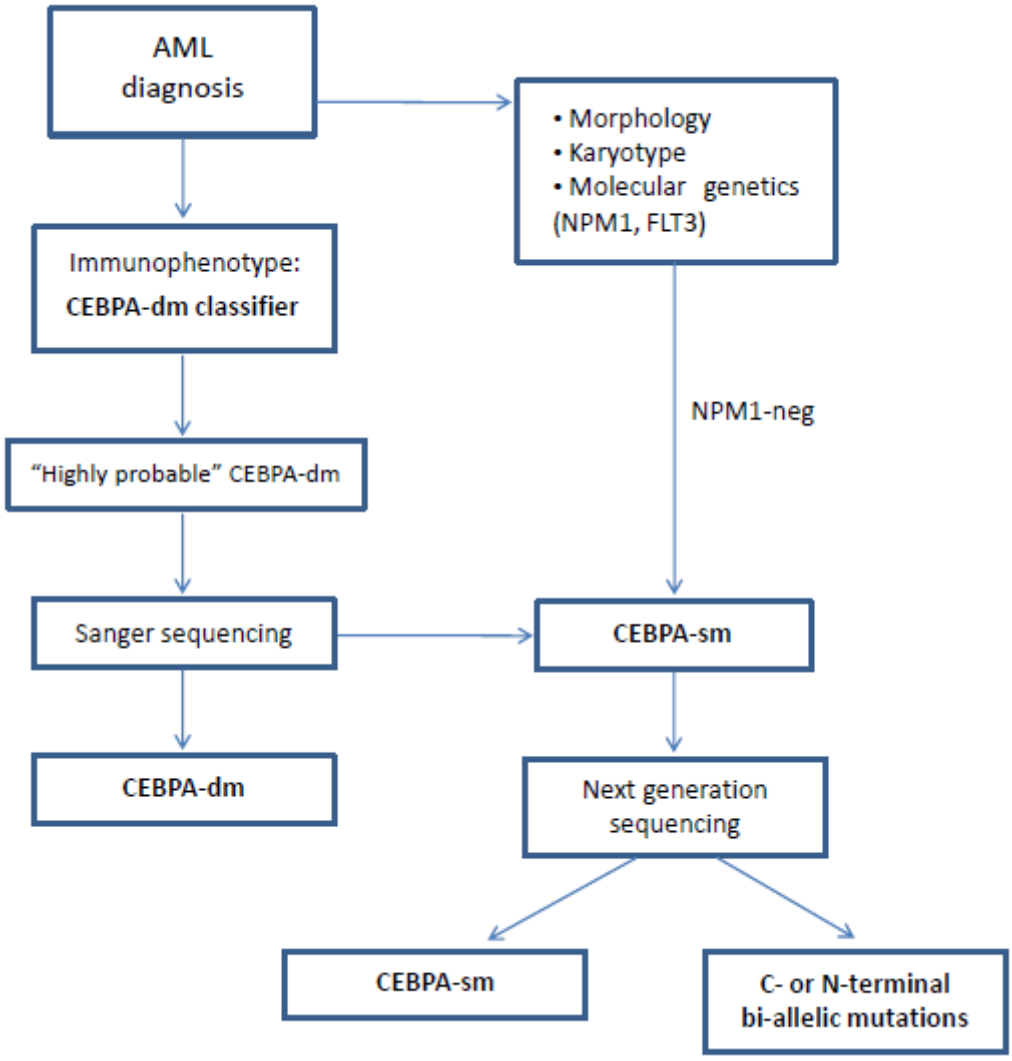


Figure S13. Proposal for integration of phenotype-based classifier in the diagnostic work-out of AML. The application of the classifier at diagnosis might provide a focused application of molecular techniques in order to reveal a CEBPA-dm status. In case of "highly probable" CEBPA-dm by phenotype and single mutant by Sanger sequencing, the application of next-generation sequencing could discriminate between an actual single mutant and bi-allelic mutations on the same part of the gene, sometimes troublesome by Sanger method.



Supplemental Tables and Results

Table SI. Phenotypic parameters in blast compartment according to CEBPA-status. Parameters are expressed as percentage of positive cells within cell compartment and/or its mean fluorescence intensity (MFI). Median values for phenotypic parameters are reported and compared by the Mann–Whitney test.

	Overall	CEBPA-wt	CEBPA-sm	CEBPA-dm	Controls	sm vs wt	sm vs crtl	sm vs dm	dm vs ctrl	dm vs wt
Blasts/CD34+ cells										
% tot	45.9 (0.1-97.7)	44.0 (0.1-97.7)	51.4 (14.7-84.5)	55.3 (23.4-74.3)	0.9 (0.2-1.8)	.355	<.0001	.889	<.0001	.14
FSC MI	461.5 (210.6-788.8)	460.5 (210.6-788.8)	458.7 (312.2-541.2)	527.5 (368.0-40.4)	417.5 (358.8-516.9)	.659	.283	.0315	.0002	.02
SSC MI	64.4 (25.9-225.6)	65.2 (25.9-225.6)	62.9 (35.2-87.0)	58.1 (50.2-76.0)	57.0 (47.1-69.4)	.739	.173	.257	.78	.0282
CD45 MFI	455.5 (133.7-1977.3)	450.4 (133.7-1977.3)	510.0 (212.7-1156.9)	482.1 (329.0-695.0)	475.4 (277.9-568.8)	.358	.094	.522	.55	.688
CD34 %	92 (0-100)	86.5 (0-100)	4 (0-100)	100 (95-100)	100 (100-100)	.0072	-	.0004	-	.0011
CD34 MFI	163.3 (1.5-2309.3)	230.1 (21.5-314.3)	16.0 (1.6-509.6)	429.5 (70.0-939.0)	256.3 (148.2-500.8)	.0011	.0004	.0002	.026	.0041
CD13 MFI	690.0 (15.4-4210.0)	783.5 (47.6-970.1)	578.9 (90.5-989.0)	543.3 (125.0-852.0)	572.8 (364.6-917.6)	.276	.929	.533	.137	.0359
CD33 MFI	734.2 (1.0-4692.0)	918.4 (71.2-1197.5)	1031.9 (102.1-3514.0)	757.0 (366.0-1480.0)	444.7 (66.6-718.5)	.368	.012	.301	.0003	.628
CD117 MFI	680.0 (11.4-4943.1)	643.6 (11.4-4943.1)	498.6 (304.0-1248.2)	979.2 (633.0-1508.7)	586.7 (438.0-899.0)	.448	.396	.0018	<.0001	.0002

HLA-DR MFI	1040.0 (14.1-6680.1)	992.8 (14.1-6680.1)	1050.8 (27.3-1962.3)	1186.5 (894.0-1924.0)	1591.0 (1001.9-2502.6)	.946	.007	.251	.0015	.171
CD7 %	10 (0-100)	6 (0-100)	8 (0-100)	93.5 (75-100)	14 (8-19)	.354	.684	.0078	<.0001	<.0001
MPO %	45 (0-100)	36.5 (0-100)	80 (0-100)	88 (45-100)	34 (26-50)	.614	.164	.212	<.0001	.0015
TdT %	0 (0-100)	0 (0-100)	0 (0-0)	0 (0-0)	0 (0-0)	-	-	-	-	-
CD56 %	0 (0-100)	0 (0-100)	0 (0-4)	5.5 (0-100)	0 (0-0)	.434	-	.0687	-	.053
CD15 %	35 (0-100)	30 (0-100)	50 (7-90)	86.5 (50-95)	23 (14-30)	.175	.032	.0127	<.0001	<.0001
CD64 %	20 (0-100)	20 (0-100)	37 (0-80)	77.5 (8-95)	22 (14-28)	.298	.331	.0055	<.0001	<.0001
CD65 %	12 (0-100)	10 (0-82)	16 (2-90)	62.5 (50-100)	21 (12-34)	.067	.625	.003	<.0001	<.0001
CD36 %	5 (0-100)	5 (0-100)	0 (0-13)	15.5 (5-20)	17 (10-26)	.125	<.0001	.0002	.088	.0001
CD11b%	2 (0-100)	3 (0-100)	0 (0-30)	0 (0-10)	2 (0-5)	.157	.927	.75	.326	.017
CD38 %	95 (0-100)	94 (0-100)	96 (40-100)	99 (95-100)	90 (84-98)	.261	.025	.118	<.0001	.0002
CD25 %	0 (0-100)	0 (0-100)	10 (0-90)	0 (0-0)	3 (0-6)	.99	.415	-	-	-
CD19 %	0 (0-100)	0 (0-100)	0 (0-100)	0 (0-0)	0 (0-0)	.737	-	-	-	-
CD22 %	0 (0-95)	0 (0-95)	0 (0-12)	0 (0-0)	8 (4-18)	.923	.004	-	-	-
CD2 %	0 (0-100)	0 (0-100)	0 (0-0)	0 (0-5)	9 (6-19)	-	-	-	<.0001	.24

Table SII. Phenotypic parameters in granulocytic compartment according to CEBPA-status. Parameters are expressed as percentage of positive cells within cell compartment and/or its mean fluorescence intensity (MFI). Median values for phenotypic parameters are reported and compared by the Mann–Whitney test.

	Overall	CEBPA-wt	CEBPA-sm	CEBPA-dm	Controls	sm vs wt	sm vs crtl	sm vs dm	dm vs ctrl	dm vs wt
Granulocytic compartment										
% tot	9.3 (0.03-71.8)	9.3 (0.03-71.8)	7.4 (1.4-45.3)	12.6 (1.1-39.8)	63.8 (50.3-77.8)	.652	<.0001	.421	<.0001	.699
FSC	586.9 (269.9-821.0)	590.8 (269.9-821.0)	563.7 (435.0-784.5)	576.4 (484.0-720.1)	660.8 (598.6-754.7)	.581	.0077	.420	.0025	.781
SSC	191.0 (23.7-687.7)	194.5 (23.7-566.0)	203.7 (77.5-687.7)	148.7 (112.6-175.5)	214.8 (202.1-285.1)	.750	.163	.0231	<.0001	<.0001
CD45	562.5 (109.5-1640.3)	576.3 (109.5-1640.3)	461.6 (273.6-1080.2)	509.5 (213.0-762.2)	720.8 (299.6-949.4)	.573	.0311	.897	.0004	.317
CD15 MFI	5900.0 (1222.8-9960)	6082.1 (1222.8-9960)	5495.4 (3155.6-9345)	5524.0 (4190-8235)	6371.5 (4455-8746)	.564	.059	.736	.052	.661
CD15 CV	43.5 (3.0-187.2)	42.0 (3.0-187.2)	49.1 (11.5-101.9)	42.7 (18.9-81.0)	27.9 (7.7-44.4)	.335	.0002	.219	.0025	.575
CD16 %	35 (0-100)	36 (0-100)	24.5 (0-50)	40.5 (20-80)	49 (36-76)	.048	<.0001	.048	.0042	.970
CD13	731.3 (81.5-3836.0)	783.0 (91.2-3836.0)	665.6 (172.6-1225.0)	451.5 (81.5-781.0)	854.3 (361.1-1385.6)	.125	.0389	.263	<.0001	.0006
CD11b	643.9 (41.6-4175)	668.0 (41.6-4175.0)	537.9 (129.7-945.4)	553.8 (396.0-686.5)	920.9 (744.5-1331.2)	.194	<.0001	.99	<.0001	.0652
CD65	915.0 (29.6-3689.6)	940.3 (29.6-3689.6)	945.8 (152.2-1603.3)	447.5 (213.1 (971.0)	1528.4 (804.9-2582)	.639	.0022	.0087	<.0001	<.0001
CD33	302.0 (14.0-3757.1)	298.0 (14.0-3757.1)	506.5 (125.7-1288.0)	316.6 (116.6-647.2)	256.3 (70.0-456.3)	.202	.0433	.310	.026	.695
CD64	328.0 (29.1-2702.6)	328.0 (29.1-2702.6)	310.2 (147.5-1331.9)	600.0 (320.4-1166.0)	248.4 (122.5-434.7)	.99	.416	.0484	<.0001	<.0001

MPO	2080.0 (121.0-8730.1)	2167.0 (121.0-8730.1)	1490.0 (481.0-6660.3)	1181.5 (912.0-2100)	2062.0 (1012.4-4655)	.068	.147	.856	.0067	.0003
CD10 %	20 (0-100)	20 (0-100)	17.5 (0-40)	35 (4-45)	50 (26-73)	.346	<.0001	.020	.0001	.202
CD123 %	15 (0-100)	15 (0-100)	45 (10-85)	23.5 (0-42)	0 (0-20)	.006	<.0001	.042	<.0001	.150
CD4 %	25 (0-100)	25 (0-100)	40 (5-75)	25 (10-33)	14 (8-27)	.137	.0005	.0038	.002	.766
CD56 %	0 (0-100)	0 (0-100)	0 (0-55)	0 (0-10)	0 (0-0)	.98	-	.97	-	.909
CD14 %	0 (0-70)	0 (0-70)	0 (0-0)	0 (0-0)	0 (0-0)	-	-	-	-	-
Stage I %	5.9 (0.01-93.7)	6.4 (0.01-93.7)	8.1 (1.5-33.0)	2.7 (1.3-11.4)	0.5 (0.1-2.1)	.582	<.0001	.027	<.0001	.073
Stage II %	18.5 (0.3-87.5)	18.7 (0.3-87.5)	27.3 (12.7-64.7)	16.3 (4.4-25.4)	10.4 (2.5-21.5)	.322	<.0001	.037	.010	.202
Stage III %	20.9 (0.3-74.4)	19.8 (0.3-74.4)	23.3 (2.6-34.0)	38.2 (12.4-55.7)	38.7 (14.0-43.3)	.59	.0001	.0008	.85	<.0001
Stage IV %	46.1 (2.4-100)	48.5 (2.4-100)	44.6 (13.8-75.3)	41.5 (28.5-76.3)	50.9 (40.2-80.9)	.583	.201	.938	.005	.382
Stage I FSC	606.9 (329.9-870.1)	600.7 (329.9-870.1)	604.3 (471.2-673.8)	630.4 (511.0-733.7)	682.8 (493.7-836.4)	.519	.003	.228	.027	.295
Stage II FSC	629.0 (273.3-907.7)	627.1 (273.3-907.7)	607.7 (455.9-781.1)	660.5 (536.0-660.5)	857.6 (632.4-958.0)	.419	<.0001	.207	<.0001	.389
Stage III FSC	640.1 (348.5-852.0)	643.5 (348.5-852.0)	618.0 (450.4-801.8)	623.8 (535.9-723.0)	713.6 (526.4-713.6)	.908	.0401	.933	.0003	.581
Stage IV FSC	203.0 (69.7-721.8)	558.0 (305.9-821.7)	526.1 (403.0-725.6)	575.1 (404.0-672.2)	648.9 (489.2-732.3)	.618	.0101	.391	.0014	.732
Stage I-III SSC	163.1 (44.0-350.5)	175.0 (44.0-350.2)	203.5 (67.4-280.0)	136.5 (104.7-170.0)	178.8 (152.6-239.6)	.804	.787	.097	<.0001	.006
Stage IV SSC	203.0 (69.7-721.8)	209.1 (69.7-521.0)	214.6 (81.6-721.8)	156.3 (127.7-201.3)	242.1 (207.0-320.3)	.883	.059	.0408	<.0001	.0003

Stage I-III CD45	442.3 (89.0-1080.0)	441.0 (89.1-1080.3)	434.6 (251.4-901.4)	471.5 (198.0-710.0)	366.0 (208.8-571.6)	.873	.053	.938	.0161	.557
Stage IV CD45	584.2 (110.1-2199.7)	580.0 (110.0-2199.7)	574.1 (214.0-1178.6)	613.7 (119.0-755.5)	653.5 (346.9-1197.0)	.851	.326	.979	.182	.680

Table SIII. Phenotypic parameters in monocytic compartment according to CEBPA-status. Parameters are expressed as percentage of positive cells within cell compartment and/or its mean fluorescence intensity (MFI). Median values for phenotypic parameters are reported and compared by the Mann–Whitney test.

	Overall	CEBPA-wt	CEBPA-sm	CEBPA-dm	Controls	sm vs wt	sm vs crtl	sm vs dm	dm vs ctrl	dm vs wt
Monocytic compartment										
% tot	5.6 (0.00-90.3)	5.9 (0.03-90.3)	11.2 (0.2-85.4)	2.6 (1.4-18.5)	4.6 (2.7-6.3)	.391	.163	.168	.048	.129
FSC	518.9 (338.0-803.7)	513.0 (338.0-803.7)	601.6 (388.0-636.2)	595.9 (424.0-679.9)	457.8 (420.2-507.1)	.149	.0017	.803	<.0001	.0024
SSC	94.0 (39.0-353.4)	95.8 (39.0-353.4)	117.0 (51.0-141.0)	80.5 (73.0-99.0)	99.0 (82.6-114.6)	.445	.125	.027	<.0001	.026
CD45	1680.2 (451.8-339.0)	1688.0 (451.8-3390.0)	1750.0 (618.0-2519.5)	1410.4 (1040.9-1905.0)	1856.7 (1182.3-2608.9)	.793	.964	.251	.0023	.067
CD14 %	75 (0-100)	76.5 (0-100)	52 (5-100)	68 (32-96)	90 (83-97)	.115	.0004	.223	.0005	.337
CD14 MFI	644.7 (8.0-3050.0)	636.6 (8.0-3050.0)	662.9 (41.7-1481.7)	666.8 (77.0-1427.0)	1352.5 (901.1-1957.1)	.911	.0019	.889	<.0001	.839
CD36	916.4 (30.4-3387.5)	940.0 (30.4-3387.5)	985.0 (149.6-3064.3)	518.3 (84.4-823.2)	1265.2 (724.4-1806.1)	.459	.625	.0028	<.0001	<.0001
CD64	1957.5 (96.0-7717.0)	1864.2 (96.0-7526.0)	2501.0 (1828.2-5343.0)	2939.6 (1885.0-7717.0)	1671.6 (1201.8-2627.1)	.0146	<.0001	.677	<.0001	<.0001
CD13 %	76.5 (0-100)	85 (0-100)	45 (15-90)	69 (30-90)	100 (95-100)	.0414	<.0001	.213	<.0001	.174
CD13 MFI	387.6 (14.1-4522.3)	478.2 (16.0-4522.3)	193.6 (14.1-746.2)	278.0 (65.0-1132.0)	1558.6 (805.7-3184.9)	.0276	<.0001	.487	<.0001	.0358
CD11b	1020.3 (168.0-6066.9)	1028.0 (168.0-6066.9)	970.0 (641.7-1691.7)	1080.2 (529.0-1612.0)	1450.5 (658.9-1841.0)	.298	.005	.419	.0034	.905
CD15	780.0 (30.0-6630.9)	757.1 (30.0-6630.9)	1543.0 (530.0-3030.0)	730.7 (279.0-1875.5)	586.1 (171.8-783.6)	.0156	<.0001	.008	.0175	.848

CD33	901.9 (50.0-3551.1)	800.0 (50.0-3551.1)	1216.8 (240.0-1393.3)	1264.5 (143.0-1794.7)	1125.5 (278.9-2218.9)	.333	.99	.328	.290	.0571
CD65	218.4 (15.0-2494.0)	221.0 (15.0-2494.0)	380.2 (27.0-668.0)	197.5 (158.0-930.0)	179.6 (51.8-732.6)	.473	.0401	.108	.250	.608

Table SIV. Phenotypic parameters in erythroid compartment according to CEBPA-status. Parameters are expressed as percentage of positive cells within cell compartment and/or its mean fluorescence intensity (MFI). Median values for phenotypic parameters are reported and compared by the Mann–Whitney test.

	Overall	CEBPA-wt	CEBPA-sm	CEBPA-dm	Controls	sm vs wt	sm vs crtl	sm vs dm	dm vs ctrl	dm vs wt
Erythroid compartment										
% tot	1.9 (0.00-55.9)	1.6 (0.00-55.9)	0.7 (0.00-25.8)	7.2 (1.9-19.1)	8.9 (1.8-19.9)	.447	.0078	.0268	.349	.0004
FSC	232.0 (146.0-468.5)	229.6 (146.0-468.5)	250.5 (162.0-351.9)	326.3 (191.0-462.2)	221.9 (209.4-253.7)	.130	.0356	.037	.0001	<.0001
SSC	22.8 (13.6-70.1)	22.3 (13.6-70.1)	24.8 (15.3-52.2)	26.6 (16.2-41.9)	19.8 (14.5-25.8)	.412	.230	.846	.0003	.0041
CD45	7.8 (1.0-553.5)	7.4 (1.0-382.0)	12.4 (2.2-553.5)	32.9 (1.0-208.1)	4.5 (2.2-13.3)	.287	.081	.301	.0064	.0038
CD117 %	0 (0-40)	0 (0-40)	0 (0-19)	13.5 (6-30)	7 (2-10)	.463	.089	.0022	<.0001	<.0001
CD36 %	94 (10-100)	95 (10-100)	87 (50-100)	65 (38-100)	100 (74-100)	.250	.013	.0314	<.0001	.0003
CD36 MFI	1342.0 (21.1-7794.0)	1478.5 (21.1-7794.0)	1243.1 (218.9-1977.0)	684.5 (509.6-2042.9)	1715.7 (1365.8-2123.7)	.498	.016	.151	.0012	.0045
CD105 %	16 (0-73)	15 (0-73)	20 (8-40)	25 (12-70)	20 (14-33)	.342	.458	.047	.0294	.0002
CD105 MFI	111.0 (12.6-1306.0)	97.6 (12.6-838.5)	270.0 (27.0-444.8)	480.5 (163.0-1306.0)	174.7 (110.0-364.2)	.211	.859	.0014	<.0001	<.0001
CD71	2710.0 (19.0-9143.4)	2951.4 (19.0-9143.4)	2236.0 (825.0-8263.3)	1332.5 (815.0-3240.0)	2596.4 (2186.1-4896.9)	.582	.226	.187	<.0001	.0028
CD71 CV	55.7 (15.1-168.5)	53.7 (15.1-168.5)	60.1 (21.5-83.0)	89.1 (46.5-115.0)	41.1 (22.4-57.4)	.829	.0246	.0014	<.0001	.0002

Table SV. TET2 and GATA2 gene mutation status in CEBPA-dm subset from the primary cohort.

ID	TET2				GATA2			
	status	**exon**	**sequence change**	**protein change**	**status**	**exon**	**sequence change**	**protein change**
11	wt				MUT	ex4	c.1396C>T	p.Leu321Phe (L321F)
63	wt				wt			
90	MUT	ex3.p4	c.1064G>A	p.Gly355Asp (G355D)	wt			
92	MUT	ex11.p3-4	c.5162T>G	p.Leu1721Trp (L1721W)	wt			
96	MUT	ex11.p4	c.5333A>G	p.His1778Arg (H1778R)	wt			
103	MUT	ex3.p10	c.2599T>C	p.Tyr867His (Y867H)	MUT	ex4	c.989G>A	p.Arg330Gly (R330Q)
109	wt				wt			
111	wt				wt			
148	MUT	ex11.p3-4	c.5162T>G	p.Leu1721Trp (L1721W)	wt			
222	wt				wt			
223	wt				wt			
228	MUT	ex11.p3-4	c.5162T>G	p.Leu1721Trp (L1721W)	wt			

ID refers to hierarchical cluster analysis (see Figure 5). All observed mutations were heterozygous. Mutation nomenclature is based on the reference sequence for TET2. GATA binding protein 2 (GATA2) transcript variant 1 (NM_001145661). Exon numbering is based also on NCBI reference transcript

Table SVI. Validation of phenotype-based classifier for CEBPA-dm cases on an independent cohort of AML cases. Results of the prospective application of euclidean distance based classifier. From the analysis of retrospective dataset, we selected a group of phenotypic parameters and developed a classifier. A classification threshold (3.26) was defined from the maximum distance between CEBPA-DM samples and a set of average values. The table reports the results of the application of the classifier on an independent cohort of 259 AML cases; the score represents the calculated distance for each case: the lower the score, the lower the distance from CEBPA-dm cases, the higher the probability of a CEBPA-dm status. Karyotype and molecular genetics is summarized for each case. ID is provided for CEBPA mutant cases (see Table SVII).

ID	**Score**	**Genetics**					**ID**	**Score**	**Genetics**			
		Karyotype	*CEBPA*	*NPM1*	*FLT3*				*Karyotype*	*CEBPA*	*NPM1*	*FLT3*
5	0.60	46, XY	DM	wt	wt			19.69	control	-	-	-
1	0.69	46, XX	DM	wt	wt			19.75	control	-	-	-
7	0.97	47, XX, +10	DM	wt	ITD			19.76	control	-	-	-
4	1.03	46, XX	DM	wt	wt			20.30	control	-	-	-
12	1.25	46, XY	DM	wt	wt			20.45	control	-	-	-
3	1.508	46, XY	DM	wt	wt			20.86	46,XY	wt	MUT	wt
11	1.509	46, XY	DM	wt	wt			21.00	complex	wt	wt	wt
6	2.31	46, XY	DM	wt	wt			22.00	46, XX, inv(16)	wt	wt	ITD
8	2.34	46, XX	DM	wt	wt			22.22	control	-	-	-
9	2.52	46, XY	DM	wt	wt			22.27	control	-	-	-
10	2.71	lack of growth	DM	wt	wt			24.19	control	-	-	-
2	3.15	46, XY	DM	wt	wt			24.41	control	-	-	-
	12.98	complex	wt	wt	wt			24.93	control	-	-	-
	13.13	control	-	-	-			25.56	complex	wt	wt	wt
	13.48	control	-	-	-			26.36	46,XX, t(8;21;21)	wt	wt	wt
18	15.14	46, XY	SM	wt	wt			26.41	47, XY, +8	wt	wt	wt
	17.09	control	-	-	-			26.94	control	-	-	-
	17.26	control	-	-	-			27.63	control	-	-	-
	18.09	control	-	-	-			27.82	46, XX	wt	wt	wt
	18.42	control	-	-	-			27.93	46;XY, t(2;22)	wt	wt	wt
	18.50	control	-	-	-			28.88	47,XX, t(16;16), +10	wt	wt	wt
	18.75	control	-	-	-			29.03	46, XX	wt	wt	wt
	18.91	46, XX	wt	MUT	wt			31.08	46, XX	wt	MUT	wt

ID	Score	Genetics					ID	Score	Genetics			
		Karyotype	*CEBPA*	*NPM1*	*FLT3*				*Karyotype*	*CEBPA*	*NPM1*	*FLT3*
	31.44	47, XY, +13	wt	wt	wt			49.35	46, XY	wt	wt	wt
	32.80	46, XY, t(3;11)(q?21;p?15)	wt	wt	wt			49.82	46, XY	wt	MUT	wt
	33.76	46, XY, t(7;12)	wt	MUT	wt			49.90	46, XY	wt	wt	wt
	34.48	46, XX	wt	MUT	wt			50.13	46, XY	wt	MUT	ITD
	35.71	46, XY	wt	wt	D835Y			50.93	46, XY	wt	MUT	wt
	38.76	complex	wt	wt	wt			51.25	46, XX	wt	wt	wt
	39.03	lack of growth	wt	MUT	ITD			52.53	complex	wt	wt	wt
	39.41	complex	wt	wt	wt			52.86	complex	wt	wt	wt
	39.54	46, XY, del(9q)	wt	MUT	ITD			53.33	46, XY	wt	MUT	wt
	40.31	46, XY	wt	MUT	D835Y			53.64	46, XY, del(20q)	wt	wt	wt
14	40.62	46, XX	SM	wt	ITD			53.67	46, XX	wt	wt	wt
	40.70	46, XX	wt	MUT	ITD			54.88	46, XY	wt	MUT	wt
	40.84	46, XY	wt	MUT	ITD			55.03	complex	wt	wt	wt
	41.17	46, XY, i(16)(q11.1)	wt	wt	ITD			55.34	46, XX, inv(16)	wt	wt	wt
	42.20	46, XY	wt	MUT	wt			56.47	complex	wt	wt	wt
	42.82	lack of growth	wt	wt	wt			56.49	46, XX, inv(16)	wt	wt	D835Y
	43.39	46, XY	wt	wt	wt			56.88	46, XY	wt	wt	wt
	43.55	46, XY, +8	wt	wt	wt			57.66	46, XY, t(8;21)	wt	wt	wt
	43.71	46, XX	wt	wt	wt		16	58.54	46, XY	SM	MUT	wt
	45.00	46, XX	wt	MUT	wt			58.73	46, XX	wt	MUT	ITD
	45.02	46, XY,del(9q)	wt	MUT	wt			58.93	46, XY, t(8;21)	wt	wt	wt
	45.44	46, XY	wt	wt	wt			60.64	46, XX, inv(16)	wt	wt	wt
	47.19	46, XY	wt	MUT	ITD			60.74	46, XY, inv(16)	wt	wt	wt
	47.20	46,XX,add(11)(p?),add(14)(p11)	wt	wt	wt			61.00	46, XX	wt	wt	wt
	48.41	47, XY, +8	wt	wt	wt			62.13	45, XY, -7	wt	wt	wt
	49.03	complex	wt	wt	wt			62.84	46, XY, t(8;21)	wt	wt	wt
	49.13	46, XX	wt	MUT	wt			62.96	46, XY	wt	wt	wt

ID	Score	Genetics					ID	Score	Genetics			
		Karyotype	*CEBPA*	*NPM1*	*FLT3*				*Karyotype*	*CEBPA*	*NPM1*	*FLT3*
	63.36	46, XY	wt	wt	wt			75.76	46, XY	wt	MUT	ITD
	64.72	46, XY, t(8;21)	wt	wt	wt			76.04	46, XY	wt	wt	wt
	64.83	46, XX	wt	MUT	ITD			76.48	46, XX	wt	wt	wt
	66.76	46, XY, t(6;9)	wt	wt	wt			77.68	46, XY	wt	MUT	wt
	67.27	46, XX, t(15;17)	wt	wt	ITD			78.85	46, XX, inv(16)	wt	wt	wt
	67.30	45, X, -Y	wt	MUT	D835Y			79.14	46, XY	wt	MUT	ITD
	68.56	complex	wt	wt	wt			79.41	46, XY, t(6;9)	wt	wt	wt
	69.13	complex	wt	wt	wt			79.65	46, XY	wt	MUT	wt
	69.90	46, XX	wt	wt	wt			80.54	46, XX, t(4;14)	wt	MUT	ITD
	70.53	46, XY	wt	wt	wt			80.85	46, XX, t(15;17)	wt	wt	ITD
	70.56	complex	wt	wt	wt			80.88	46, XX	wt	wt	wt
	70.57	46, XY	wt	wt	wt			81.15	46, XX	wt	wt	ITD
	70.79	complex	wt	wt	wt			81.37	45, XX, del(5q)	wt	wt	wt
13	71.11	47, XX, +10	SM	wt	wt			82.03	47, XY, +13	wt	wt	wt
	71.94	46, XY, t(6;9)	wt	wt	ITD			82.47	46, XY	wt	wt	wt
	72.03	46, XY, del(9q)	wt	MUT	wt			84.39	48, XX, +2mar	wt	wt	wt
	72.20	46,XY, t(8;21)	wt	wt	wt			84.47	46, XY	wt	wt	wt
	72.27	46, XY	wt	wt	wt			85.28	45, X, -Y	wt	wt	wt
	73.10	46, XY	wt	MUT	wt			86.75	45,X, -Y, t(8;21)	wt	wt	wt
	73.16	46, XX	wt	wt	wt			87.03	46, XX	wt	MUT	wt
	73.78	46, XX	wt	wt	wt			87.30	46, XX	wt	MUT	D835Y
	74.21	47, XY, +8	wt	wt	wt			87.32	46, XY	wt	MUT	wt
	74.27	46, XX, t(2;11q22), del(5q22)	wt	wt	wt			87.57	complex	wt	wt	wt
	74.52	46, XX	wt	MUT	ITD			87.70	47, XX, +8	wt	wt	wt
	74.78	46, XY	wt	wt	wt			87.75	46, XY	wt	wt	ITD
	75.03	46, XY	wt	wt	wt			88.89	47, XY, +13	wt	wt	D835Y
	75.20	46, XX, t(15;17)	wt	wt	wt			89.96	46, XX	wt	wt	wt
	75.37	46, XY, t(8;21)	wt	wt	wt			90.14	47, XY, +8	wt	wt	wt
	75.67	46, XY	wt	MUT	wt			91.86	46, XY	wt	MUT	ITD

ID	Score	Genetics					ID	Score	Genetics			
		Karyotype	*CEBPA*	*NPM1*	*FLT3*				*Karyotype*	*CEBPA*	*NPM1*	*FLT3*
	92.17	46, XX	wt	MUT	ITD			107.41	46, XY, t(15;17)	wt	wt	ITD
	93.34	46, XX	wt	wt	wt			107.74	46, XY	wt	wt	wt
	94.22	46, XY	wt	MUT	wt			108.00	46, XY, t(15;17)	wt	wt	wt
	94.33	46, XY	wt	wt	ITD			108.15	complex	wt	wt	wt
	94.46	46, XY	wt	wt	D835Y			108.63	46, XY, t(15;17)	wt	wt	ITD
	95.27	46, XY	wt	MUT	D835Y			108.92	46, XX	wt	wt	wt
	95.69	complex	wt	wt	wt			109.15	46, XY, t(15;17), del(9q)	wt	wt	wt
	96.01	46, XX	wt	MUT	ITD			109.37	46, XX	wt	MUT	D835Y
	96.28	47, XX, +21	wt	wt	wt			109.93	46, XX	wt	MUT	wt
	96.32	46, XY	wt	MUT	ITD			110.23	46, XY, t(6;11)	wt	wt	wt
	97.06	complex	wt	wt	wt			110.89	46, XY, t(15;17)	wt	wt	wt
	97.16	46, XY	wt	wt	wt			112.23	46, XY	wt	MUT	wt
	97.64	46, XY	wt	wt	wt			112.67	46, XY	wt	MUT	wt
	98.04	46, XX, del(11q23)	wt	MUT	wt			113.28	47, XX,+21, del(20q)	wt	wt	ITD
	98.39	46, XX	wt	wt	ITD			113.93	46, XY	wt	MUT	D835Y
	98.79	46, XY, t(11;19)	wt	wt	wt			114.06	46, XY, t(15;17)	wt	wt	wt
	98.85	complex	wt	wt	wt			114.17	46, XX	wt	MUT	wt
	100.02	NA	wt	MUT	ITD			114.34	46, XY	wt	wt	wt
	100.64	NA	wt	wt	wt			114.99	46, XX, del(7q)	wt	wt	wt
	100.73	46, XY, del(9q)	wt	MUT	wt			115.10	46, XX	wt	wt	wt
	101.15	complex	wt	wt	wt			115.16	46, XX	wt	wt	wt
	101.39	46, XY	wt	MUT	ITD			115.28	46, XY	wt	MUT	wt
	101.55	complex	wt	wt	wt			115.62	46, XX, t(8;21)	wt	wt	ITD
	101.61	47, XX, +14	wt	wt	wt			116.61	46, XY	wt	MUT	ITD
	103.93	46, XX	wt	wt	wt			116.75	47, XY, +8	wt	wt	ITD
17	104.46	46, XY	SM	wt	wt			117.51	NA	wt	wt	wt
	104.54	46, XY	wt	MUT	wt			117.82	46, XY	wt	wt	wt
	104.65	46, XX	wt	MUT	D835Y			118.65	46, XY	wt	wt	wt
	106.07	46, XX	wt	MUT	D835Y			119.18	46, XX	wt	MUT	wt
	106.98	46, XX, t(9;22)	wt	wt	wt			119.90	46, XY, t(15;17)	wt	wt	wt

ID	Score	Genetics					ID	Score	Genetics			
		Karyotype	*CEBPA*	*NPM1*	*FLT3*				*Karyotype*	*CEBPA*	*NPM1*	*FLT3*
	119.98	47, XX, +8	wt	wt	D835Y			138.96	complex	wt	MUT	wt
	121.20	46, XX	wt	MUT	wt			140.25	46, XY	wt	wt	wt
	121.41	46, XY, del(11q23)	wt	wt	wt			140.78	46, XY	wt	wt	wt
	123.17	46, XY, t(15;17)	wt	wt	ITD			141.29	46, XX, del(6q)	wt	MUT	wt
	124.07	46, XX	wt	MUT	ITD			141.94	46, XX	wt	MUT	ITD
	124.19	46, XX	wt	wt	wt			142.02	47, XY, +21	wt	wt	wt
	124.44	46, XY	wt	MUT	wt			142.22	46, XY	wt	MUT	ITD
	125.63	46, XX, del(9q)	wt	MUT	ITD			144.06	46, XX	wt	MUT	ITD
	125.97	46, XY, del(7q22)	wt	wt	wt			144.24	46, XX, del(17p)	wt	MUT	wt
	126.04	46, XY	wt	MUT	wt			144.85	46, XX	wt	MUT	wt
	127.67	46, XX, t(8;21)	wt	wt	wt			147.38	46, XY	wt	MUT	wt
	128.05	46, XX, inv(16)	wt	wt	ITD			149.35	47, XY, +8	wt	wt	wt
	128.60	complex	wt	wt	wt			150.27	46, XX, inv(12)(p13.3q13.1)	wt	wt	wt
	129.14	46, XX, inv(4)(p?14q.21.3)	wt	MUT	wt			152.37	46, XY	wt	MUT	wt
	129.82	45, XY, -7	wt	wt	wt			153.27	46, XY	wt	MUT	ITD
15	131.11	46, XY	SM	MUT	wt			153.68	45, XY, -7	wt	wt	D835Y
	131.33	46, XX	wt	MUT	ITD			154.22	46, XY, t(15;17)	wt	wt	wt
	131.78	46, XY	wt	MUT	D835Y			159.04	46, XY	wt	wt	ITD
	131.89	46, XY	wt	wt	ITD			160.07	46, XY, t(9;11)	wt	wt	wt
	132.25	46, XY	wt	wt	wt			160.48	46, XX	wt	MUT	ITD
	133.21	46, XY	wt	wt	wt			161.19	46, XX	wt	MUT	wt
	133.22	complex	wt	wt	ITD			161.45	46, XY	wt	wt	wt
	136.08	46, XX	wt	wt	ITD			164.00	46, XY	wt	MUT	ITD
	137.14	46, XY	wt	wt	ITD			165.29	46, XY	wt	MUT	ITD
	138.65	46, XY	wt	wt	wt			169.43	46, XX	wt	MUT	ITD

ID	Score	Genetics					ID	Score	Genetics			
		Karyotype	*CEBPA*	*NPM1*	*FLT3*				*Karyotype*	*CEBPA*	*NPM1*	*FLT3*
	171.20	46, XY	wt	MUT	wt			193.11	46, XX	wt	MUT	wt
	177.78	46, XX	wt	MUT	wt			194.25	46, XX	wt	MUT	wt
	182.70	46, XX	wt	MUT	ITD			195.78	46, XY	wt	MUT	ITD
	186.52	46, XY	wt	MUT	ITD			200.08	46, XX	wt	MUT	wt
	189.76	46, XY	wt	MUT	ITD			210.23	46, XX	wt	MUT	ITD
	189.93	47, XY, +8	wt	wt	ITD			228.15	46, XX	wt	MUT	wt

Abbreviations: NA, not available; ITD, internal tandem duplication; MUT, mutated; SM, single mutation; DM, double mutation; wt, wild type. Lack of growth means no metaphases.

Table SVII. Characterization of CEBPA mutations in the validation cohort.

ID, sm/dm	Mutation 1 – Position in CDS		Mutation 1- Position in protein	Mutation 2 – Position in CDS		Mutation 2 – Position in protein
	N-term, nt-29_518	C-term, nt-816_1171	AA consequence	Middle, nt-469_858	C-term, nt-816_1171	AA consequence
1, dm	c.196_197insT		p.A66Vfs*107		c.902_934dup33bp	p.312_322insHKAKQ RNVETQ
2, dm	c.232_271delinsGA GGT		p.L78Efs*95		c.937_939delAAG	p.K313del
3, dm	c.241delC		p.L81Cfs*158		c.890G>C	p.R297P
4, dm	c.284_285insACG G		p.G96Rfs*108		c.913_927delCAGCGCAACGTGG AG	p.Q305_E309del
5, dm	c.221_222insAA		p.N74Kfs*160		c.937_938 TGTCGCAGCGCATCATGGAGA CCCATCACA	p.Q312_K313insMSQ RIMETHH
6, dm	c.177_194dup ACGTCCATCGA CATCGG		p.S6Rfs*165		c.912_929del GCAACGTGGAGAC	p.R306_Q311del
7, dm	c.146dupC		p.E50Gfs*107		c.913_918dupCAGCGC, A919delinsG	p.Q 305_R306dup, N307delinsD
8, dm	c.68delC		p.P23Rfs*159		c.917_918GGAGCG	p.R306_N307insRE
9, dm	c.68delC		p.P23Rfs*159		c.937_939dupAAG	p.K313dup
10, dm	c.87C>A, 92_126delTCGGC TTTCCCCGGGG CGCGGGCCCCG CGCAGCCT		p.F31Sfs*94		c.917_918insTTG	p.R306_N307insC
11, dm	c.348_349delCG		p.G117Rfs*167	c.569C>A		p.S190*
12, dm		c.937_939dupAAG	p.K313dup		c.971T>A	p.L324Q
13, sm	c.134delC		p.P46Lfs*158			
14, sm	c.166dupT		p.C56Lfs*107			

15, sm	c.16_20dupGACTT		p.F6Lfs*17			
16, sm		c.925_935delinsCT	p.309E_312QdelinsL			
17, sm		c.916dupC	p.P306fs*319			
18, sm		c.913_927delCAGCGCAACGTGGAG	p.Q305_E309del			

Abbreviations: AA amino acid number; CDS coding DNA sequence, ins, insertion; del, deletion; dup, duplication, nt, nucleotide. Nucleotides numbered from the major translational start codon at nucleotide position mutated. NCBI Reference Sequence is NM_004364.4. The description of sequence variants is according to nomenclature of HVGS site. The part in the Middle (nt-469_858) of the coding sequence is not reported for Mutation 1 because no patients with single mutation in this region were observed.