SUPPLEMENTARY APPENDIX

Variable outcome and methylation status according to *CEBPA* mutant type in double-mutated acute myeloid leukemia patients and the possible implications for treatment

Dima El-Sharkawi,[1] Duncan Sproul,[2] Christopher G. Allen,[1] Andrew Feber,[3] Melissa Wright,[4] Robert K. Hills,[4] David C. Linch[1] and Rosemary E. Gale[1]

[1]Department of Haematology, UCL Cancer Institute, London; [2]MRC Human Genetics Unit and Edinburgh Cancer Research Centre, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh; [3]Medical Genomics, UCL Cancer Institute, London and [4]Centre for Trials Research, Cardiff University, UK

 doi:10.3324/haematol.2017.173096

Received: May 19, 2017.

Accepted: October 10, 2017.

Pre-published: October 12, 2017.

Correspondence: rosemary.gale@ucl.ac.uk

SUPPLEMENTARY MATERIAL

Therapy, clinical end points and statistical methods

Details of the clinical protocols have been published elsewhere.[1-3] CR was defined as a normocellular bone marrow (BM) containing <5% blasts and showing evidence of normal maturation of other marrow elements. Persistence of myelodysplastic features did not preclude the diagnosis of CR. OS was the time from trial entry to death. For patients achieving CR, relapse-free survival (RFS) was the time from the date of first CR to an event (death in first CR or relapse) and relapse rate (RR) was the cumulative probability of relapse, censoring at death in CR.

Mantel-Haenszel and chi-squared tests were used to test for differences in demographic and clinical data by genotype. Kaplan-Meier curves were constructed for survival data and compared by means of the log-rank test. Surviving patients were censored on August 2010 for AML10 and AML12, March 2015 for AML15. Median follow-up for survival was 9.5 years (range, 0.1-22 years). Multivariable Cox models were used to analyze OS, RFS and RR, adjusting for age, white cell count, WHO performance status, secondary leukemia and trial, with additional variables of interest (*CEBPA* mutant type, *FLT3*^{ITD} and *DNMT3A* genotype). Models were fitted using forward selection, with variables added to the model if they had a *P* value, derived using the deviance statistic, of <0.05. Odds ratios (OR) or hazard ratios (HR) and 95% confidence intervals (CI) are quoted for endpoints. In all cases a ratio of <1 indicates benefit. All *P* values are two-tailed.

Bisulfite conversion and methylation-specific PCR to assess conversion efficiency

An aliquot of 350-500ng DNA was bisulfite-converted using the EZ DNA Methylation-Gold Kit (Zymo Research, California, USA) according to the manufacturer's instructions except that the incubation at 50^{o}C for 16 hours included a denaturation step at 95^{o}C for 30 seconds at the beginning of each hour. This periodic cycling has been reported to improve conversion efficiency.[4]

Several samples were randomly selected from each converted batch and subjected to two methylation-specific PCRs of the *HLA-B* gene, one using primers that would only

amplify bisulfite-converted DNA and the other using primers that would only amplify unconverted DNA (see Table S5). For each PCR, 1µl of converted DNA was added to 1x manufacturer's buffer, 3mM $MgCl_2$, 200µM dNTPs, 25pmols each primer and 1U GoTaq polymerase (Promega) in a total volume of 25µl. An initial denaturation step at 95°C for 6 minutes was followed by 36 cycles of PCR, each of 94°C for 30 seconds, annealing temperature for 30 seconds and 72°C for 30 seconds, with 2 cycles at annealing temperatures of 60°C, 59°C and 58°C, then 30 cycles at 57°C, followed by a final extension step of 72°C for 15 minutes. PCR products were run on a 2% agarose gel. Samples were considered to be successfully bisulfite-converted if they had a PCR product with the primers for the bisulfite-converted DNA but no product with the primers for the unconverted DNA.

Data processing for the methylation arrays

Data from the Illumina Infinium HumanMethylation27 and HumanMethylation450 BeadChip arrays (Illumina Inc, California, USA) were exported from Illumina's genome studio and subsequent analysis was performed using R. Beta values were used as a measure of the methylation level at a given CpG probe as derived from the intensity of the methylated (I_{meth}), and unmethylated (I_{unmeth}) allele probes (I_{meth} /(I_{meth} + I_{unmeth}). The data from 450K arrays was normalized to correct for systematic biases between the two types of probes present on these arrays as described.[5] Beta values were then filtered to remove unreliable data points based on the detection *P* value from the Infinium arrays (threshold 0.01) by setting their values to NA. Finally, probes displaying gender-specific biases were filtered out, defining them as those with Wilcoxon tests *P* values <0.05 between genders. This resulted in the exclusion of 1,362 probes from the first cohort (Infinium 27K array) and 56,356 probes from the follow-up cohort (Infinium 450K array). Beta values were converted to percentage methylation levels by multiplying by 100.

Pyrosequencing assays

Pyrosequencing assays to interrogate specific CpG sites were designed using the PyroMark Assay Design Software 2.0 (Qiagen, Crawley, UK). For each PCR, 25ng DNA was added to 1x manufacturer's buffer, 3.5mM $MgCl_2$, 200µM dNTPs, 0.2µM primers, one biotin-labeled and one unlabeled, and 1.25U GoTaq polymerase (Promega) in a total

volume of 20μl. The mixes were denatured at 95°C for 5 minutes and then subjected to 50 cycles of amplification using primers and annealing temperatures as specified in Table S5. Products were sequenced on a PyroMark MD system, (Qiagen) using PyroMark Q96 reagents and protocols and the appropriate primer as specified in Table S5. All samples were assayed in duplicate and the mean methylation level expressed as a percentage of the total alleles for a particular site. For each assay, titration curves were created using standards containing 0%, 10%, 25%, 50%, 75%, 90% and 100% methylated DNA, prepared from bisulfite-converted fully methylated and unmethylated DNA (Epitect Control DNA set, Qiagen), to check for accuracy and sensitivity.

REFERENCES

1. Hann IM, Stevens RF, Goldstone AH, et al. Randomized comparison of DAT versus ADE as induction chemotherapy in children and younger adults with acute myeloid leukemia. Results of the Medical Research Council's 10th AML trial (MRC AML10). Adult and Childhood Leukaemia Working Parties of the Medical Research Council. Blood. 1997;89(7):2311-2318.
2. Burnett AK, Hills RK, Milligan DW, et al. Attempts to optimize induction and consolidation treatment in acute myeloid leukemia: results of the MRC AML12 trial. J Clin Oncol. 2010;28(4):586-595.
3. Burnett AK, Russell NH, Hills RK, et al. Optimization of chemotherapy for younger patients with acute myeloid leukemia: results of the medical research council AML15 trial. J Clin Oncol. 2013;31(27):3360-3368.
4. Thirlwell C, Eymard M, Feber A, et al. Genome-wide DNA methylation analysis of archival formalin-fixed paraffin-embedded tissue using the Illumina Infinium HumanMethylation27 BeadChip. Methods. 2010;52(3):248-254.
5. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium Methylation 450K technology. Epigenomics. 2011;3(6):771-784.

SUPPLEMENTARY TABLES

Table S1. Full details of *CEBPA* mutations in the patients studied

Coh	Genotype	Mutant 1	Mutant 2	Type of mutation	Predicted proteins
Clin	Classic DM	p.H18fs	p.E309del	Classic N + C	p30 + C-LOF
		p.H18fs	p.Q312del		
		p.S21fs	p.N307_Q312del		
		p.S21fs	T310dup		
		p.P23fs	K304_Q305insL		
		p.P23fs	p.N307_E309delinsK		
		p.P23fs	p.L315_E316insT		
		p.H24fs	p.K304_Q305insL		
		p.H24fs	p.K304_R306dup		
		p.H24fs	p.V308dup		
		p.H24fs	p.Q311_Q312insL		
		p.H24fs	p.Q311_Q312insL		
		p.H24fs	p.K313Nins14		
		p.H24fs	p.K313_V314insQK		
		p.A29fs	p.L315dup		
		p.A30fs	p.Q312Hins15		
		p.F31fs	p.K313dup		
		p.F33fs	p.K326delins23		
		p.R35fs	p.E309dup		
		p.G36fs	p.R297_Q311dup		
		p.G36fs	p.D301_L315dup		
		p.G36fs	p.K304_Q312dup		
		p.G36fs	p.R306_N307insKQR		
		p.G36fs	p.E309delinsKTQ		
		p.A37fs	p.E309_T310insSQ		
		p.G38fs	p.T310_Q312delinsK		
		p.G38fs	p.D320_N321ins16		
		p.A40fs	p.E309_L317dup		
		p.P43fs	p.Q312dup		
		p.P49fs	p.Q305_E316dup		
		p.E50fs	p.V314_L315ins13		
		p.G53fs	p.K313dup		
		p.G54fs	p.Q305_R306insQQ		
		p.I55fs	p.Q312_K313ins12		
		p.C56fs	p.K302_R306dup		
		p.E59fs	p.E309dup		
		p.E59fs	p.T318_S319insI		
		p.I62fs	p.E309delinsAQ		
		p.I62fs	p.Q312dup		
		p.I62fs	p.E329_Q330ins35		
		p.A66fs	p.T310_Q311insP		
		p.A66fs	p.E316_L317insR		
		p.Y67X	p.R297_V308dup		
		p.I68fs	p.R306_N307insRR		
		p.I68fs	p.K313dup		
		p.A72fs	p.K302_E309dup		
		p.F73fs	p.R306dup		
		p.F77fs	p.R306_Q311del		
		p.L78fs	p.R300_L324dupinsR		
		p.L78fs	p.V308_R327dup		
		p.L78fs	p.K313dup		
		p.A79fs	p.Q312dup		
		p.F82fs	p.K304_V308dup		
		p.Q83fs	p.R300_Q312dup		
		p.Q83fs	p.Q312_K313insE		
		p.Q83fs	p.K313dup		
		p.Q83fs	p.K313dup		
		p.H84fs	p.K302_K304dup		
		p.Q88fs	p.K304_Q305insL		
		p.A91fs	p.S319delinsRL		
		p.V95fs	p.V308I+Q312dup		
		p.G96fs	p.T310_Q311ins10		
		p.G96fs	p.K313dup		
		p.G96fs	p.L317_T318insM		
		p.T98fs	p.T310dup		
		p.G99fs	p.K304_Q305insL		
		p.G101fs	p.Q312dup		
		p.F106fs	p.D301_V308dup		
		p.F106fs	p.K302_R306dup		
		p.D107fs	p.S299_T318dup		
		p.D107X	p.D301del		
		p.D107fs	p.E309delinsGQ		
		p.Y108X	p.E309_T310insK		
		p.G110fs	p.Q312dup		

Clin (cont)		p.G110fs p.P112fs p.A113fs p.G114fs p.G114fs	p.Q312dup p.K304_Q305insL p.L315_E316insQ p.Q312dup p.K313dup		
	Non-classic DM	p.A44fs p.I68fs p.L78fs p.A79fs	p.A295P p.R297P p.R300P p.Q305P	Classic N + C-missense Classic N + C-missense Classic N + C-missense Classic N + C-missense	p30 + C-LOF
		p.H24fs p.H24fs p.G36fs p.G38fs p.G54fs p.E89fs (Hom) p.V95fs p.P121fs	p.Q209fs p.G223fs p.Y181X p.K313fs p.E167fs p.A238fs p.Q312X	Classic N + mid-frameshift Classic N + mid-frameshift Classic N + mid-frameshift Classic N + C-frameshift Classic N + mid-frameshift Classic N Classic N + mid-frameshift Classic N + C-frameshift	p30 + null
		p.L178fs p.R300_D301selinsQN (Hom) p.R306_V314dup p.Q312delinsPK p.K313dup p.K313dup p.K313dup (Hom) p.V314G (Hom) p.E316_L317insR p.E316_R325dup (Hom) p.L317Q (Hom) p.L317_K326dup (Hom) p.N321S (Hom)	p.K276R p.N321S p.R343fs p.R343fs p.S349fs p.R343fs	Mid-frameshift + C-missense Classic C Classic C + C-missense Classic C + C-frameshift Classic C + C-frameshift Classic C + C-frameshift Classic C C-missense Classic C + C-frameshift Classic C C-missense Classic C C-missense	C-LOF
M1	Classic DM	p.H24fs p.G54fs p.A66fs p.I68fs p.L78fs p.H84fs p.A91fs p.G114fs	p.Q311_Q312insL p.Q305_R306insQQ p.E316_L317insR p.R306_N307insRR p.E309_V328dup p.K302_K304dup p.S319delinsRL p.Q312dup	Classic N + C	p30 + C-LOF
	Non-classic DM	p.K313dup p.V314G (Hom)	p.R343fs	Classic C + C-frameshift C-missense	C-LOF
M2	Classic DM	p.P23fs p.H24fs p.H24fs p.H24fs p.A40fs p.P43fs p.E50fs p.I55fs p.E59fs p.Y67X p.A72fs p.A79fs p.Q88fs p.G96fs p.D107fs p.P112fs p.A113fs	p.N307_E309delinsK p.K304_Q305insL p.Q311_Q312insL p.K313Nins14 p.E309_L317dup p.Q312dup p.V314_L315ins13 p.Q312_K313ins12 p.E309dup p.R297_V308dup p.K302_E309dup p.Q312dup p.K304_Q305insL p.T310_Q311ins10 p.S299_T318dup p.K304_Q305insL p.L315_E316insQ	Classic N + C	p30 + C-LOF
	Non-classic DM	p.A44fs p.I68fs p.L78fs	p.A295P p.R297P p.R300P	Classic N + C-missense Classic N + C-missense Classic N + C-missense	p30 + C-LOF
		p.H24fs p.H24fs p.G36fs p.G38fs p.G54fs p.V95fs	p.Q209fs p.G233fs p.Y181X p.K313fs p.E167fs p.A238fs	Classic N + mid-frameshift Classic N + mid-frameshift Classic N + mid-nonsense Classic N + C-frameshift Classic N + mid-frameshift Classic N + mid-frameshift	p30 + null
		p.R300_D301delinsQN (Hom) p.E316_R325dup (Hom) p.N321S (Hom) p.L317Q (Hom) p.R306_V314dup	 p.N321S	Classic C Classic C C-missense C-missense Classic C + C-missense	C-LOF
	SM	p.P23fs p.H24fs p.P46fs p.E59X p.T60fs p.S61fs p.A66fs p.I68fs p.Q83fs		Classic N	p30 + WT

M2 (cont)		p.R300_D301delinsQY p.K313dup (x 2) p.R323del p.N356_C357del		Classic C	C-LOF + WT
		p.G122fs p.G123fs p.E166X p.D168fs p.P180fs p.Q215X p.L253fs p.S266fs p.K275fs		Mid-frameshift/nonsense	Null + WT
		p.N292fs p.N307fs		C-frameshift	Null + WT
		p.Y285S p.R289C p.V296E p.L331Q p.I341V		C-missense	C-LOF + WT
		p.P187_P189del p.H193_P196del p.P239_A240del		Mid-indel	UNK + WT
		p.P183Q p.P233R p.G242S (x 2) p.K276R		Mid-missense	UNK + WT

Null indicates mid-region or C-terminal mutants with a truncating frameshift or nonsense mutation.

Abbreviations: C, C-terminal mutation; Clin, Clinical cohort; Coh, cohort; DM, double *CEBPA* mutant; del, deletion; dupl, duplication; fs, frameshift; Hom, homozygous; indel, insertion/deletion; ins, insertion; LOF, loss-of-function; M1, Methylation cohort 1; M2, Methylation cohort 2; Mid, mid-region; N, N-terminal mutation; SM, single *CEBPA* mutant; UNK, unknown; WT, wild-type

Table S2. Demographic details of the 104 patients in the clinical analysis

Parameter	Classic *CEBPA*^{DM} (n=79)	Non-classic *CEBPA*^{DM} (n=25)	*P**
Median age, years (IQR)	35 (25 to 46)	47 (42 to 55_	0.001**
Median WBC, x10^9/L (IQR)	27.6 (10.5-67.4)	15.9 (7.7-92.9)	0.7**
Sex:			0.8
Male	44 (56%)	15 (60%)	
Female	35 (44%)	10 (40%)	
Trial:			0.7
AML10	16 (21%)	6 (24%)	
AML12	31 (39%)	11 (44%)	
AML15	32 (41%)	8 (32%)	
WHO Performance status:			0.1
0	55 (70%)	13 (52%)	
1	11 (14%)	8 (32%)	
2	12 (15%)	3 (12%)	
3	1 (1%)	1 (4%)	
AML type:			1.0
De novo	77 (97%)	25 (100%)	
Secondary	2 (3%)	0 (0%)	
Transplant status:			0.8
No transplant	48 (61%)	14 (56%)	
Allograft	19 (24%)	7 (28%)	
Autograft	9 (11%)	4 (16%)	
Other	3 (4%)	0 (0%)	
Stage of transplant:			0.7
First remission	18 (58%)	6 (55%)	
First relapse	1 (3%)	1 (9%)	
Second remission	12 (39%)	4 (36%)	
Cytogenetics:			0.05
FR	0 (0%)	1 (5%)	
IR	68 (100%)	17 (89%)	
NK	53 (78%)	12 (63%)	
AK	15 (22%)	5 (26%)	
AR	0 (0%)	1 (5%)	
No result	11	6 (24%)	
FLT3^{ITD}	8 (10%)	1 (4%)	0.7
NPM1^{MUT}	2 (3%)	1 (4%)	0.6
DNMT3A^{MUT}	2/76 (3%)	4 (16%)	0.03
IDH1^{MUT}	0/60 (0%)	1/21 (5%)	0.3
IDH2^{MUT}	1 (1%)	5 (20%)	0.003
WT1^{MUT}	7/71 (10%)	0/20 (0%)	0.3
TET2^{MUT}	3/30 (10%)	5/18 (28%)	0.1
GATA2^{MUT}	28/72 (39%)	4/23 (17%)	0.1

**P* values are Fishers exact test unless otherwise stated; **Wilcoxon rank sum test
Abbreviations: AK, intermediate-risk abnormal karyotype; AR, adverse risk; FR, favorable risk; IQR, interquartile range; IR, intermediate risk; NK, normal karyotype; WBC, white blood cell count

Table S3. The top 25 most differentially methylated sites between *CEBPA*$^{\text{Classic-DM}}$ and *CEBPA*$^{\text{WT}}$ samples. The probes were selected by Wilcoxon tests and the median difference between the samples (mean rank of both parameters) and are ordered by their median beta value in *CEBPA*$^{\text{Classic-DM}}$ samples.

Genomic Location				Gene			Signature Details					
Probe ID	Chr	Position	Dist. to TSS	ENSEMBL Gene ID	Symbol	Description	In CpG Island	WT Median β	Classic DM Median β	Wilcoxon P-value	Median Δ	Rank
cg21237418	17	24069170	-157	ENSG00000109113	RAB34	RAB34, member RAS oncogene family	FALSE	0.153	0.928	4.91E-07	0.775	2
cg14338887	6	43036478	0	ENSG00000124713	GNMT	Glycine N-methyltransferase	TRUE	0.170	0.918	3.93E-06	0.748	3
cg17186163	10	44794323	7	ENSG00000165507	C10orf10	Chromosome 10 open reading frame 10	FALSE	0.155	0.907	4.09E-08	0.752	1
cg24101359	6	43036473	5	ENSG00000124713	GNMT	Glycine N-methyltransferase	TRUE	0.199	0.904	1.89E-05	0.705	20
cg13105904	14	23969884	-903	ENSG00000100441	KHNYN	KH and NYN domain containing	TRUE	0.306	0.889	5.56E-06	0.583	10
cg01274660	7	100303561	-675	ENSG00000087077	TRIP6	Thyroid hormone receptor interactor 6	FALSE	0.317	0.888	1.80E-05	0.572	25
cg04355435	1	43508877	117	ENSG00000179178	TMEM125	Transmembrane protein 125	FALSE	0.349	0.879	7.69E-06	0.531	22
cg10056627	6	43036751	-273	ENSG00000124713	GNMT	Glycine N-methyltransferase	TRUE	0.249	0.877	1.42E-05	0.629	16
cg27588902	6	43036129	349	ENSG00000124713	GNMT	Glycine N-methyltransferase	TRUE	0.265	0.854	2.74E-06	0.589	4
cg25651505	2	85665534	-492	ENSG00000168899	VAMP5	Vesicle associated membrane protein 5	TRUE	0.347	0.837	2.74E-06	0.489	18
cg23696834	6	43036323	155	ENSG00000124713	GNMT	Glycine N-methyltransferase	TRUE	0.102	0.822	3.93E-06	0.720	5
cg24081819	8	27404857	-295	ENSG00000120915	EPHX2	Epoxide hydrolase 2, cytoplasmic	TRUE	0.243	0.817	3.93E-06	0.574	9
cg08965235	11	65081734	541	ENSG00000168056	LTBP3	Latent TGF beta binding protein 3	TRUE	0.266	0.804	2.74E-06	0.538	11
cg16068833	1	26517102	-104	ENSG00000169442	CD52	CD52 molecule	FALSE	0.227	0.763	1.06E-05	0.536	24
cg19764555	11	62071695	-787	ENSG00000124942	AHNAK	AHNAK nucleoprotein	TRUE	0.272	0.763	2.74E-06	0.491	17
cg00350296	11	65841417	-343	ENSG00000174807	CD248	CD248 molecule, endosialin	FALSE	0.253	0.738	2.74E-06	0.485	19
cg10798171	7	8268826	-59	ENSG00000003147	ICA1	Islet cell autoantigen 1	TRUE	0.249	0.715	4.09E-08	0.466	13
cg15032239	15	20443395	709	ENSG00000068793	CYFIP1	Cytoplasmic FMR1 interacting protein 1	TRUE	0.195	0.708	4.09E-08	0.513	6
cg16155382	1	24518778	-135	ENSG00000158055	GRHL3	Grainyhead-like transcription factor 3	FALSE	0.101	0.678	3.69E-06	0.577	8
cg13490971	5	141468305	203	ENSG00000131507	NDFIP1	Nedd4 family interacting protein 1	TRUE	0.203	0.653	4.09E-08	0.450	14
cg21697134	17	78287128	-331	ENSG00000167363	FN3K	Fructosamine 3 kinase	FALSE	0.090	0.614	1.04E-07	0.524	7
cg08897388	6	112682398	44	ENSG00000112769	LAMA4	Laminin subunit alpha 4	TRUE	0.135	0.575	1.84E-06	0.440	23
cg12417466	3	35658823	30	ENSG00000172995	ARPP21	CAMP regulated phosphoprotein 21kDa	FALSE	0.720	0.179	5.56E-06	-0.541	15
cg05615150	3	35658819	34	ENSG00000172995	ARPP21	CAMP regulated phosphoprotein 21kDa	FALSE	0.668	0.113	5.56E-06	-0.555	12
cg18920397	1	159032429	123	ENSG00000122224	LY9	Lymphocyte antigen 9	FALSE	0.625	0.071	1.06E-05	-0.554	21

Table S4. Details of primers and conditions for methylation-specific PCR and pyrosequencing assays.

	Gene	Primer	Primer Sequence	Size (bp)	Annealing Temp (°C)
Methylation-specific PCR for bisulfite conversion	*HLA-B* converted	F	5'-TTTTAAGTTTTATTTTTGTGGGGTA-3'	300	Touchdown
		R	5'-AAATCCCAACTAATAACTATTTTTCAA-3'		(see text)
	HLA-B unconverted	F	5'-CCCAAAGTCCACTAACATTAGAA-3'	464	Touchdown
		R	5'-GCTGAGAAAATAGCCTCAGAATA-3'		(see text)
Pyrosequencing assays for array validation	*SOCS2*	F (biotinylated)	5'-AGGTGGGAAGTAAAGAATAAGATGGA-3'	128	62
		R	5'-CCAAACCTAAATCCCTAAAAAACCACTTT-3'		
		Sequencing	5'-CCTAAAAAAACCACTTTCCT-3'		
	WNT2	F	5'-GTGTATGAAATGATGGTAAGAGATGTT-3'	246	62
		R (biotinylated)	5'-ATACATAATAATCTCCTTATCCCCTAACC-3'		
		Sequencing	5'-GGGAAGGGGGAATATYGTTGTATG-3'		
	PRF1 A	F	5'-AGTAGGGTTATTTTTTTGTTTTTGATGT-3'	150	60
		R (biotinylated)	5'-CCTACCAATCCACACTACTAATACA-3'		
		Sequencing	5'-GTTATTTTTTTGTTTTTGATGTATA-3'		
	PRF1 B	F	5'-TAGGAAGTGTTGTGATTTATAAGATAAG-3'	163	60
		R (biotinylated)	5'CTTTAATATCAACACTTACAAAACCTTAA-3'		
		Sequencing	5'-TAAGATAAGATATTTGGGTTA-3'		
Pyrosequencing assays for good-risk patients	*LY9*	F (biotinylated)	5'-TGTTTTAGAGGGAGGGTTGTTTATA-3'	100	58
		R	5'-AATCACAAATAAAACCCTAAATAAACTTA-3'		
		Sequencing	5'-TAAAACCTCTACCTACC-3'		
	KHNYN	F (biotinylated)	5'-GGGTTTTTTAGTTGTAGTTAGATGTG-3'	192	60
		R	5'-ACTAAAAACAACAACCATACCTAC-3'		
		Sequencing	5'-ACCCCATATAAAACCCATCTTC-3'		
	VAMP5	F	5'-GTGTTYGTTTATTAGGTAGAGGTGTTA-3'	281	59
		R (biotinylated)	5'-CCCRCCTAAACCCTCACCATC-3'		
		Sequencing	5'- GTTTATYGTTTTYGATTTGATTTGG-3'		

SUPPLEMENTARY FIGURES

Figure S1. Flowchart showing the patients investigated in the different analyses performed.

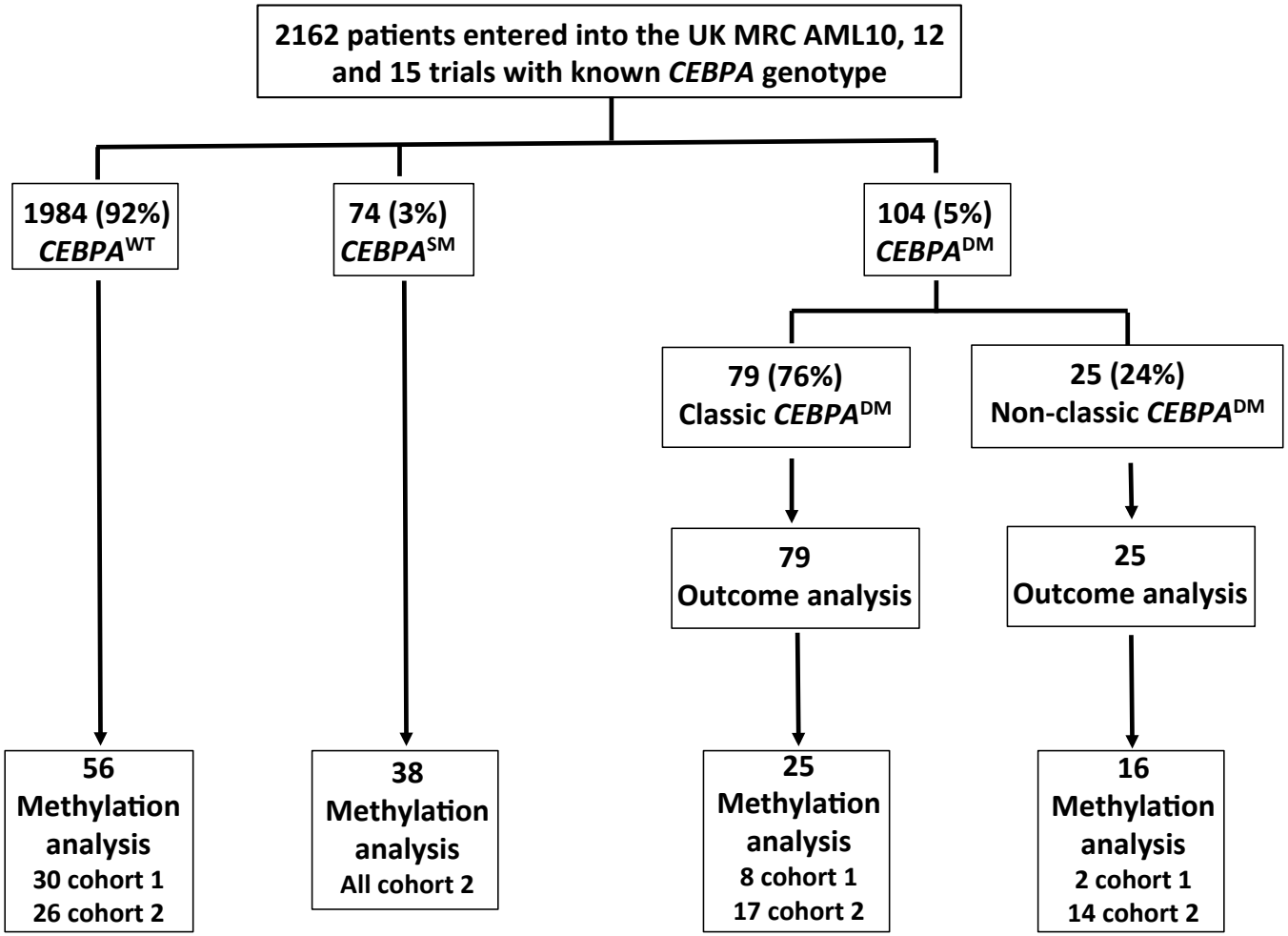

Figure S2. Comparison of methylation quantification values obtained at four CpG sites investigated on the arrays and by pyrosequencing using the preliminary set of 40 samples. (a-d) β values obtained from the methylation array versus % methylation by pyrosequencing. (e-h) Bland-Altman plots showing the mean result for each sample compared with the difference in values between the two assays. The dotted line indicates the 95% confidence intervals for each probe. Details of PCR primers and assay conditions are given in Table S4.

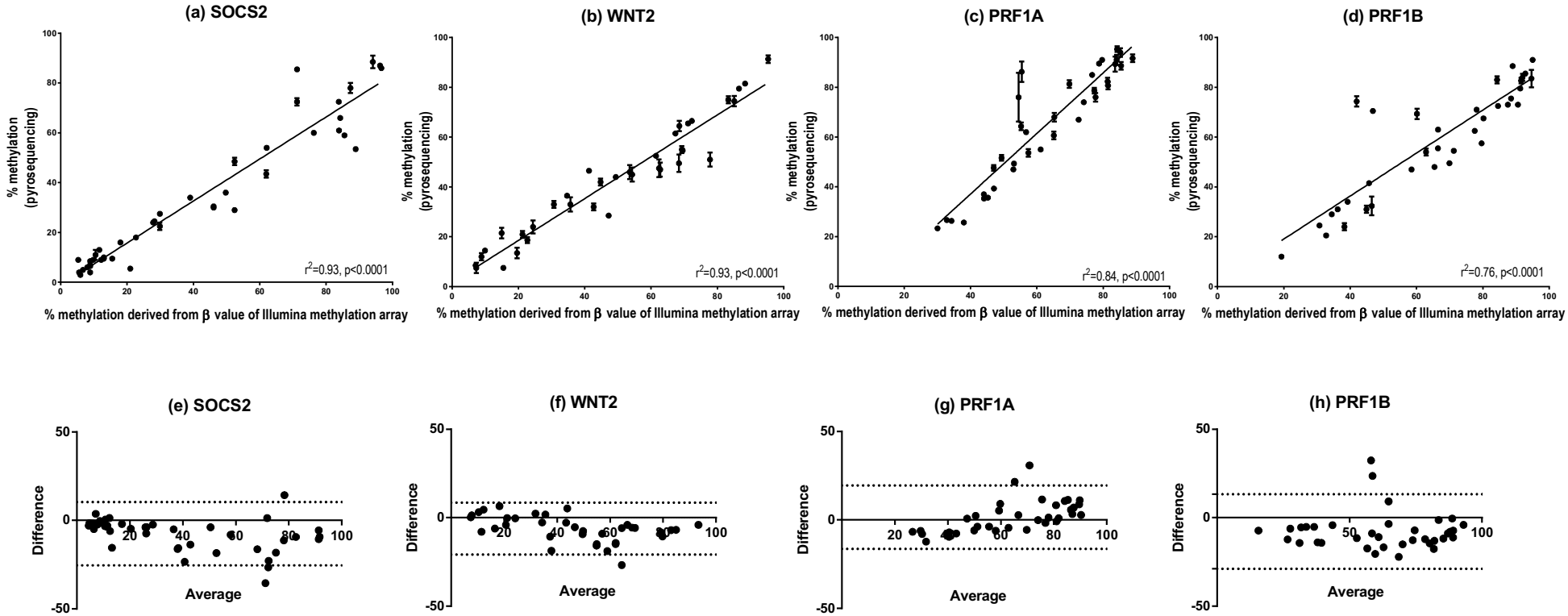

Figure S3. Unsupervised cluster analysis of the follow-up cohort. Each column represents a patient. Genotype is given in the upper panel. Samples were clustered based on their methylation levels at 157,797 variable probes and the heatmap in the middle panel shows the variable CpG probes located within CGIs. The latter were used to calculate the mean % CpG methylation shown in the lower panel; red and blue bars indicate a predominantly hyper- or hypo-methylated profile respectively.

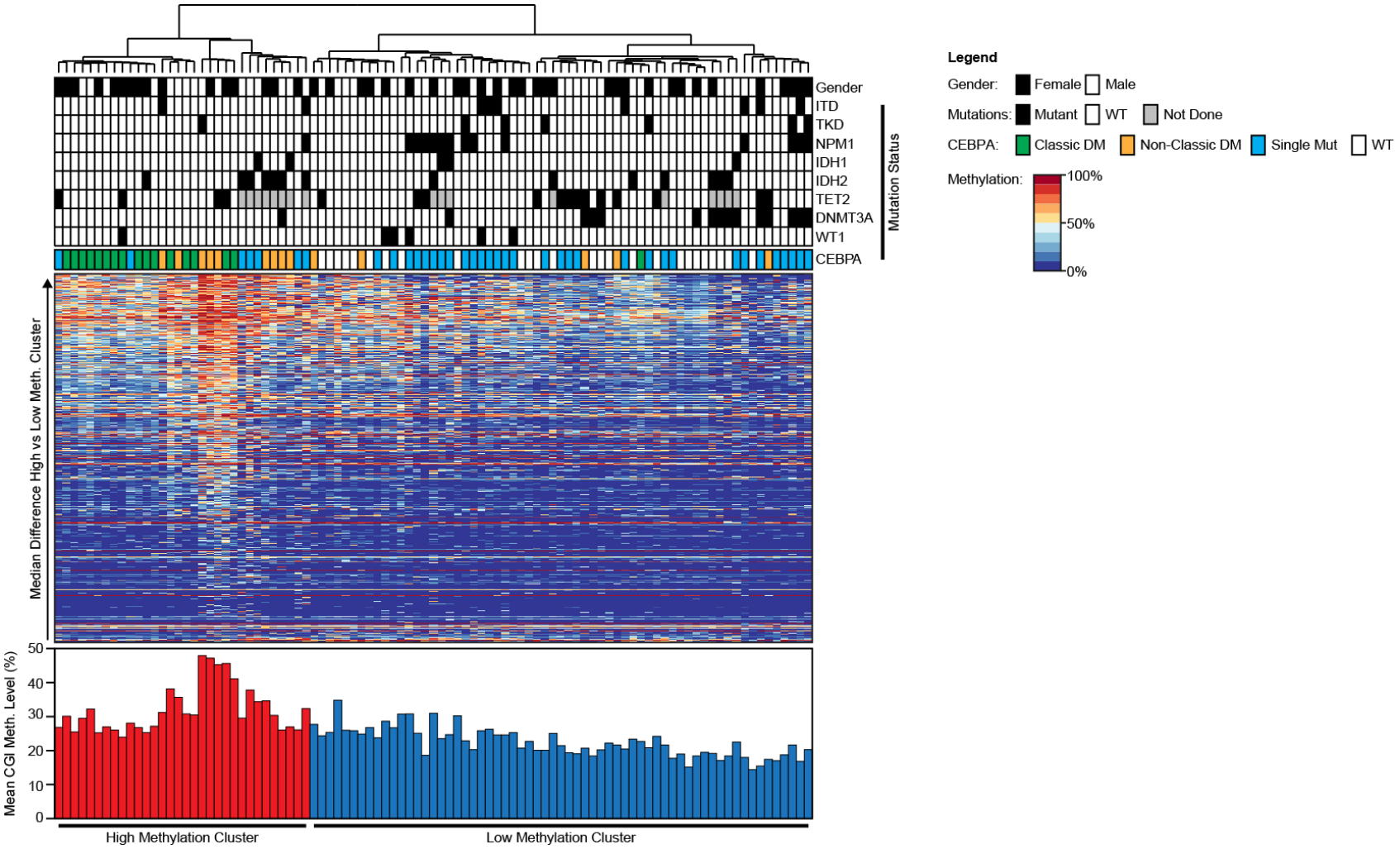

Figure S4. Methylation levels in other good-risk groups compared to the *CEBPA*$^{\text{Classic-DM}}$ cases at three differentially methylated CpG sites. The 24 *CEBPA*$^{\text{Classic-DM}}$ (excluding the outlier) and 56 *CEBPA*$^{\text{WT}}$ results were beta values from the arrays; results for the three comparative groups were obtained by pyrosequencing. Significance refers to difference from the *CEBPA*$^{\text{Classic-DM}}$ group (**P*≤0.05, ***P*≤0.01, ****P*≤0.001)