Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome

Running title: Heterogeneity of CEBPA mutations in AML

Bas J. Wouters¹, Bob Löwenberg¹, Claudia A.J. Erpelinck-Verschueren¹, Wim L.J. van Putten², Peter J.M. Valk¹, Ruud Delwel¹

Departments of ¹Hematology and ²Trials and Statistics, Erasmus University Medical Center, Rotterdam, The Netherlands

Correspondence:
Ruud Delwel
Erasmus University Medical Center
Department of Hematology, room Ee1342
P.O. Box 2040
3000 CA Rotterdam
The Netherlands
E-mail: h.delwel@erasmusmc.nl
Phone: +31-10-7043843
Fax: +31-10-7044745

Scientific heading: Myeloid Neoplasia
Abstract

Mutations in CCAAT/enhancer binding protein alpha (CEBPA) are seen in 5-14% of acute myeloid leukemia (AML) and have been associated with a favorable clinical outcome. Most AMLs with CEBPA mutations simultaneously carry two mutations (CEBPA\textsuperscript{double-mut}), usually biallelic, while single heterozygous mutations (CEBPA\textsuperscript{single-mut}) are less frequently seen. Using denaturing high performance liquid chromatography and nucleotide sequencing we identified among a cohort of 598 newly diagnosed AMLs a subset of 41 CEBPA mutant cases, i.e. 28 CEBPA\textsuperscript{double-mut} and 13 CEBPA\textsuperscript{single-mut} cases. CEBPA\textsuperscript{double-mut} associated with a unique gene expression profile as well as favorable overall and event-free survival, retained in multivariable analysis that included cytogenetic risk, FLT3-ITD and NPM1 mutation, white blood cell count and age. In contrast, CEBPA\textsuperscript{single-mut} AMLs did not express a discriminating signature and could not be distinguished from wild type cases as regards clinical outcome. These results demonstrate significant underlying heterogeneity within CEBPA mutation positive AML with prognostic relevance.
Introduction

Mutations in the transcription factor CCAAT/enhancer binding protein alpha (CEBPA) are found in 5-14% of acute myeloid leukemia (AML).\textsuperscript{1-9} CEBPA mutations have been associated with a relatively favorable outcome, and have therefore gained interest as a prognostic marker.\textsuperscript{4,6,10} While variable sequence variations have been described, two prototypical classes of mutations are most frequent. N-terminal mutations are located between the major translational start codon and a second ATG in the same open reading frame. These mutations introduce a premature stop of translation of the p42 CEBPA protein while preserving translation of a p30 isoform that has been reported to inhibit the function of full length protein.\textsuperscript{2} Mutations in the C-terminal basic leucine zipper (bZIP) region, in contrast, are in-frame, and may impair DNA binding and/or homo- and heterodimerization.\textsuperscript{8} The remaining mutations are mostly found between the N-terminus and bZIP region.\textsuperscript{11}

Most CEBPA mutant AMLs exhibit two mutations, which most frequently involves a combination of an N-terminal and a bZIP gene mutation.\textsuperscript{7,8,11,12} In AMLs with two CEBPA mutations, the mutations are typically on different alleles.\textsuperscript{11} Hence, in these cases no wild type CEBPA protein is expressed. A similar condition is found in AMLs carrying a homozygous CEBPA mutation.\textsuperscript{13} However, there are also AMLs that only show one single heterozygous mutation, and thus retain expression of a wild type allele.\textsuperscript{7,11,12}

To obtain better insight into the distribution of the various types of CEBPA mutations in \textit{de novo} adult AML and their impact on clinical outcome, we examined a cohort of 598 cases. Following denaturing high performance liquid chromatography (dHPLC) and nucleotide sequencing, we distinguished cases with two different
mutations or one homozygous mutation (further referred to as double mutations; $CEBPA^{\text{double-mut}}$) as well as cases with only one single heterozygous mutation ($CEBPA^{\text{single-mut}}$). Genome-wide gene expression profiling revealed that $CEBPA^{\text{double-mut}}$ AMLs expressed a highly characteristic signature, while $CEBPA^{\text{single-mut}}$ cases did not. In addition, favorable prognosis appeared uniquely associated with $CEBPA^{\text{double-mut}}$ AML.
Materials and methods

*AML samples, mRNA isolation, dHPLC analysis and nucleotide sequencing*

Bone marrow aspirates or peripheral blood samples of 598 cases of *de novo* AML were collected, blast cells were purified, and mRNA was isolated as reported. The entire *CEBPA* coding region was investigated by dHPLC and nucleotide sequencing. For details on patient characteristics and experimental procedures, see Supplementary Materials & Methods. All studies were approved by the Erasmus University Medical Center Institutional Review Board and patient informed consent was obtained in accordance with the Declaration of Helsinki.

*Statistical analysis*

Survival was estimated according to the method by Kaplan and Meier. The log rank test was used to assess statistical significance. Multivariable analysis was performed using Cox’s proportional hazards models. Definitions of outcome parameters and cytogenetic risk groups have been described. Further details are given in Supplementary Materials & Methods.

*Gene expression profiling analysis*

Gene expression profiles were obtained using Affymetrix (Santa Clara, CA) HGU133Plus2.0 GeneChips. Details on data processing and analysis are given in Supplementary Materials & Methods.
Results and discussion

In a cohort of 598 cases of adult de novo AML we identified 65 cases with an aberrant profile in at least one of the three investigated amplicons of the CEBPA coding sequence (Figure 1A-B). The presence of a CEBPA sequence variation was confirmed by nucleotide sequencing. Cases that only carried an insertion polymorphism or variation(s) that did not lead to amino acid changes were considered wild type. Two additional specimens were not considered in further analysis because they carried in-frame variations of unknown significance in the N-terminus (Table S1). As a result, 41/598 unambiguous CEBPA\textsuperscript{mut} AML cases (6.9%) were considered. These included 13 CEBPA\textsuperscript{single-mut} cases and 28 CEBPA\textsuperscript{double-mut} cases. Four of the CEBPA\textsuperscript{double-mut} cases carried homozygous mutations whereas the remaining 24 cases showed 2 heterozygous mutations (Table S1). Additional screening of the remaining 547 AML cases using a combination of agarose gel analysis and nucleotide sequencing as described\textsuperscript{6} did not reveal mutations that had been missed by dHPLC.

To investigate whether CEBPA mutations related to gene expression, we examined genome-wide gene expression data of 524 AML cases, that included 26 CEBPA\textsuperscript{double-mut} and 12 CEBPA\textsuperscript{single-mut} cases. Clinical and molecular characteristics of the AML cases are reported in Table S4-S5. Using Prediction Analysis for Microarrays (PAM)\textsuperscript{18} according to a supervised approach, we derived a 19-probe set signature predictive of CEBPA mutations (Figure 1C). This classifier showed a high specificity (99%), but a limited sensitivity (67%) in cross-validation, indicating a limited ability to recognize all CEBPA\textsuperscript{mut} specimens. Strikingly, misclassification was almost entirely due to CEBPA\textsuperscript{single-mut} cases, whereas CEBPA\textsuperscript{double-mut} AMLs were
predicted with an accuracy that was near perfect (Figure 1C, S1). In line with this, we were able to derive a specific 21-probe set classifier for \( CEBA^{\text{double-mut}} \) AMLs within the entire AML cohort with a cross-validated sensitivity of 100% (specificity 98%) (Table S3). In further support, unsupervised analysis of the expression data derived from the \( CEBA^{\text{mut}} \) subset indicated an underlying variability in gene expression that correlated with either double or single mutation status (Figure S2).

We next assessed how these differences between \( CEBA^{\text{double-mut}} \) and \( CEBA^{\text{single-mut}} \) related to clinical outcome. In line with previous data, overall survival and event-free survival were significantly better for \( CEBA^{\text{mut}} \) cases compared to cases with wild type \( CEBA \) (\( CEBA^{\text{wt}} \)) (Figure 1D and not shown). Separate analyses for the \( CEBA^{\text{double-mut}} \) and \( CEBA^{\text{single-mut}} \) subgroups, however, revealed a favorable outcome that was specific for \( CEBA^{\text{double-mut}} \) cases. We failed to find a favorable prognostic effect in relation to the \( CEBA^{\text{single-mut}} \) cases. In fact, \( CEBA^{\text{single-mut}} \) AMLs showed a significantly worse outcome than \( CEBA^{\text{double-mut}} \) cases, including a poor rate of complete remission (Figure 1E-F). These findings were also apparent in multivariable analysis (Table 1). When only patients less than 60 years of age or only patients with normal cytogenetics were considered, similar results were found, although in the latter subgroup with smaller numbers only the pair-wise comparison for overall survival between \( CEBA^{\text{double-mut}} \) and \( CEBA^{\text{single-mut}} \) reached statistical significance (Figure S3, Table S6).

Based on our previous analyses\(^6\) and based on the literature\(^11\) it is likely that in the vast majority of the \( CEBA^{\text{double-mut}} \) AML studied, both \( CEBA \) alleles were affected. A plausible hypothesis is therefore that absence of wild type \( CEBA \) mRNA is directly involved in the \( CEBA^{\text{double-mut}} \) gene expression profile. This may be further supported by our previous and current observations that indicate a high degree
of similarity between the profiles of $CEBPA^{\text{double-mut}}$ AML and a specific subgroup of leukemias characterized by epigenetic $CEBPA$ silencing (Figure S1).\textsuperscript{19} It is possible that analysis of larger patient series will lead to further refinement of this subclassification, for instance based on the location of the mutations. For example, our data indicated a tendency of $CEBPA^{\text{single-mut}}$ cases with mutations in the bZIP region to be potentially less distinct from the $CEBPA^{\text{double-mut}}$ AMLs (cases #7185, #7324, #2237; Figure 1C, S2). Of note, a subset of the $CEBPA^{\text{mut}}$ AMLs studied here was included in the cohort of 285 cases of AML that we previously investigated by gene expression profiling.\textsuperscript{14} In that study, all $CEBPA^{\text{double-mut}}$ AMLs were found in two particular clusters, while $CEBPA^{\text{single-mut}}$ AMLs did not specifically aggregate.\textsuperscript{14,19}

Studies to date have associated $CEBPA$ mutations with outcome\textsuperscript{4-6,9}, but have not applied subdivisions into single and double mutants. It is unclear why $CEBPA^{\text{double-mut}}$ AMLs would have a better outcome than those with a single heterozygous mutation. One explanation could be that a single mutant $CEBPA$ allele is not sufficient for leukemogenesis, and requires cooperating mutations which may be in $CEBPA$ itself or in other genes. Of note, recent data indicate that germ line $CEBPA$ mutations predispose to AML and the acquisition of a second, somatic $CEBPA$ mutation may then contribute to AML development.\textsuperscript{20} In fact, we found a tendency towards more $FLT3$-ITD, $FLT3$-TKD and $NPM1$ mutations in $CEBPA^{\text{single-mut}}$ compared to $CEBPA^{\text{double-mut}}$ cases (Table S5). Yet unknown abnormalities may associate with $CEBPA^{\text{single-mut}}$ AML as well and predispose to inferior outcome. It is however evident that these findings and their clinical significance warrant confirmation in independent cohorts of AML.

In summary, the data presented here indicate that $CEBPA^{\text{mut}}$ AML should at least be distinguished according to the presence of $CEBPA^{\text{double-mut}}$ and $CEBPA^{\text{single-mut}}$. 
mut. Screening using dHPLC, followed by nucleotide sequencing, appears useful for rapidly identifying mutant cases. In addition, gene expression based classification, for instance using the classifiers described here, enables the accurate identification of $CEBPA^{\text{double-mut}}$ AML cases.
Acknowledgements

This work was supported by grants from the Dutch Cancer Society “Koningin Wilhelmina Fonds” and the National Institutes of Health (NIH). We are indebted to Gert J. Ossenkoppele, M.D. (Free University Medical Center, Amsterdam, The Netherlands), Jaap Jan Zwaginga M.D. (Sanquin, The Netherlands), Edo Vellenga, M.D. (University Hospital, Groningen, The Netherlands), Leo F. Verdonck, M.D. (University Hospital, Utrecht, The Netherlands), Gregor Verhoef, M.D. (Hospital Gasthuisberg, Leuven, Belgium) and Matthias Theobald, M.D. (Johannes Gutenberg-University Hospital, Mainz, Germany) who provided AML samples. We thank Sonja van der Poel for help with dHPLC analysis. We also thank our colleagues from the bone marrow transplantation group and molecular diagnostics group in the department of Hematology of Erasmus University Medical Center for storage of samples and molecular analysis, respectively.

Authorship


Disclosure: B.L., P.J.M.V. and R.D. have declared ownership interests in Skyline, a spin-off company of Erasmus University Medical Center (ErasmusMC), held in a Special Purpose Foundation of ErasmusMC. The other authors have declared no competing financial interests.
References


Table 1. Multivariable analysis of $CEBA_{\text{double-mut}}$ and $CEBA_{\text{single-mut}}$ as prognostic markers for overall survival and event-free survival.

<table>
<thead>
<tr>
<th></th>
<th>Overall survival</th>
<th>Event-free survival</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (95% CI)</td>
<td>P</td>
</tr>
<tr>
<td>$CEBA_{\text{single-mut}}$†</td>
<td>1.18 (0.58 - 2.40)</td>
<td>0.65</td>
</tr>
<tr>
<td>$CEBA_{\text{double-mut}}$†</td>
<td>0.32 (0.17 - 0.61)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Intermediate‡</td>
<td>2.21 (1.52 - 3.22)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Poor‡</td>
<td>3.35 (2.27 - 4.94)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age [decades]</td>
<td>1.17 (1.08 - 1.28)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>WBC§</td>
<td>1.33 (1.05 - 1.68)</td>
<td>&lt;0.019</td>
</tr>
<tr>
<td>FLT3-ITD†</td>
<td>1.56 (1.20 - 2.03)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>NPM1†</td>
<td>0.55 (0.41 - 0.74)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Complete data for multivariable analysis were available for 511 cases.
HR indicates hazard ratio; CI, confidence interval; WBC, white blood cell count; FLT3, fms-related tyrosine kinase 3; ITD, internal tandem duplication; and, NPM1, nucleophosmin.

* P value < 0.05
† $CEBA$ status versus $CEBA^{wt}$
‡ Cytogenetic risk versus cytogenetic good risk
§ WBC higher than 20 x 10⁹/L versus lower than 20 x 10⁹/L
| FLT3-ITD versus no FLT3-ITD
| NPM1 mutation versus no NPM1 mutation
Figure legend

Figure 1. Schematic overview of dHPLC analysis, gene expression profiling analysis and survival estimates. A. Schematic representation of the CEBPA gene and location of amplicons a, b and c for PCR, used for dHPLC analysis. Functional regions are depicted, i.e. two transactivation domains (TAD1 and TAD2) in the N-terminal part, and the basic leucine zipper (bZIP) region in the C-terminal part. Nucleotide (nt) position is indicated relative to the main translation start site. Amino acid (aa) numbering and the alternative translation start site at position nt 358 (aa 120) are also depicted. B. Representative profiles of dHPLC analysis of one of the three investigated fragments, i.e. amplicons b, in a random selection of 90 samples. Heteroduplexes (various colors) are released earlier than homoduplexes (green), and can therefore be recognized as distinct peaks. Time is depicted on the x-axis, and absorbance on the y-axis. C. A gene expression prediction signature for CEBPAmut AML (irrespective of single or double mutant status) was derived in a data set of 524 AMLs, including 38 CEBPAmut cases. Prediction accuracy for each of the 38 CEBPAmut cases was estimated using repeated 10-fold cross-validation as detailed in Supplementary Materials and Methods. The proportion of correct predictions for the selected 38 CEBPAmut specimens is indicated (upper panel). Mutation status is color coded, i.e CEBPAsingle-mut (blue) or CEBPAdouble-mut (red). The heat map in the lower panel depicts the 19 probe sets in the resulting CEBPAmut gene expression classifier (see Table S2 for probe set information). Intensity values (log2) were mean centered over the cohort of 524 AML cases and for visualization purposes the genes were hierarchically clustered (Euclidian distance, average linkage). Cells represent relative log2 expression values, and have been color coded on a scale ranging from bright
green (-3) to bright red (+3), with black indicating no change relative to the mean. **D.** Kaplan Meier estimates of overall survival (OS) among $CEBPA^{\text{mut}}$ and $CEBPA^{\text{wt}}$ AML, log rank test $P=0.027$. **E.** OS among $CEBPA^{\text{double-mut}}$ versus $CEBPA^{\text{wt}}$ AML, $P=0.004$, and versus $CEBPA^{\text{single-mut}}$ AML, $P=0.005$; pooled $P=0.012$. **F.** Event-free survival (EFS) among $CEBPA^{\text{double-mut}}$ and $CEBPA^{\text{wt}}$ AML, $P=0.005$, and versus $CEBPA^{\text{single-mut}}$ AML, $P=0.004$; pooled $P=0.008$. The cumulative proportion of survival at the intercept, i.e. the point where a line crosses the Y-axis, reflects the proportion of patients reaching complete remission. Analyses similar to those depicted in panels D-F were performed after splitting the group of $CEBPA^{\text{wt}}$ AMLs into those with favorable cytogenetics and those with other cytogenetics. These additional analyses can be found in Figure S4.
Double *CEBPA* mutations, but not single *CEBPA* mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome

Bas J. Wouters, Bob Lowenberg, Claudia A.J. Erpelinck-Verschueren, Wim L.J. van Putten, Peter J.M. Valk and Ruud Delwel