GATA2 zinc finger 1 mutations associated with biallelic CEBPA mutations define a unique genetic entity of acute myeloid leukemia*

Philipp A Greif\textsuperscript{1,2,13}, Annika Dufour\textsuperscript{1,13}, Nikola P Konstandin\textsuperscript{1,2}, Bianka Ksienzyk\textsuperscript{1}, Evelyn Zellmeier\textsuperscript{1,2}, Belay Tizazu\textsuperscript{1,2}, Jutta Sturm\textsuperscript{1,2}, Tobias Benthaus\textsuperscript{1}, Tobias Herold\textsuperscript{1}, Marjan Yaghmaie\textsuperscript{1,2,12}, Petra Dörge\textsuperscript{3}, Karl-Peter Hopfner\textsuperscript{4}, Andreas Hauser\textsuperscript{5}, Alexander Graf\textsuperscript{6}, Stefan Krebs\textsuperscript{6}, Helmut Blum\textsuperscript{6}, Purvi M Kâkadıa\textsuperscript{1,12}, Stephanie Schneider\textsuperscript{1}, Eva Hoster\textsuperscript{1}, Friederike Schneider\textsuperscript{1}, Martin Stanulla\textsuperscript{3}, Jan Braess\textsuperscript{7}, Maria Cristina Sauerland\textsuperscript{8}, Wolfgang E Berdel\textsuperscript{9}, Thomas Büchner\textsuperscript{9}, Bernhard J Woermann\textsuperscript{10}, Wolfgang Hiddemann\textsuperscript{1,2}, Karsten Spiekermann\textsuperscript{1,2,14} and Stefan K Bohlander\textsuperscript{1,2,11,14}

(1) Department of Internal Medicine 3, Ludwig-Maximilians-Universität (LMU) Munich, Munich, Germany
(2) Clinical Cooperative Group ‘Leukemia’, Helmholtz Zentrum München, German Research Center for Environmental Health, Munich, Germany
(3) Department of Pediatrics, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany
(4) Department of Biochemistry, and
(5) Scientific Computing, and
(6) Laboratory for Functional Genome Analysis at the Gene Center, Ludwig-Maximilians-Universität (LMU) Munich, Munich, Germany
(7) Oncology and Hematology, St. John-of-God Hospital, Regensburg, Germany
(8) Institute of Biostatistics and Clinical Research, and
(9) Department of Medicine A - Hematology, Oncology and Pneumology, Universität Münster, Münster, Germany
(10) German Society of Hematology and Oncology, Berlin, Germany
(11) Center for Human Genetics, Philips Universität, Marburg, Germany
(12) Present addresses: P.M.K.: Center for Human Genetics, Philips Universität, Marburg, Germany; M.Y.: Hematology - Oncology and Stem Cell Transplantation Research Center, Tehran University of Medical Sciences, Tehran, Iran
(13) These authors contributed equally to this work
(14) These authors jointly supervised this work

To whom correspondence should be addressed:
Prof. Stefan K. Bohlander, MD
Clinical Cooperative Group ‘Leukemia’, Helmholtz Zentrum München, Marchioninistr. 25
81377 Munich Germany
+49 (0)89 7099 357 (phone)
+49 (0)89 7099 400 (fax)
e-mail: bohlander@helmholtz-muenchen.de

*Presented at the 53rd ASH-Annual Meeting, San Diego, Dec 11\textsuperscript{th} 2011 (abstract #72)
Abstract

Cytogenetically normal acute myeloid leukemia (CN-AML) with biallelic CEBPA gene mutations (biCEBPA) represents a distinct disease entity with a favorable clinical outcome. So far, it is not known if other genetic alterations cooperate with biCEBPA mutations during leukemogenesis. To identify additional mutations, we performed whole exome sequencing of five biCEBPA patients and detected somatic GATA2 zinc finger 1 (ZF1) mutations in 2 out of 5 cases. Both GATA2 and CEBPA are transcription factors crucial for hematopoietic development. Inherited or acquired mutations in both genes have been associated with leukemogenesis. Further mutational screening detected novel GATA2 ZF1 mutations in 13 of 33 biCEBPA positive CN-AML patients (13/33: 39.4%). No GATA2 mutations were found in 38 CN-AML patients with a monoallelic CEBPA mutation and in 89 CN-AML patients with wild-type CEBPA status. In reporter gene assays, all tested GATA2 ZF1 mutants showed reduced capacity to enhance CEBPA-mediated activation of transcription, suggesting that the GATA2 ZF1 mutations may collaborate with biCEPBA mutations to deregulate target genes during malignant transformation. We thus provide evidence for a genetically distinct subgroup of CN-AML. The AMLCG trials 1999 and 2008 are registered with the identifiers NCT00266136 and NCT01382147 at www.clinicaltrials.gov.

Introduction

In 2008, the WHO classification of Tumours of the Haematopoietic and Lymphoid Tissues included “AML with mutated CEBPA” as a provisional favorable prognostic entity and a routine CEBPA mutation screening was recommended for all AML patients with no detectable chromosomal abnormalities.\textsuperscript{1,2} However, only patients with mutations in both CEBPA alleles (biCEBPA mutated AML) have a prognostically
favorable outcome.\textsuperscript{3-7} Bi\textit{CEBPA} patients typically have a combination of an N-terminal frameshift mutation leading to a 30 kDa dominant negative \textit{CEBPA} isoform and a C-terminal in-frame mutation in the bZIP region, disrupting dimerization and DNA binding activities of \textit{CEBPA}. In murine bone-marrow transplantation models, distinct but collaborative roles of both types of \textit{CEBPA} mutations have been detected.\textsuperscript{8} Furthermore, a combination of an N- and C-terminal disruption of the \textit{CEBPA} gene synergistically resulted in fast and efficient development of leukemia in mice.\textsuperscript{8-10} These experiments suggest that a biallelic disruption of \textit{CEBPA} might be responsible for both the differentiation block and for enhanced proliferation of progenitor cells and thus be sufficient for leukemogenesis. Indeed, bi\textit{CEBPA} mutant AML represents as a very homogeneous AML subgroup with a distinct immunophenotype\textsuperscript{11} and a characteristic gene expression profile.\textsuperscript{3,6,7,12} Furthermore, bi\textit{CEBPA} mutant AML patients rarely have mutations in genes which are frequently mutated in CN-AML like the nucleophosmin (\textit{NPM1}) gene, internal tandem duplications (ITD) of the \textit{FLT3} gene, partial tandem duplications (PTD) of the \textit{MLL} gene (\textit{MLL-PTD}),\textsuperscript{3-5} mutations in the \textit{TET2}, isocitrate dehydrogenase (\textit{IDH})\textsubscript{1} and 2 or \textit{DNMT3A} genes.\textsuperscript{13-15} In order to identify potentially cooperating mutations associated with bi\textit{CEBPA} mutated AML, we performed whole exome sequencing. Thereby, we discovered a high frequency of N-terminal zinc finger (ZF1) mutations of \textit{GATA2} in patients with bi\textit{CEBPA} mutated AML. \textit{GATA2} is a zinc finger transcription factor important for hematopoietic stem cell proliferation\textsuperscript{16-19} and normal megakaryocytic development.\textsuperscript{20} Somatic mutations affecting the C-terminal zinc finger (ZF2) of \textit{GATA2} are associated with the progression of chronic myeloid leukemia (CML),\textsuperscript{21} whereas hereditary \textit{GATA2} ZF2 mutations predispose to AML and myelodysplastic syndrome (MDS).\textsuperscript{22} Mutations targeting either of the two zinc fingers were described
as a rare event in the M5 subtype of AML.\textsuperscript{23} In the present study we report a unique association of \textit{GATA2 ZF1} mutations and bi\textit{CEBPA} mutations in AML.

\section*{Methods}

\subsection*{Patients}

In this analysis we included diagnostic bone-marrow or peripheral blood samples from 160 adult AML patients with a normal karyotype of which 146 AML patients were enrolled in the German AML cooperative group (AMLCG) 1999 multicenter treatment trial. These patients have also been investigated in a previous publication with a different objective.\textsuperscript{3} Four patients with bi\textit{CEBPA} mutations were treated within the AMLCG 2008 study. Sample selection for exome sequencing was dependent on availability of remission samples. Available clinical characteristics were age, sex, \textit{de novo} versus secondary AML, French-American-British (FAB) subtype, white blood cell and platelet counts, the amount of bone marrow blasts, hemoglobin and LDH levels. All patients treated within the AMLCG 1999 trial received intensive induction therapy with either thioguanine, cytarabine and daunorubicine (TAD) as standard therapy or high-dose cytarabine and mitoxantrone (HAM) in the experimental arm followed by one course of HAM and consolidation therapy. In patients with an age of $\geq 60$ years, a second induction was only administered in case of inadequate response to initial intensive induction treatment. Details of the trial protocols have been published elsewhere.\textsuperscript{24} Patients treated within the AMLCG 2008 trial received dose-dense sequential HAM versus age dependent standard double induction and TAD-9 consolidation.\textsuperscript{25} The study protocols were approved by the ethics committees of the participating centers, and all patients provided written informed consent to the
scientific use of surplus samples (e.g. bone marrow, peripheral blood) in accordance with the Declaration of Helsinki.

Sample preparation and high throughput sequencing

Genomic DNA was extracted from patients’ bone marrow or peripheral blood samples using Qiacube technology (Qiagen, Hilden, Germany). 5µg of genomic DNA were fragmented to an average size of 150 bp using the bioruptor sonicator (Diagenode). Ultra-sound was applied during 3 cycles of 15 min each at low power as described before. Sequencing libraries were prepared using DNA sample prep reagent set 1 (NEBNext). In brief, library preparation included end repair, adapter ligation and PCR-enrichment. Exon-coding sequences were then captured using SureSelect human all exon 50M Kit (Agilent) according to the manufacturer’s instructions. Exome libraries were then sequenced by performing 76 bp paired-end reads on the Genome Analyzer IIX platform (Illumina).

Sequence alignment

Short read alignment to the human genome assembly (build NCBI36/hg18) was performed using the BWA sequence alignment program with the default parameters. For each sample, we generated at least 90,000,000 paired-end reads of 76 bp length (7 Gbp), of which at least 84,000,000 (6.4 Gbp) could be aligned to the reference sequence (supplemental Table S1). Read mapping, subsequent assembly and variant calling were performed using the resequencing software packages BWA and SAMtools. During alignment, apparently duplicated reads were removed. Global exon coverage and histograms (supplemental Table S1 and supplemental FigureS1) were calculated using the BEDtools software.
Variant detection

Variant calling was performed using the VarScan software.\textsuperscript{31} For the AML samples we excluded all positions with a read depth lower than 10 and required (i) each putative variant to have a median quality value of the variant bases of at least 15; (ii) that at least 20\% of all reads covering the position show the variant allele with an absolute minimum number of 5 variant reads and (iii) that reads showing the variant allele are from both strands. For the remission samples we included all potential variants with a minimum frequency of 5\% regardless of read depth, base quality, absolute number of variant reads or number of affected strands. Analysis of the effect of the variants on the protein level was performed with VarScan scripts using datasets from Agilent, Ensembl and the Santa Cruz Genome Browser as described before.\textsuperscript{32,33} Variants were annotated using a Variant Effect Predictor perl script.\textsuperscript{34} Alignment inspection was performed using the Integrative Genomics Viewer.\textsuperscript{35}

Sanger Sequencing

Candidate non-synonymous somatic variants (present in the AML but not in the remission sample) were verified by bidirectional DNA sequencing using ABI 3100 Avant technology of the corresponding gene after PCR amplification using exon-spanning primers.

High resolution melting curve analysis

To verify the frequency of selected gene mutations in a larger cohort of biCEBPA mutated patients (n=33), we scanned all coding exons 1-5 of the \textit{GATA2} gene, and all coding exons 1-7 of the \textit{IKZF1} gene and exons 15-23 of the \textit{DNMT3A} gene for sequence variations using high resolution melting analysis. Due to limited sample
availability, we confined the DNMT3A mutation screening to exons 15-23 in which most mutations have been detected in AML so far.\textsuperscript{13,36,37} Exon-spanning primers were designed to generate amplicons of up to 300bp length. Approximately 15-30 ng of genomic DNA was amplified using the LightCycler High Resolution Melting master (Roche Diagnostics, Penzberg, Germany) following the manufacturer’s instructions. PCR products were directly examined on a 1.5% Gel Red (Biotium, distributed by Biotrend, Köln, Germany) agarose gel. Differences in the high resolution melting profile in comparison to the wildtype genes in HL-60 cells were confirmed by bidirectional DNA sequencing.

**Multiplex Ligation-dependent Probe Amplification (MLPA) analysis**

MLPA analysis was performed using the MLPA SALSA kit P335-A4 ALL-IKZF1 (MRC-Holland, Amsterdam, The Netherlands) according to the manufacturers’ protocol. The assay includes probes for each of the eight exons of the IKZF1 gene and is therefore able to detect deletions of the whole gene as well as all types of focal intragenic deletions in exons. Separation of MLPA products performed on an Applied Biosystems 3730 DNA Analyzer with GeneScan\textsuperscript{TM} 500 LIZ\textsuperscript{®} as internal size standard (both Applied Biosystems, Carlsbad, CA, USA). Data were analyzed using Peak Scanner v1.0 and Coffalyser v9.4 software. Relative copy number was calculated after intra-sample normalization against control fragments and inter-sample normalization against control samples (healthy blood donor DNA). A ratio of 1±0.3 represents a normal copy number of 2, ratios <0.7 and <0.3 represent heterozygous and homozygous deletions, respectively, while ratios >1.3 indicate amplifications.
Plasmids

The pFLAG-CMV2-GATA2 construct was purchased from Addgene (Cat# E7398). For the following GATA2 point mutants (A318T, G320D, L321V, R293Q, V296G, L359V) a 560bp fragment was synthesized by GeneArt (Regensburg) and inserted into pFLAG-CMV2-GATA2 using restriction sites SacII und HpaI. The pcDNA3-CEBPA expression plasmids (wild type and p30) and the 4xCAAT-binding-site-TK-LUC reporter are a gift from Daniel G. Tenen (Boston). The GATA-LUC reporter was purchased from Panomics (Cat# LR0027).

Immunoblotting analysis

Cellular lysates from HEK293T cells were electrophoresed on 10-12 % SDS-PAGE and transferred to PVDF membrane (Hybond P™, Amersham Pharmacia biotech). The membranes were blocked for 30 min with 5 % nonfat dried milk at room temperature and probed with polyclonal rabbit GATA2 antibodies (H116; sc-9008, Santa Cruz biotechnology) or mouse monoclonal β-actin antibodies (Clone AC-15; Cat# A5441, Sigma-Aldrich), followed by secondary antibodies conjugated to horseradish peroxidase. Proteins were detected with enhanced chemiluminescence (ECL™, Amersham, GE Health Care). Quantification of Western Blot signals was performed using ImageJ software.

Immunostaining and confocal laser scanning fluorescence microscopy

For intracellular localization studies, HEK 293T cells were grown on coverslips and co-transfected with FLAG-GATA wt or FLAG-GATA2 mutants. After 24 hours, cells were fixed with PBS 2% paraformaldehyde for 10 min, permeabilized with PBS 0.1%
Triton X for 10 min and blocked with PBS 10% FCS for 1 hour. Coverslips were incubated with monoclonal mouse FLAG antibodies (M2, Sigma-Aldrich) over night at 4° C. Following extensive washing with PBS, Alexa 555 conjugated secondary antibodies were added for 1h at room temperature. After further washing steps, cells were stained with DAPI and mounted using fluorescent mounting medium (DAKO). Finally, immunostained species were analyzed in a confocal fluorescence laser scanning system (TCS-SP2 scanning system and DM IRB inverted microscope, Leica, Solms, Germany).

**Reporter gene assays**

HEK 293T cells were seeded in 24 well plates and co-transfected with 100 ng of luciferase reporter plasmid, 10 ng of pRL co-reporter vector (Promega), variable amounts of pFLAG-GATA2 wt, pFLAG-GATA2 mutants or pcDNA3-CEBPA. The total amount of transfected DNA in each well was kept constant by adding pcDNA3 empty vector. After 24 h, cells were harvested and assayed for firefly and Renilla luciferase activities using a Dual-Luciferase Reporter Assay System (Promega). Measurements of Renilla luciferase activity were used for normalization. Experiments were performed in triplicates and repeated at least three times. All results are reported as mean ± standard error of the mean. Pairwise comparisons were performed using a Student’s t-test.

**Statistical analyses**

Comparisons of different categoric parameters according to the GATA2 mutation status were performed using the χ²-test. Fisher’s exact test was used instead of the χ²-test when at least one cell had an expected frequency of less than 5. For
continuous variables, we used the Mann Whitney U test. Outcome variables were overall survival (OS) and event-free survival (EFS). OS was calculated from randomization to death from any cause or to the latest follow-up date. EFS was defined as the period from the start of therapy until lack of a complete remission (CR), relapse after CR or death without relapse. Patients treated within the AMLCG 2008 study were excluded from the survival analysis due to differences in treatment protocols and a short follow-up up time. Survival curves were calculated according to the Kaplan-Meier method and compared with the log rank test. All tests were two-tailed and \( p \)-values <0.05 were considered statistically significant.

Statistical analyses were performed with SPSS version 16.0 (SPSS Inc. Chicago, ILL USA) and the R 2.7.2 software package (R Foundation for Statistical Computing, Vienna, Austria).

**Structural model of GATA2**

The model is based on the crystal structure of the highly homologous GATA3\(^{38}\) (PDB accession 3DVF). The backbone was left unaltered. All displayed side chains that harbor mutations except one are conserved between GATA2 and GATA3 and were left unaltered. The remaining side chain was modeled with COOT\(^{39}\) using the most common rotamer.

**Results**

**Exome sequencing of 5 bi\(\text{CEBPA}\) AML cases**

To identify collaborating mutations, we sequenced the exomes (protein coding regions) of five bi\(\text{CEBPA}\) mutated CN-AML leukemia samples and the corresponding
remission samples representing the patients’ germlines. We generated a minimum of 7 Gbp of sequence for each exome. This allowed us to cover at least 80% of RefSeq coding exon positions with a minimum read depth of 10 (supplemental Table S1 and supplemental Figure S1). Comparison of the AML exome sequence with the remission exome sequence and exclusion of annotated polymorphisms led to the identification of leukemia-specific variants. Between 2 and 8 non-synonymous coding somatic mutations per patient were confirmed using Sanger sequencing (Table 1). Thus, we detected tumor-specific mutations (nonsense and missense) in a total of 21 genes. \textit{DNMT3A} and \textit{GATA2} were found to be mutated in 2 of the 5 bi\textit{CEBPA} CN-AML samples. \textit{DNMT3A} encodes a DNA-methyltransferase with an overall mutation frequency of about 20% in AML.

The \textit{IKZF1} G158S mutation in patient #1 has previously been identified in acute lymphoblastic leukemia (ALL) and was functionally characterized as dominant negative.\textsuperscript{40,41} Deletions of \textit{IKZF1} are frequent events in acute lymphocytic leukemia and also occur at the progression of chronic myeloid leukemia to lymphoid blast crisis.\textsuperscript{41,42} Thus, three key regulators of hematopoiesis were targeted by somatic mutations in patient #1, namely \textit{CEBPA}, \textit{GATA2} and \textit{IKZF1} (\textit{Ikaros}) (Figure 1 a, Table 1). In patient #2 we found a mutation of \textit{STAG2} (Table 1) which is part of the cohesin complex. Alterations of the cohesin genes were previously reported in myeloid diseases.\textsuperscript{43} \textit{KRAS} mutations (patient #1) occur in 5% of AML patients and are mainly associated with FAB M4.\textsuperscript{44}

**High frequency of \textit{GATA2} ZF1 mutations in bi\textit{CEBPA} AML**

Based on the exome sequencing results, \textit{GATA2}, \textit{DNMT3A} and \textit{IKZF1} were selected for mutational screening in 33 bi\textit{CEBPA} mutated AML patients. Strikingly, we detected 15 heterozygous missense mutations in coding exon 4 of \textit{GATA2} in 13 out
of 33 biCEBPA patients (39.4%). Two patients were found to carry two different mutations in GATA2. The GATA2 mutations affected 7 different amino acids positions (Figure 1 b). All mutations were in the highly conserved N-terminal zinc finger domain (ZF1-domain) of GATA2 (Figure 1 b). A mutational hotspot within the ZF1 domain between amino acids A317 and G321 surrounding cysteine 319 accounted for 9 of the 15 mutations. This is a highly conserved region in different GATA factors and among different species (supplemental Figure S2). The missense mutations A318T and G320D were detected in 6 out of 13 GATA2 mutated biCEBPA patients (3 with A318T and 3 with G320D). 4 out of 13 biCEBPA patients with GATA2 mutations were also analyzed during remission. All 4 patients had lost the GATA2 mutation at this point (Figure 1 a and data not shown). Furthermore, no GATA2 mutations were found in 38 patients with a monoallelic CEBPA mutation (moCEBPA) and in 89 patients with wild-type CEBPA. The association between GATA2 ZF1 mutations and biCEBPA mutations in a total cohort of 160 screened CN-AML patients was highly significant (p<0.001; Chi-square test).

High resolution melting curve analysis of all exons of IKZF1 detected only one additional patient with a non-synonymous substitution in exon 4 (Pat #21, supplemental Table S2). Unfortunately, no remission sample from this patient was available. No IKZF1 sequence variations were identified in 38 moCEBPA patients. Whole gene and intragenic deletions of IKZF1 were ruled out by performing Multiplex Ligation-dependent Probe Amplification (MLPA) in 30 out of 33 biCEBPA patient samples. We detected an additional mutation in DNMT3A (Table 1) resulting in an overall DNMT3A mutation frequency of 9.1% (3/33) in biCEBPA patients. The frequency distribution of additional mutations in biCEBPA mutant AML is shown in Figure 1 c. Interestingly, FLT3-ITD (6/33) was mutually exclusive with GATA2 mutations in biCEBPA mutant AML (P=0.029, Fisher’s Exact Test).
GATA2 ZF1 mutants show reduced capacity to enhance CEBPA-dependent activation of transcription

To study the functional consequences of the GATA2 ZF1 mutations, we performed reporter gene assays in HEK293T cells. Western blotting confirmed that the GATA2 mutants were expressed at similar levels (Figure 2 a). All mutants showed nuclear localization like the wild-type GATA2 protein (Figure 2 b and supplemental Figure S3). Upon co-transfection of a GATA-responsive reporter together with expression plasmids for GATA2 wild-type or GATA2 mutants a reduced activation was observed for mutants A318T and L321F. In contrast, an increased activation was observed for the G320D mutant (Figure 3 a). As a control we included the L359V mutant of GATA2, located in ZF2, which is found in about 10% of patients at the progression of chronic myeloid leukemia to blast crisis. The L359V mutant did not show a significant difference in transcriptional activation compared to wild type GATA2. GATA2 and CEBPA are protein-protein interactors. In co-immunoprecipitation experiments the GATA2 mutants A318T, G320D, L321F and L359V were still able to interact with wildtype CEBPA (supplemental Figure S5). To further analyze the role of GATA2 mutations in the context of this interaction, we used a CEBPA-responsive luciferase reporter (4xCAAT-binding-site-TK-LUC) together with expression plasmids for wild type CEBPA or the truncated CEBPA p30 mutant and GATA2 wild-type or GATA2 mutants. CEBPA-dependent activation was enhanced by co-expression of GATA2 wild type. Interestingly, this enhancement was significantly reduced for all the GATA2 ZF1 mutants identified in biCEBPA CN-AML, but not for the CML-associated L359V mutant (Figure 3 b). Residual activation of the reporter by the p30 CEBPA mutant could also be enhanced by co-expression of wild-type GATA2, but all GATA2 mutants tested showed a markedly reduced enhancement (Figure 3 b).
**GATA2 ZF1 mutations and clinical characteristics**

There were no significant differences in the clinical parameters (WBC, age, sex etc.) between biCEBPA mutated patients with (n=13) and without (n=20) mutations in GATA2 (supplemental Table S3).

In Kaplan Meier survival analysis, the presence of GATA2 mutations (n=10) did not negatively impact on the favorable overall survival (OS) and event free survival (EFS) of 26 biCEBPA patients (OS: \( P=0.274 \); EFS: \( P=0.185 \)) (Figure 4).

In a three group comparison there was a trend towards a better OS (\( P=0.012 \)) and EFS (\( P=0.172 \)) for biCEBPA patients with GATA2 mutations (n=10) compared with biCEBPA patients without GATA2 mutation (n=16) and 117 patients with wildtype GATA2 status which included 89 CN-AML patients with wildtype CEBPA and 28 patients with monoallelic CEBPA mutations (Figure 4).

**Structural modelling of the GATA2 ZF1 mutants**

To gain further insights into the consequences of the GATA ZF1 mutations on the protein level we performed structural modelling. Based on the homology model (Figure 5), amino acids N317, A318, L321 and R330 are directly implicated in DNA binding and mutations in these residues likely alter DNA affinity. G320 is important for proper attachment of an adjacent beta-hairpin loop that provides additional DNA binding contacts. Q328 is not directly involved in DNA binding. However, the Q328P mutation could either affect the backbone fold of ZF1 and indirectly alter DNA binding by perturbing the adjacent R330, or alternatively be involved in interactions with other domains such as ZF2.
CEBPA target genes are enriched in biCEBPA mutated patients with concurrent GATA2 mutations.

To assess the impact of GATA2 mutations on gene expression we performed microarray analysis. Nine biCEBPA mutated CN-AML patient samples were analyzed. Samples were selected based on availability. Four patients had an additional GATA2 mutation (44.4 %). No patient had an additional NPM1 mutation or an MLL-PTD. One patient without GATA2 mutation had an additional FLT3-ITD (11.1%). Gene set enrichment analysis (GSEA) was performed to assess significant changes in gene expression levels. We were able to identify 18 pathways significantly enriched in the GATA2-mutated subgroup (False Discovery Rate (FDR) < 25% and Nominal p-value < .05). These belonged e.g. to the cell cycle, ERBB, MTOR, p53 signaling and apoptosis pathway (supplemental table S4). We then focused on validated CEBPA and GATA2 target genes (supplemental Table S5). Very interestingly, CEBPA target genes were significantly enriched in the GATA2-mutated subgroup (Nominal p-value=.003; FDR=1.7%; Normalized Enrichment Score (NES)=1.62; supplemental Figure S4). Genes enriched in the GATA2 mutated subgroup were e. g. MYC, JUNB, FOS and CEBPB (supplemental Table S6, supplemental Figure S4). GATA2 target genes showed no significant enrichment.

Discussion

Ours is the first report of recurrent GATA2 mutations associated exclusively with biCEBPA mutations in AML.

The specific occurrence of GATA2 mutations in the biCEBPA mutant AML subgroup, which accounts for approximately 4% of AML, might explain why GATA2 mutations...
remained undetected in a previous full-length GATA2 mutation screen of 50 unselected AML patients.\textsuperscript{46} Other studies limited GATA2 mutational screening to GATA2 exon 5 \textsuperscript{22} or to specific GATA2 mutations, e.g. L359V in GATA2 ZF2 \textsuperscript{21,22,46} and thus would have missed GATA2 ZF1 mutations located in exon 4. GATA2 ZF1 and 2 mutations were recently detected at a low frequency of 3.6\% (4/112) in the M5 subtype of AML.\textsuperscript{23} However, the CEBPA status was not reported in this study. We found that biCEBPA mutations are significantly associated with AML M1 or M2 subtype and rarely occur in the AML FAB M5 subtype.\textsuperscript{3}

Notably, all mutations found in the present study affected exon 4 encoding the N-terminal zinc finger domain (ZF1) of GATA2, whereas recently GATA2 germline mutations were reported in the C-terminal zinc finger (ZF2) in familial cases of myelodysplastic syndrome (MDS) and AML.\textsuperscript{22,47-49} The ZF2 domain of GATA2 is also targeted by somatic mutations in blast crisis of CML.\textsuperscript{21,46} While the GATA2 ZF2 mutations in CML are associated with poor prognosis,\textsuperscript{21} the GATA2 ZF1 mutants found in the present study do not appear to have a negative impact on clinical outcome (Figure 4). Analysis of a larger cohort of biCEBPA mutated patients will elucidate whether concurrent GATA2 and biCEBPA mutations define a prognostically distinct subgroup of AML.

The exclusiveness of GATA2 mutations and FLT3-ITD within biCEBPA mutated patients (Figure 1 c) suggests alternative mechanisms of leukemogenesis in these genetic subgroups. A potential pathogenetic relevance of FLT3-ITD in CEBPA mutated AML has been suggested in a bone-marrow transplantation model in which a C-terminal CEBPA mutant collaborated with a FLT3-ITD in the process of leukemogenesis.\textsuperscript{8} Considering the direct protein-protein interaction between GATA2 and CEBPA,\textsuperscript{45} and their prominent expression in early hematopoietic progenitor
mutations of both genes may affect the same protein complex in a synergistic manner during hematopoietic differentiation leading to leukemogenesis. We show that the GATA2 ZF1 mutants found in our patients have a reduced capacity to cooperate with wild type CEBPA to activate transcription (Figure 3 b). This is in contrast to the behavior of the GATA2 ZF2 mutant L359V found in CML blast crisis, which exhibits gain of function properties.\textsuperscript{21}

The ZF1 domain of GATA2 is known to contribute to the stabilization and specificity of DNA-binding and mediates the interaction with the transcriptional co-factor Friend of GATA2 (FOG1),\textsuperscript{51} whereas the interaction with CEBPA was mapped to the C-terminus of GATA2.\textsuperscript{45} Thus, the novel GATA2 ZF1 mutants likely influence DNA-binding properties (Figure 5) or protein-protein interactions of GATA2. In co-immunoprecipitation experiments all GATA2 mutants tested were able to interact with CEBPA (supplemental Figure S5). However, we cannot exclude that the GATA2 mutations may result in subtle differences in the binding affinity between GATA2 and CEBPA that might be very challenging to quantify. The structural model (Figure 5) may provide an explanation for the differences observed in the activation of the GATA-LUC reporter by GATA ZF1 mutants (Figure 3A): The amino acids affected by those mutants which show reduced activation (A318T and L321F) appear to be more directly involved in DNA-binding whereas the mutant G320D showing increased activation might alter the conformation of ZF1 in a way which is favorable for DNA-binding or transactivation. Additional experiments will be necessary to analyze the impact of GATA2 ZF1 mutants on DNA-binding.

The gene expression profiling (GEP) results have to be interpreted carefully due to the limited sample size. Considering the high degree of homogeneity within the analyzed groups (Nine patients with normal karyotype and biCEBPA mutations...
including four patients with additional GATA2 mutations and only one patient with an additional FLT3-ITD) the results show distinct differences between the GATA2 mutated and GATA2 wild type groups. These differences affect many well known oncogenic pathways in AML. Very interestingly CEBPA target genes showed significant enrichment in GSEA, whereas GATA2 target genes showed no significant enrichment. Interpreting our GEP analysis in the context of our reporter gene assays, it seems that GATA2 mutants are able to modulate the effect of the mutated CEBPA transcription factor, rather than affecting GATA2 function and GATA2 target genes.

It is intriguing to speculate that the ZF1 mutations of GATA2 may further reduce the residual activity of CEBPA p30 (Figure 3 b) which results from the N-terminal nonsense mutations in biCEBPA AML. This would support the recent hypothesis that a "dosage window" of nuclear factors is critical for the process of hematopoietic differentiation and leukemogenesis. For example, the complete absence of PU.1 does not lead to the development of leukemia in knock-out mice whereas lowering the levels of PU.1 to 20% of normal levels will result in leukemia.52

In summary, we describe the specific association of biallelic CEBPA mutations in cytogenetically normal AML with novel mutations in the N-terminal zinc finger of GATA2. In contrast to the results of most other high throughput sequencing studies in AML, which have painted a picture of increasing genetic complexity, our results suggest that there are indeed striking associations of defined mutations in subgroups of AML. The specific association of mutations affecting two interacting regulators of hematopoiesis introduces a novel concept for leukemogenesis: The simultaneous mutational targeting of two transcription factors that function in the same differentiation pathway in AML. Studies like ours might help to eventually define the various pathways that are critical to AML leukemogenesis.
Acknowledgements

This work was supported by a Deutsche Krebshilfe grant (109031) to P.A.G. and S.K.B. and a grant from the BMBF (NGFN Plus, PKL-01-GS0876-6) to S.K.B.

P.A.G., K.P.H., K.S. and S.K.B. acknowledge support by the DFG (SFB 684 Molecular Mechanisms of Normal and Malignant Hematopoiesis). We thank the participating centers of the AMLCG clinical trial and Klaus H. Metzeler for support with the gene expression analysis.

Authorship

References


Figure legends

Figure 1 Additional genetic aberrations identified in biCEBPA mutant CN-AML

(a) Somatic mutations affected three key regulators of hematopoiesis, namely CEBPA, GATA2 and Ikaros (IKZF1) in one CN-AML patient analyzed by exome sequencing (see also Table 1). Chromatograms are shown for both diagnostic AML samples and corresponding follow-up samples at complete remission from the same patients. (b) Structure of the human GATA2 gene with five coding exons is shown in black and non-coding exons in grey (upper panel). The GATA2 protein structure includes the N- and C-terminal zinc finger domains (ZF1, ZF2) and the nuclear localization signal (NLS) (middle panel). The amino acid sequence of the N-terminal ZF1 domain of GATA2 is detailed with cysteine residues that are responsible for zinc binding are shown in bold (bottom panel). 15 heterozygous missense mutations were detected in coding exon 4 of GATA2 in 13 out of 33 biCEBPA patients (overall frequency: 39.4%). All GATA2 mutations clustered in the highly conserved N-terminal zinc finger domain (ZF1-domain) of GATA2. Positions A318, G320 and G321 were recurrently affected. The ZF1 is also illustrated as polypeptide chain with the amino acids (circles) targeted by mutations in black (bottom panel).

(c) Frequency distribution of additional genetic aberrations in 33 biCEBPA patients. Each box indicates one patient. Dark grey boxes are indicative for patients who are positive for the respective mutation; light grey boxes indicate wild type status. Missing information is shown as a white space. All 33 biCEBPA mutated patients were NPM1 wildtype and did not carry an additional FLT3-TKD or MLL-PTD.

Abbreviations: FLT3-ITD, internal tandem duplications in the FLT3 gene; FLT3-TKD, FLT3 mutations in the tyrosine kinase domain; MLL-PTD, partial tandem duplications in the MLL gene; NPM1, nucleophosmin
**Figure 2** Expression of GATA2-mutants in HEK 293T cells

(a) Expression of FLAG-tagged GATA2-mutants confirmed by 3 independent Western Blot experiments using anti-GATA2 antibodies (upper panels) and anti-ß-actin antibodies (lower panels). Graph shows mean values of the GATA2 signals normalized to ß-actin signals as fold change compared to wild type GATA2. Error bars indicate standard error of the mean.

(b) Immunofluorescence of GATA2 using anti-FLAG antibodies and Alexa555-labeled secondary antibodies. DNA was counter stained using DAPI. Confocal laser scans show nuclear localization for GATA2 wild type (upper panels) and the recurring GATA2-mutants A318T (middle panels) or G320D (lower panels). Similar results were obtained for all GATA2-mutants tested (supplemental figure S3).

**Figure 3** Analysis of GATA2 mutants by transcription assays in HEK293T cells. Bars indicate fold activation of the luciferase reporters in at least three independent experiments each performed in triplicates. Errors bars represent standard error of the mean. P-values are indicated for pairwise comparisons using Student’s t-test.

(a) A GATA2-responsive reporter (GATA-LUC) containing GATA-binding elements derived from the TCRδ enhancer was co-transfected with expression plasmids for either GATA2 wild-type or GATA2 mutants. A reduced activation was observed for mutants A318T and L321F as compared to wild type GATA2. In contrast, an increased activation was observed for the G320D mutant. The L359V mutant of GATA2, located in ZF2, which is found in about 10% of patients at the progression of chronic myeloid leukemia to AML\(^21\), did not show a significant difference in transcriptional activation compared to wild type GATA2. (b) A CEBPA-responsive reporter (4xCAAT-BS-TK-LUC) containing CEBPA-binding elements was co-
transfected with expression plasmids for wild type CEBPA or a p30 CEBPA truncation mutant in combination with either GATA2 wild-type or GATA2 mutants. Enhancement of CEBPA-dependent transcriptional activation by wild-type GATA2 was significantly reduced for all the ZF1-mutants but not for L359V ZF2-mutant. Residual activation of the reporter by the p30 CEBPA truncation mutant was enhanced by co-expression of wild-type GATA2 (dashed line), but this enhancement was markedly reduced for all of the GATA2 mutants tested.

**Figure 4** Survival according to GATA2 mutational status in 143 CN-AML patients. (a) Overall survival and (b) Event-free survival within biCEBPA patients according to the GATA2 mutational status and compared to other GATA2 wildtype CN-AML patients. Abbreviations: biCEBPA, biallelic CEBPA mutations. \( P \): Log rank test

**Figure 5** Model of the N-terminal zinc finger of GATA2 (ZF1; yellow ribbon with brown DNA and grey zinc ion), based on the crystal structure of the DNA complex of the highly homologous zinc finger of GATA3\(^{38} \) (PDB accession number 3DFV). Mutated residues (magenta) are annotated and displayed with side chains. The mutations cluster at the DNA binding side of ZF1, suggesting they perturb DNA binding. Based on the homology model, N317, A318, L321 and R330 are directly implicated in DNA binding, so mutations in these residues likely alter the affinity to DNA or prevent DNA binding. G320 is important for proper attachment of an adjacent beta-hairpin loop that provides additional DNA binding contacts. Q328 is not directly involved in DNA binding. However, Q328->P could either affect the backbone fold of ZF1 and indirectly alter DNA binding by perturbing the adjacent R330, or alternatively is involved in interaction with other domains such as ZF2. All in all, the location of the mutations suggests that they influence DNA binding of ZF1.
Table 1: Confirmed tumor-specific somatic mutations identified by exome sequencing in 5 patients with biCEBPA mutated AML

<table>
<thead>
<tr>
<th>Patient</th>
<th>Gene</th>
<th>Genomic Position (hg18)</th>
<th>Reference Genotype</th>
<th>Variant Genotype</th>
<th>Amino Acid</th>
<th>Ensembl transcript</th>
<th>Read depth in AML</th>
<th>Variant Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pat #1</td>
<td>RBM39chr20:33765551</td>
<td>G/G</td>
<td>G/A</td>
<td>R356C</td>
<td>ENST00000361162</td>
<td>92</td>
<td>40.22%</td>
<td></td>
</tr>
<tr>
<td>Pat #1</td>
<td>IKZF1chr7:50417782</td>
<td>G/G</td>
<td>G/A</td>
<td>G158S</td>
<td>ENST00000359197</td>
<td>37</td>
<td>43.24%</td>
<td></td>
</tr>
<tr>
<td>Pat #1</td>
<td>FBLN7chr2:112639148</td>
<td>G/G</td>
<td>G/A</td>
<td>R112Q</td>
<td>ENST00000331203</td>
<td>36</td>
<td>25%</td>
<td></td>
</tr>
<tr>
<td>Pat #1</td>
<td>GATA2chr12:129685449</td>
<td>G/G</td>
<td>G/C</td>
<td>L321V</td>
<td>ENST00000341105</td>
<td>22</td>
<td>40.91%</td>
<td></td>
</tr>
<tr>
<td>Pat #1</td>
<td>KRASchr12:25289548</td>
<td>C/C</td>
<td>C/T</td>
<td>G13D</td>
<td>ENST00000256078</td>
<td>28</td>
<td>21.43%</td>
<td></td>
</tr>
<tr>
<td>Pat #2</td>
<td>GALNT10chr5:153769409</td>
<td>G/G</td>
<td>G/A</td>
<td>R427S</td>
<td>ENST00000297107</td>
<td>85</td>
<td>40%</td>
<td></td>
</tr>
<tr>
<td>Pat #2</td>
<td>STAG2chrX:123006890</td>
<td>A/A</td>
<td>*-C</td>
<td>L221Tfs</td>
<td>ENST00000371145</td>
<td>154</td>
<td>46.1%</td>
<td></td>
</tr>
<tr>
<td>Pat #3</td>
<td>GABREchrX:150874947</td>
<td>G/G</td>
<td>G/A</td>
<td>R276W</td>
<td>ENST00000370328</td>
<td>57</td>
<td>96.49%</td>
<td></td>
</tr>
<tr>
<td>Pat #3</td>
<td>ZNF318chr6:43431833</td>
<td>G/G</td>
<td>G/A</td>
<td>S406F</td>
<td>ENST00000318149</td>
<td>44</td>
<td>40.91%</td>
<td></td>
</tr>
<tr>
<td>Pat #3</td>
<td>STK19chrX:13386347</td>
<td>A/A</td>
<td>A/T</td>
<td>T76S</td>
<td>ENST00000375333</td>
<td>13</td>
<td>46.15%</td>
<td></td>
</tr>
<tr>
<td>Pat #3</td>
<td>DNMT3Achr2:25323985</td>
<td>G/G</td>
<td>*-A</td>
<td>G332Rfs</td>
<td>ENST00000321117</td>
<td>51</td>
<td>27.45%</td>
<td></td>
</tr>
<tr>
<td>Pat #4</td>
<td>TRPM8chr2:234543699</td>
<td>C/C</td>
<td>C/T</td>
<td>L749F</td>
<td>ENST00000324695</td>
<td>121</td>
<td>48.76%</td>
<td></td>
</tr>
<tr>
<td>Pat #4</td>
<td>AFTPHchr2:64633197</td>
<td>C/C</td>
<td>C/T</td>
<td>T362I</td>
<td>ENST00000238855</td>
<td>110</td>
<td>45.45%</td>
<td></td>
</tr>
<tr>
<td>Pat #4</td>
<td>PSD3chr18:18703129</td>
<td>G/G</td>
<td>G/A</td>
<td>R1118X</td>
<td>ENST00000286845</td>
<td>76</td>
<td>50%</td>
<td></td>
</tr>
<tr>
<td>Pat #4</td>
<td>DSC1chr18:26979591</td>
<td>G/G</td>
<td>G/A</td>
<td>R307I</td>
<td>ENST00000257198</td>
<td>83</td>
<td>44.58%</td>
<td></td>
</tr>
<tr>
<td>Pat #4</td>
<td>ASAP3chr1:23632600</td>
<td>G/G</td>
<td>G/A</td>
<td>Q678X</td>
<td>ENST00000336689</td>
<td>78</td>
<td>34.62%</td>
<td></td>
</tr>
<tr>
<td>Pat #4</td>
<td>B4GALNT1chr12:56306985</td>
<td>G/G</td>
<td>G/A</td>
<td>L471F</td>
<td>ENST00000341156</td>
<td>24</td>
<td>54.17%</td>
<td></td>
</tr>
<tr>
<td>Pat #4</td>
<td>FAM117Achr17:45152254</td>
<td>G/G</td>
<td>G/A</td>
<td>A192V</td>
<td>ENST00000240364</td>
<td>12</td>
<td>50%</td>
<td></td>
</tr>
<tr>
<td>Pat #4</td>
<td>DNMT3Achr2:25310746</td>
<td>C/C</td>
<td>C/T</td>
<td>R882H</td>
<td>ENST00000321117</td>
<td>32</td>
<td>46.88%</td>
<td></td>
</tr>
<tr>
<td>Pat #5</td>
<td>FLJ35848chr17:4099629</td>
<td>G/G</td>
<td>G/C</td>
<td>G275A</td>
<td>ENST00000409122</td>
<td>177</td>
<td>49.72%</td>
<td></td>
</tr>
<tr>
<td>Pat #5</td>
<td>PCDH13chr1:410243097</td>
<td>T/T</td>
<td>T/C</td>
<td>S354P</td>
<td>ENST00000289272</td>
<td>92</td>
<td>40.22%</td>
<td></td>
</tr>
<tr>
<td>Pat #5</td>
<td>NEUROD4chr12:53706952</td>
<td>C/C</td>
<td>C/T</td>
<td>R88X</td>
<td>ENST00000242994</td>
<td>78</td>
<td>42.31%</td>
<td></td>
</tr>
<tr>
<td>Pat #5</td>
<td>GATA2chr3:129685421</td>
<td>C/C</td>
<td>C/T</td>
<td>R330Q</td>
<td>ENST00000341105</td>
<td>17</td>
<td>58.82%</td>
<td></td>
</tr>
</tbody>
</table>
Somatic mutations affect 3 key regulators of hematopoiesis in patient #1

- **IKZF1** c.472 G>A
- **GATA2** c.961C>G
- **CEBPA** c.86delC
- **CEBPA** c.224_225ins CCC
**Fig. 1 b**
Missense mutations affecting the N-terminal zinc finger domain of GATA2

**Fig. 1 c**
Frequency distribution of additional mutations in biCEBPA mutant AML
Expression of GATA2-mutants confirmed by Western Blot

Non-transfected

Wild type

R293Q

V296G

A318T

G320D

L321F

L359V

α GATA2

α actin

Western Blot densitometry

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th>R293Q</th>
<th>V296G</th>
<th>A318T</th>
<th>G320D</th>
<th>L321F</th>
<th>L359V</th>
</tr>
</thead>
<tbody>
<tr>
<td>fold change</td>
<td>1</td>
<td>0.8</td>
<td>1</td>
<td>0.8</td>
<td>1</td>
<td>0.8</td>
<td>1</td>
</tr>
</tbody>
</table>

For personal use only.
Fig. 2b  Subcellular localization of GATA2-mutants

<table>
<thead>
<tr>
<th></th>
<th>DAPI</th>
<th>Alexa 555</th>
<th>Overlay</th>
</tr>
</thead>
<tbody>
<tr>
<td>GATA2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wild type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GATA2 A318T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GATA2 G320D</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Altered activation of the GATA-LUC-reporter by the GATA2 ZF1-mutants

![Graph showing the fold activation of different GATA2 mutants compared to the WT and Vector. The graph includes error bars and statistical significance levels (P<0.05, P<0.06, P=0.06, P=0.15).]

- **GATA2**
  - GATA-binding-site
  - TATA-box-promoter
  - luciferase
Reduced synergism between CEBPA and the GATA2-ZF1-mutants

Fig. 3 b
Survival according to GATA2 status in 143 CN-AML patients

a

![Graph showing overall survival](image1.png)

- biCEBPA + GATA2\textsuperscript{mutated} (n=10)
- biCEBPA + GATA2\textsuperscript{wildtype} (n=16)
- other GATA2\textsuperscript{wildtype} genotypes (n=117)

\(p = 0.274\)
\(p = 0.012\)

b

![Graph showing event-free survival](image2.png)

- biCEBPA + GATA2\textsuperscript{mutated} (n=10)
- biCEBPA + GATA2\textsuperscript{wildtype} (n=16)
- other GATA2\textsuperscript{wildtype} genotypes (n=117)

\(p = 0.185\)
\(p = 0.172\)
Fig. 5

Model of the N-terminal zinc finger of GATA-2
GATA2 zinc finger 1 mutations associated with biallelic CEBPA mutations define a unique genetic entity of acute myeloid leukemia

Philipp A. Greif, Annika Dufour, Nikola P. Konstandin, Bianka Ksienzyk, Evelyn Zellmeier, Belay Tizazu, Jutta Sturm, Tobias Benthaus, Tobias Herold, Marjan Yaghmaie, Petra Dörge, Karl-Peter Hopfner, Andreas Hauser, Alexander Graf, Stefan Krebs, Helmut Blum, Purvi M. Kakadia, Stephanie Schneider, Eva Hoster, Friederike Schneider, Martin Stanulla, Jan Braess, Maria Cristina Sauerland, Wolfgang E. Berdel, Thomas Büchner, Bernhard J. Woermann, Wolfgang Hiddemann, Karsten Spiekermann and Stefan K. Bohlander