Whole exome sequencing reveals a C-terminal germline variant in CEBPA-associated acute myeloid leukemia: 45-year follow up of a large family

Anand Pathak,[1] Katja Seipel,[2] Alexander Pemov,[1] Ramita Dewan,[1] Christina Brown,[3] Sarangan Ravichandran,[4] Brian T. Luke,[4] Michael Malasky,[5] Shalabh Suman,[5] Meredith Yeager,[5] NCI DCEG Cancer Genomics Research Laboratory, NCI DCEG Cancer Sequencing Working Group, Richard A. Gatti,[3,6] Neil E. Caporaso,[7] John J. Mulvihill,[8] Lynn R. Goldin,[7] Thomas Pabst,[2] Mary L. McMaster,[7*] and Douglas R. Stewart[1*]

[1]Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; [2]Departments of Medical Oncology and Clinical Research, University Hospital and University of Berne, Switzerland; [3]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA; [4]Advanced Biomedical Computing Center, Leidos Biomedical Research Inc., Frederick National Laboratory for Cancer Research, Frederick, MD, USA; [5]Cancer Genomics Research Laboratory, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Rockville, MD, USA; [6]Department of Human Genetics, David Geffen UCLA School of Medicine, Los Angeles, CA, USA; [7]Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; and [8]Department of Pediatrics, Section of Genetics, The University of Oklahoma College of Medicine, OK, USA

*MLMcM and DRS contributed equally.

ABSTRACT

Familial acute myeloid leukemia is rare and linked to germline mutations in *RUNX1, GATA2* or CCAAT/enhancer binding protein-α (*CEBPA*). We re-evaluated a large family with acute myeloid leukemia originally seen at NIH in 1969. We used whole exome sequencing to study this family, and conducted *in silico* bioinformatics analysis, protein structural modeling and laboratory experiments to assess the impact of the identified CEBPA Q311P mutation. Unlike most previously identified germline mutations in *CEBPA*, which were N-terminal frameshift mutations, we identified a novel Q311P variant that was located in the C-terminal bZip domain of C/EBPα. Protein structural modeling suggested that the Q311P mutation alters the ability of the CEBPA dimer to bind DNA. Electrophoretic mobility shift assays showed that the Q311P mu-tant had attenuated binding to DNA, as predicted by the protein modeling. Consistent with these findings, we found that the Q311P mutation has reduced transactivation, consistent with a loss-of-function mutation. From 45 years of follow up, we observed incomplete penetrance (46%) of *CEBPA* Q311P. This study of a large multi-generational pedigree reveals that a germline mutation in the C-terminal bZip domain can alter the ability of C/EBP-α to bind DNA and reduces transactivation, leading to acute myeloid leukemia.

Introduction

Familial acute myeloid leukemia (AML) is a rare disease and is linked to mutations in *RUNX1*,[1] *GATA2*[2,3] and CCAAT/enhancer binding protein-α (*CEBPA*).[4] *CEBPA* is located on 19q13.1 and is a single exon gene.[4] Its protein product, C/EBP-α, is a granulocyte differentiation factor. The protein consists of two N-terminal transactivating domains, a basic DNA binding domain and a C-terminal bZIP leucine-zipper dimerization domain. C/EBP-α is a critical protein in myeloid differentiation and regulates the expression of granulocyte specific genes.[5] *CEBPA* knock-out mice develop a phenotype similar to AML with a block in granulocyte maturation.[6]

Germline *CEBPA* mutations segregate in familial AML in an autosomal dominant manner. Smith *et al.* reported the first germline mutation in *CEBPA* in a family.[4] The causative mutation was an N-terminal (212delC) mutation that abrogated the production of the wild-type 42-kDa transcript, leading to the production of a 30-kDa product from an alternative start site. This mutation resulted in the loss of a transactivation domain. Five additional pedigrees have been identified with similar germline N-terminal frameshift insertions and deletions, leading to the production of the alternate dominant-negative 30-kDa product (*Online Supplementary Table S1*).[7-10] These germline mutations have also been frequently associated with somatic C-terminal insertions or deletions. This evidence suggests that the pathogenesis of AML caused by *CEBPA* may require bi-allelic inactivation of the gene. In addition, a previous study has uncovered germline C-terminal mutations in *CEBPA* associated with sporadic AML.[11] Somatic *CEBPA* mutations also occur in sporadic AML with a frequency of between 5%-14%.[12] Fasan *et al.* analyzed 2296 sporadic AML cases and found that 244 patients (10.6%) carried *CEBPA* mutations: 140 were *CEBPA* single mutant cases, 104 were *CEBPA* double mutant cases.[13] Patients with the *CEBPA* double mutation had a more favorable prognosis than the *CEBPA* single mutation.[13]

In 1970, Snyder *et al.* reported a kindred of 6 family members with AML in three generations evaluated at the National Cancer Institute.[14] A follow-up report on the family in 1977 revealed that a previously healthy sister had also developed AML.[15] Cells from 2 affected sisters and their mother all had increased cellular transformation on exposure to the SV40 virus.[15] This family was also studied for sister chromatid exchange, along with other familial leukemia families.[16] We have continued to follow up both the nuclear and extended family since their original description up to the present time. This large pedigree has 10 AML patients as well as 4 obligate carriers (Figure 1). Since there are multiple genes presumed to be associated with AML, we applied whole exome sequencing to 3 AML patients as well as to 2 obligate carriers to search for mutations in known and novel genes.

Methods

Detailed clinical descriptions of the patients are presented in the *Online Supplementary Appendix*. Patients from this family were enrolled into NIH protocol "Clinical, Laboratory and Epidemiologic Characterization of Individuals and Families at High Risk of Cancer" (#78-C-0039; NCT-00001163), which was subsequently merged into a specific familial leukemia/lymphoma protocol "Clinical, Laboratory and Epidemiologic Characterization of Individuals and Families at High Risk of Hematologic Cancer" (#02-C-0210; NCT-00039676). The study was approved by the NCI Institutional Review Board and written informed consent was obtained from all participants. Briefly, in 1970, 2 brothers had died of AML at 4 (IV.2) and 8 (IV.3) years of age and a sister (age 12 years, IV.4) was affected (Figure 1). Twin brothers (IV.5 and IV.6) were healthy, though a healthy sister (IV.1) developed AML in 1976.[15] Through the mother (III.2), an obligate carrier, an uncle (II.5), an aunt (II.8) and a cousin (III.7) were also affected; they had died at 63, 53 and 38 years of age, respectively. The mother (III.2) had a history of cervical cancer in 1963 that had been treated with radiotherapy; she had no history of leukemia. Three additional relatives have

developed AML (III.8, IV.7 and V2). Cytogenetic studies on peripheral bone marrow from IV.1, IV.4 and V.2 were all normal. A high percentage of peripheral blasts were observed in patients III.7 (68%) and V.2 (77%), as is typical of CEBPA mutated AML.

Discovery phase

Exome sequencing and annotation: whole exome sequencing on the AML family was performed at the National Cancer Institute Division of Cancer Epidemiology and Genetics Cancer Genomics Research Laboratory (NCI DCEG CGR). The procedure used has been described previously.[17,18] Briefly, 1.1 µg of genomic DNA was extracted by standard methods from low-passage cultured fibroblasts (IV.1, IV.4 and III.2) or from lymphocytes (III.5 and III.8). Pretreatment leukemic cells were not available for any of the tested AML patients; all testing was performed on healthy tissue and diseased cells were not available for testing.

Variant filtering: the variants were filtered by the following criteria: 1) present in all 5 individuals in the pedigree, 3 AML affected patients (III.8, IV.1 and IV.4) and 2 obligate carriers (III.2 and III.5); 2) appearing in less than or equal to 10 individuals in an in-house internal control group ("NCI CGR Out Group") of 1170 exomes in individuals from families with cancers other than hematologic cancers; 3) appearing in less than or equal to 1.0% in the NHLBI Exome Sequencing Project (ESP) European-American population (the ethnic background of this family); and 4) not present in areas of repeats or segmental duplications.

In silico analysis: we utilized SIFT,[19] PolyPhen-2,[20] Mutation Taster,[21] Mutation Assessor,[22] FATHMM,[23] and the Combined Annotation Dependent Depletion-scaled C-score,[24] for assessment of deleteriousness of mutations. For information on the degree of conservation at the site, we used GERP[25] and PhyloP.[26]

Full details of the materials and methods are provided in the *Online Supplementary Methods*.

Results

Whole exome sequencing and Sanger validation identifies a *CEBPA* Y11525.1:c.932A>C mutation (p.Gln311Pro, Q311P) in the NCI AML family

Exome sequence data were generated for the 3 affected members (IV.1, IV.4 and III.8) of the pedigree and the 2 obligate carriers (III.2 and III.5). There were a total of 479,900 variants (both synonymous and non-synonymous) identified in this family. We restricted analysis to non-synonymous mutations (missense, nonsense, splice site mutations) and frame-shift mutations, resulting in 31,448 variants. After filtering for genetic segregation in all 5 members of the family, and rarity (in-house NCI CGR outgroup database ≤10 and NHLBI ESP EA ≤1.0%), and removal of variants in regions of repeats or segmental duplication, there were only 3 variants remaining in the following genes: *ANGEL1*, *CEBPA* and *COL4A6*. Even with a more permissive filtering strategy (eliminating the NCI CGR Outgroup filter), there were no mutations in *FLT3, KRAS, NRAS, NPM1, RUNX1, MLL, KIT, PTPN11, TP53, GATA2, IDH* or *DNMT3A* detected, as segregating with this pedigree. Based on *in silico* bioinformatic predictions (*Online Supplementary Table S2*) and literature review, a rare missense variant in the *CEBPA* gene (CEBPA; g.19.33792389T>G; p.Gln311Pro, NM_004364:c.A932C) was deemed to be the most likely causative mutation of the familial AML clustering. This variant has not been previously reported in 1000 Genomes (2577 exomes), the NHLBI ESP database (6503 exomes) or the Broad Institute Exome Aggregation Consortium (61,486 exomes). In

Table 1. Sanger sequencing of the extended NCI AML family.

Pedigree ID	Age at Dec 2013 (years)	Cause of death	Cell/tissue type	NM_004364: c.932A>C (*CEBPA* Q311P)	AML	Age at AML diagnosis (years)
II.5	Deceased 1970, age 63	AML	N/A	N/A	Yes	62
II.8	Deceased 1956, age 53	AML	N/A	N/A (obligate carrier)	Yes	53
III.2	Deceased 1980, age 44	Adenocarcinoma of the rectum	Fibroblasts	Yes (heterozygous)	No	N/A
III.5	88	N/A	Lymphocytes	Yes (heterozygous)	No	N/A
III.6	Deceased 2008, age 68	Renal cancer	Lymphocytes	No	No	N/A
III.7	Deceased 1961, age 38	AML	N/A	N/A	Yes	36
III.8	Deceased 2012, age 69	AML	Lymphocytes	Yes (heterozygous)	Yes	58
III.10	Deceased 2013, age 80	Myocardial infarction	Lymphocytes	Yes (heterozygous)	No	N/A
IV.1	Deceased 1978, age 22	Sepsis	Fibroblasts	Yes (heterozygous)	Yes	20
IV.2	Deceased 1956, age 4	Gastrointestinal hemorrhage	N/A	N/A	Yes	3
IV.3	Deceased 1961, age 8	AML	N/A	N/A	Yes	7
IV.4	Deceased 1970, age 12	Anaphylaxis	Fibroblasts	Yes (heterozygous)	Yes	11
IV.5	54	N/A	Lymphocytes	Yes (heterozygous)	No	N/A
IV.6	54	N/A	Lymphocytes	Yes (heterozygous)	No	N/A
IV.7	Deceased 1996, age 43	AML	N/A	N/A (obligate carrier)	Yes	41
IV.9	64	N/A	Lymphocytes	No	No	N/A
IV.10	52	N/A	Lymphocytes	No	No	N/A
IV.11	60	N/A	Lymphocytes	Yes (heterozygous)	No	N/A
IV.12	51	N/A	Lymphocytes	No	No	N/A
IV.13	47	N/A	Lymphocytes	No	No	N/A
V.1	27	N/A	Lymphocytes	Yes (heterozygous)	No	N/A
V.2	24	N/A	Lymphocytes	Yes (heterozygous)	Yes	22

N/A: not applicable.

addition, this variant was not present in our in-house NCI CGR cancer exome database. This mutation was present in all 4 family members with AML for whom we had a DNA sample. We used bi-directional Sanger sequencing to validate the *CEBPA* variant initially detected by exome sequencing. *Online Supplementary Figure S1* shows that patient IV.1 had a c.932A>C heterozygous mutation, resulting in one allele coding for proline (codon CCG), instead of glutamine (codon CAG). Sanger validation was performed on fibroblasts from IV.1, IV.4 and III.2 and on lymphocytes on III.5 and III.8. We sequenced lymphocytes from 11 additional blood relatives from the family (Table 1); we were able to detect the *CEBPA* Q311P in the only other AML patient (V.2) in this group. While AML patients III.8, IV.1, IV.4, and V.2 all exhibited the mutation, subjects III.2, III.5, III.10, IV.5, IV.6, IV.11, and V.1 carried the mutation and did not develop AML.

CEBPA *Q311 is highly conserved*

We investigated the conservation of the *CEBPA* Q311P residue using GERP and Phylo-P. We found that Q311P is conserved by both GERP (score=4.7) and Phylo-P (score=4.5). Alignment of *CEBPA* orthologs (*Online Supplementary Figure S2, top panel*) shows absolute conservation of the glutamine 311 residue. In addition, the residue was conserved in four out of five *CEBPA* paralogs (*Online Supplementary Figure S2, bottom panel*).

Molecular modeling predicts reduced transcriptional activity of CEBPA *Q311P*

To regulate transcription, two protein chains of C/EBP-α dimerize in a crisscross orientation at its

"leucine zipper" domain (residues 317-345); the DNA strand is held using the basic motif (residues 286-313) in a "scissors-grip"[27-29] (*Online Supplementary Figure S3A*). The mutated residue, Q311, is located at the interface between the two C/EBP-α chains and is in an optimal position to form hydrogen bonds with residue T310 from the opposing chain (*Online Supplementary Figure S3B*) as previously suggested.[29] The Q311 mutation to proline, a rigid amino acid and a protein "helix-breaker", is thus predicted to perturb the leucine zipper dimer and its complex with DNA. Configurations of the wild-type and mutant homo- and heterodimers are presented in *Online Supplementary Figure S4.*

The computed binding energies of wild-type and mutant dimers were estimated using molecular dynamics simulations. Using the average of the energies from two simulations (*Online Supplementary Table S3*), the C/EBP-α dimerization energy and DNA binding energies were calculated (*Online Supplementary Table S4A and B*, respectively). These binding energies represent the amount of energy that would be required to separate the dimer into the two component monomers, and to remove the DNA from the C/EBP-α dimer, respectively. The binding energy of the heterodimer (Q:P) was 3.9 kcal/mol greater than for the wild-type dimer (Q:Q) and the binding energy of two mutated C/EBP-α (P:P) was 13.2 kcal/mol greater than the wild-type dimer (Q:Q), suggesting that the P:P dimer is the most stable and the wild-type Q:Q dimer is the least stable. This means that a mutated C/EBP-α monomer would rather bind to another mutated C/EBP-α. Modeling suggests that DNA was bound more tightly to the wild-type dimer than to the mutant
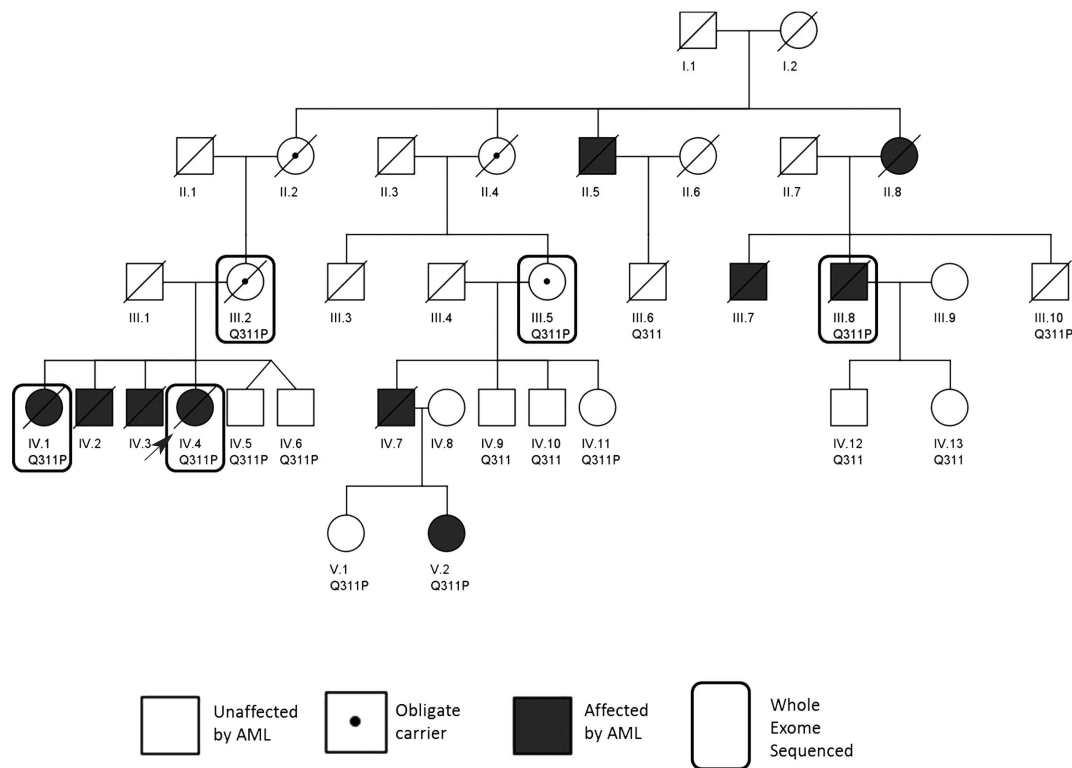
Figure 1. Pedigree of NCI AML family. *CEBPA* Q311P genotyping was performed on all individuals with available DNA. Filled symbols: AML cases; symbol with a dot in the center: obligate carrier. Boxed symbols: exome-sequenced samples; Q311 is the wild-type allele and Q311P is the variant at this position. Arrow: proband.

dimer (Figure 2 and *Online Supplementary Table S4B*). This DNA binding energy was 14.8 kcal/mol larger than that of the mutant dimer.

Given that the dimerization and DNA binding are reversible processes, the energy differences were used to calculate the Boltzmann probabilities of occurrence (*Online Supplementary Table S3*). These probabilities represent the relative abundance of the different complexes when the cell has achieved a steady-state equilibrium. These results show that the great majority of mutated C/EBP-α monomers are more likely to bind to another mutated C/EBP-α than to the wild-type protein. Therefore, in C/EBP-α-haploinsufficient cells, the predominant mutant dimer species is P:P. The wild-type monomers will have to form Q:Q dimers, since there will be very few mutated monomers available for wild-type dimerization. The modeling predicts that since the P:P dimer does not bind DNA nearly as tightly as the wild-type Q:Q dimer, we expect that its ability to initiate transcription will also be greatly reduced (Figure 3). Therefore, in a cell with a single Q311P mutation, the overall transcription activity should be reduced by about a factor of 2 since there are only half as many dimers Q:Q available for binding.

Experimental evidence that CEBPA *Q311P is a loss-of-function allele*

Molecular modeling predicts that Q311P, located in the hinge region between the basic DNA binding region and the leucine zipper region of the C/EBP-α protein, interferes with DNA binding and dimerization. The Q311P point mutation was introduced into the *CEBPA* expression plasmid by site-directed mutagenesis and was tested for transcriptional activator function in MOLM-16 AML cells (Figure 4A). As predicted, *CEBPA* Q311P was a loss-of-function allele since it was unable to activate reporter gene expression by itself (pcDNA/Q311 *vs.* pcDNA/Q311P; *P*=0.0002), despite comparable levels of wild-type and mutant protein (Figure 4B). In the heterozygous state (Q311+Q311P), the reporter gene induction was higher than in the hemizygous wild-type (pcDNA+Q311) (*P*=0.0011). If the Q311P was a dominant-negative, we would expect the heterozygous (Q311+Q311P) induction to be lower than the hemizygous wild-type (pcDNA+Q311). These data indicate that the Q311P mutation has a loss of transcriptional activator function, but has not acquired a dominant negative function, similar to other germline N-terminal frameshift mutations that are associated with familial AML. The homozygous WT state (Q311+Q311) exhibited significantly more transactivation than the heterozygous state (Q311+Q311P) (*P*=0.0343). There were no apparent differences in the C/EBP-α protein-DNA complexes of C/EBP-α wild-type and Q311P mutant *in vitro* on a high affinity CEBP binding site (Figure 3A). On a lower affinity site, however, only the wild-type C/EBP-α was able to bind, indicating that Q311P exhibits comparatively reduced DNA binding (Figure 3B).

Discussion

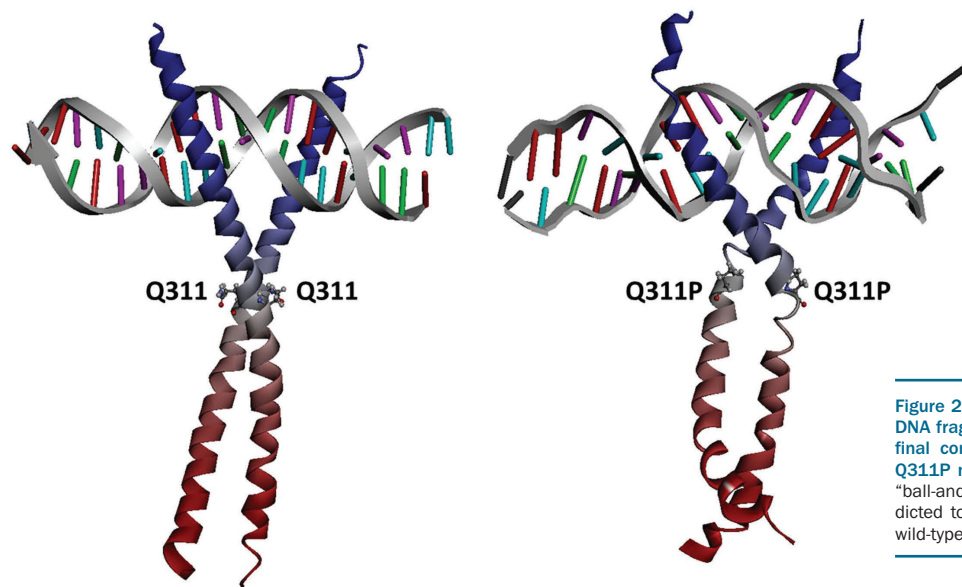Members of this family met the diagnostic criteria for "acute myeloid leukemia with mutated *CEBPA*", accord-

Figure 2. Comparison of the wild-type C/EBP-α - DNA fragment (PDB: 1NWQ; left) complex to the final conformation from the simulated double Q311P mutant (right). Residue 311 is shown in "ball-and-stick style". The mutant C/EBPα is predicted to bind to DNA with less avidity than the wild-type, affecting transcriptional regulation.

ing to the WHO Classification of Tumors of the Hematopoietic and Lymphoid Tissues.[30] AML with mutated *CEBPA* has higher peripheral blood blast cell counts than *CEBPA* non-mutated AML and a more favorable prognosis. Germline mutations in *CEBPA* are associated with familial AML.[4,7-10] To date, the majority of reported variants have been N-terminal frameshift mutations that create a premature stop codon and thus trigger transcription from an alternate start site; this results in the production of a truncated 30-kDa isoform, which lacks a transactivation domain. In addition, a study by Taskesen *et al*. has also identified germline C-terminal mutations associated with AML.[11] In a large, multi-generational pedigree with AML and 45 years of follow up, we report a C-terminal germline missense mutation (Q311P) in the C-terminal bZip domain of *CEBPA*, distinct from the previously reported causative variants. The majority of reported *CEBPA* somatic mutations in familial AML are in the C-terminus, located in close proximity to glutamine 311, the amino acid affected in this family. In addition, in the COSMIC database,[31] there are 13 missense/frameshift mutations at residue Q311, although Q311P was not observed.

Given the location of Q311P, we hypothesized that C/EBP-α dimerization and DNA binding ability would be affected. Protein molecular modeling pointed to preferential P:P (mutant:mutant) dimerization and a decreased ability of the Q311P dimer to bind DNA compared to the wild-type dimer. Experimental evidence showed that the ability of the Q311P mutant to transactivate a CEBP luciferase construct was significantly reduced, indicating that it is a loss-of-function allele. In the heterozygous state, with both the normal Q311 and mutant Q311P expressed, transactivation was no further reduced than the pcDNA and normal Q311, indicating that this mutant allele is not a dominant negative, as has been observed with other *CEBPA* mutations.[32] In accordance with the protein molecular modeling prediction, an electrophoretic mobility shift assay showed comparatively reduced binding of the Q311P mutant to DNA *versus* the wild-type Q311. This explains how the Q311P mutant loses its ability to activate transcription from a CEBP promoter.

Notably, a somatic in-frame duplication of amino acids 312-317 and a deletion of the lysine 312 in C/EBP-α also lead to the loss of DNA binding and reduced transactivation.[32]

We have 603 person-years of follow up in this AML family. The age of onset of AML ranged from three years to 62 years. The mean age of onset of AML in this family was 31.3 years (median 29 years). Unlike the generally good prognosis observed in *CEBPA* mutation carriers, the overall survival of the majority of our patients was only 1-2 years. This may be due to the diagnosis and treatment of the majority of the family members in the 1950s-1970s, when the survival after AML diagnosis was extremely poor. Family members more recently affected with AML survived longer. One patient, diagnosed at 57 years of age in 2001, survived 11 years (III.8) with chemotherapy, and another member recently diagnosed at 22 years of age (V.2) is doing well two years after a double umbilical cord transplant. It is possible that a C-terminal bZip mutation such as Q311P may have a worse prognosis than the N-terminal frameshift mutation, a pattern that has been previously reported.[4] In familial AML, it is well documented that a second somatic mutation in *CEBPA* may be required for leukemia development;[4,7-10] however, we did not have any appropriate tumor tissue available to test for this and only tested healthy tissue in this study. Thus, we cannot speculate whether the prognosis of our patients is associated with a single mutation or a double mutation.[13]

We observed a penetrance for AML of 46% [6 of 13 confirmed mutation carriers (III.8, IV.1, IV.4 and V.2) or obligate carriers (II.8 and IV.7)]; this rate is 59% (10 of 17) if we include family members (e.g. II.5, III.7, IV.2, IV.3) who died of AML but before DNA was collected and who presumably had the Q311P mutation. This is lower than the almost complete penetrance reported in the literature for the N-terminal frameshift mutations,[4,7-10] although the sample sizes are limited in these studies.

The mother (III.2) was diagnosed with cervical cancer at 27 years of age and received radium and cobalt therapy of uncertain dose and intensity in 1963. Pathological materials from this diagnosis were not available to for review. She developed poorly differentiated adenocarcinoma of

the rectum 16 years later (pathology materials reviewed by NCI Laboratory of Pathology) at 43 years of age and died of its complications at 44 years of age. There are 7 family members with the Q311P mutation who to date have not developed AML, other leukemia or lymphoma (III.5, IV.5, IV.6, IV.11, V.1; age range 27-88 years, mean 57 years).

In summary, we report an incompletely penetrant *CEBPA* Q311P mutation in all tested, affected members of the largest multi-generational AML pedigree reported to date with 45 years of longitudinal follow up. The Q311P mutation was predicted to be highly deleterious by *in silico* algorithms and the Q311 position was highly conserved among *CEBPA* orthologs and paralogs. Protein structural modeling suggested that the Q311P mutation alters the ability of the C/EBP-α dimer to bind DNA. The
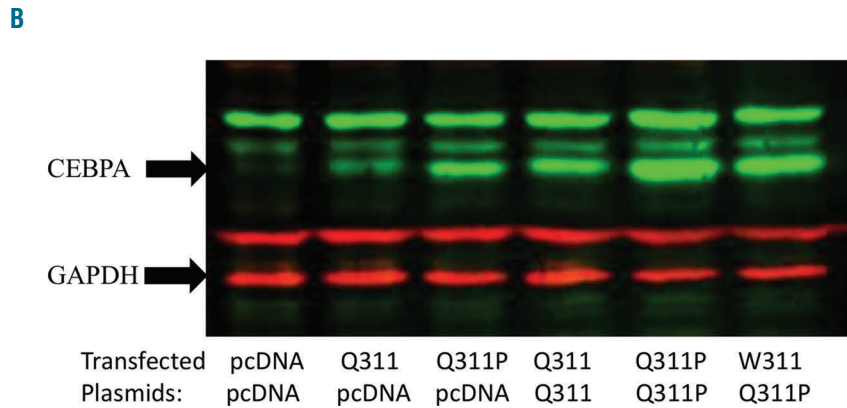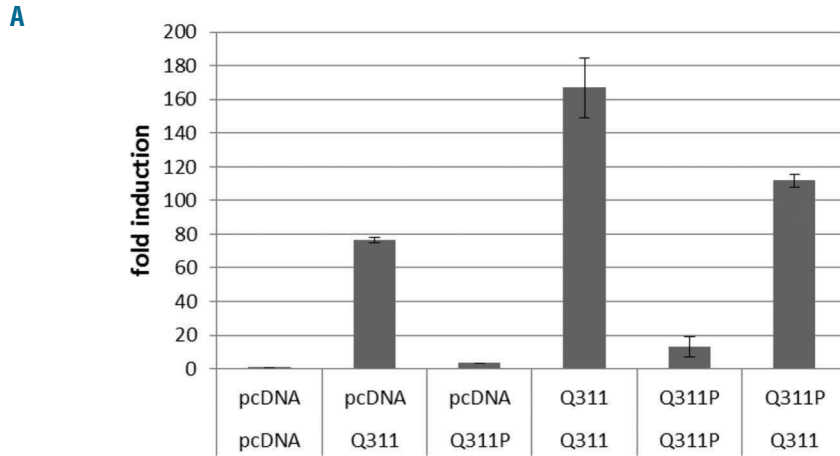
Figure 3. Transactivation potential of CEBPA Q311P *versus* wild-type. (A) CEBP promoter transactivation assays in MOLM16 cells for pcDNA, CEBPA Q311 (WT) and CEBPA Q311P (and all combinations), show a loss-of-function of CEBPA Q311P (experiments were performed in triplicate). *P*-value paired *t*-test values: pcDNA/Q311 *vs.* pcDNA/Q311P, *P*=0.0002; pcDNA/Q311 *vs.* Q311/Q311P, *P*=0.0011; Q311/Q311 *vs.* Q311P/Q311P, *P*= 0.0021; Q311/Q311 *vs.* Q311/Q311P, *P*=0.0343; Q311P/Q311P *vs.* Q311/Q311P, *P*=0.0013. (B) anti-CEBPA and anti-GAPDH Western blot of total extracts of transiently transfected MOLM16 cells.
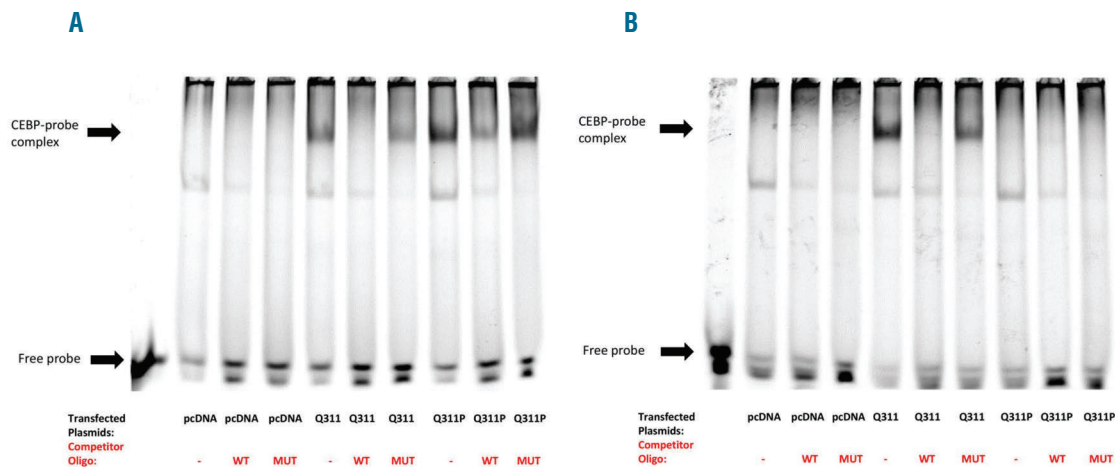

Figure 4. DNA binding potential of *CEBPA* Q311P *versus CEBPA* Q311 (wild-type). Electrophoretic mobility shift assay performed with nuclear extracts of H1299 cells transiently transfected with equal amounts of pcDNA, CEBPA Q311 (wild-type) and CEBPA Q311P expression plasmids. (A) Binding to high affinity site and (B) binding to low affinity site.

transactivation potential of C/EBP-α Q311P was reduced, suggesting it is a loss-of-function mutation. Furthermore, EMSA studies showed that the Q311P mutant had attenuated binding to DNA, providing a mechanism for its loss-of-function. Clinically, the *CEBPA* Q311P mutation was incompletely penetrant for AML; 46% of the carriers developed the disease. Thus, we conclude that a germline mutation in the C-terminal bZip domain, distinct from previously reported mutations, can alter the ability of C/EBPα to bind DNA and reduces transactivation, leading to AML, though with lower penetrance than the canonical risk variants.

References

1. Song WJ, Sullivan MG, Legare RD, et al. Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. Nat Genet. 1999;23(2):166-175.
2. Gao J, Gentzler RD, Timms AE, et al. Heritable GATA2 mutations associated with familial AML-MDS: a case report and review of literature. J Hematol Oncol. 2014; 7(1):36.
3. Hahn CN, Chong CE, Carmichael CL, et al. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. Nat Genet. 2011;43(10):1012-1017.
4. Smith ML, Cavenagh JD, Lister TA, Fitzgibbon J. Mutation of CEBPA in familial acute myeloid leukemia. N Engl J Med. 2004;351(23):2403-2407.
5. Radomska HS, Huettner CS, Zhang P, Cheng T, Scadden DT, Tenen DG. CCAAT/enhancer binding protein alpha is a regulatory switch sufficient for induction of granulocytic development from bipotential myeloid progenitors. Mol Cell Biol. 1998;18(7):4301-4314.
6. Zhang P, Nelson E, Radomska HS, et al. Induction of granulocytic differentiation by 2 pathways. Blood. 2002;99(12):4406-4412.
7. Nanri T, Uike N, Kawakita T, Iwanaga E, Mitsuya H, Asou N. A family harboring a germ-line N-terminal C/EBPalpha mutation and development of acute myeloid leukemia with an additional somatic C-terminal C/EBPalpha mutation. Genes Chromosomes Canc. 2010;49(3):237-241.
8. Pabst T, Eyholzer M, Haefliger S, Schardt J, Mueller BU. Somatic CEBPA mutations are a frequent second event in families with germline CEBPA mutations and familial acute myeloid leukemia. J Clin Oncol. 2008;26(31):5088-5093.
9. Renneville A, Mialou V, Philippe N, et al. Another pedigree with familial acute myeloid leukemia and germline CEBPA mutation. Leukemia. 2009;23(4):804-806.
10. Sellick GS, Spendlove HE, Catovsky D, Pritchard-Jones K, Houlston RS. Further evidence that germline CEBPA mutations cause dominant inheritance of acute myeloid leukaemia. Leukemia. 2005; 19(7):1276-1278.
11. Taskesen E, Bullinger L, Corbacioglu A, et al. Prognostic impact, concurrent genetic mutations, and gene expression features of AML with CEBPA mutations in a cohort of 1182 cytogenetically normal AML patients: further evidence for CEBPA double mutant AML as a distinctive disease entity. Blood. 2011;117(8):2469-2475.
12. Pabst T, Eyholzer M, Fos J, Mueller BU. Heterogeneity within AML with CEBPA mutations; only CEBPA double mutations, but not single CEBPA mutations are associated with favourable prognosis. Br J Cancer. 2009;100(8):1343-1346.
13. Fasan A, Haferlach C, Alpermann T, et al. The role of different genetic subtypes of CEBPA mutated AML. Leukemia. 2014;28 (4):794-803.
14. Snyder AL, Henderson ES, Li FP, Todaro GJ. Possible inherited leukaemogenic factors in familial acute myelogenous leukaemia. Lancet. 1970;1(7647):586-589.
15. McKeen EA, Miller RW, Mulvihill JJ, Blattner WA, Levine AS. Familial leukaemia and SV40 transformation. Lancet. 1977; 2(8032):310.
16. Cheng WS, Mulvihill JJ, Greene MH, Pickle LW, Tsai S, Whang-Peng J. Sister chromatid exchanges and chromosomes in chronic myelogenous leukemia and cancer families. Int J Cancer. 1979;23(1):8-13.
17. Stewart DR, Pemov A, Johnston JJ, et al. Dubowitz syndrome is a complex comprised of multiple, genetically distinct and phenotypically overlapping disorders. PLoS ONE. 2014;9(6):e98686.
18. Shi J, Yang XR, Ballew B, et al. Rare missense variants in POT1 predispose to familial cutaneous malignant melanoma. Nat Genet. 2014;46(5):482-486.
19. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073-1081.
20. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7 (4):248-249.
21. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods. 2010;7(8):575-576.
22. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. 2011;39(17):e118.
23. Shihab HA, Gough J, Cooper DN, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum Mutat. 2013;34(1):57-65.
24. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46(3):310-315.
25. Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglou S, Sidow A. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. 2005; 15(7):901-913.
26. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 2010;20(1):110-121.
27. Ellenberger TE, Brandl CJ, Struhl K, Harrison SC. The Gcn4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha-helices - crystal-structure of the protein-DNA Complex. Cell. 1992; 71(7):1223-1237.
28. Vinson CR, Sigler PB, McKnight SL. Scissors-grip model for DNA recognition by a family of leucine zipper proteins. Science. 1989;246(4932):911-916.
29. Miller M, Shuman JD, Sebastian T, Dauter Z, Johnson PF. Structural basis for DNA recognition by the basic region leucine zipper transcription factor CCAAT/enhancer-binding protein alpha. J Biol Chem. 2003;278(17):15178-15184.
30. Swerdlow SH, Campo E, Harris NL, et al. (eds.) World Health Organization Classification of Tumours of the Haematopoietic and Lymphoid Tissues. 4th Edition ed. Lyon: International Agency for Research on Cancer, 2008.
31. Forbes SA, Bindal N, Bamford S, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res. 2011; 39(Database issue):D945-950.
32. Gombart AF, Hofmann WK, Kawano S, et al. Mutations in the gene encoding the transcription factor CCAAT/enhancer binding protein alpha in myelodysplastic syndromes and acute myeloid leukemias. Blood. 2002;99(4):1332-1340.