Mutational spectrum of myeloid malignancies with inv(3)/t(3;3) reveals a predominant involvement of RAS/RTK signaling pathways

Stefan Gröschel,1,2,3 Mathijs A. Sanders,1 Remco Hoogenboezem,1 Annelieke Zeilemaker,1 Marije Havermans,1 Claudia Erpelinck,1 Eric M. J. Bindels,1 H. Berna Beverloo,4,5 Hartmut Döhner,3 Bob Löwenberg,1 Konstanze Döhner,3 Ruud Delwel,1 and Peter J. M. Valk1

1Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands; 2Department of Translational Oncology, National Center for Tumor Diseases and German Cancer Research Center, Heidelberg, Germany; 3Department of Internal Medicine III, University Hospital Ulm, Ulm, Germany; 4Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands; and 5Dutch Working Group on Hemato-Oncologic Genome Diagnostics, Rotterdam, The Netherlands

Key Points

- inv(3)/t(3;3) disease exhibits high rates of activated RAS/RTK signaling, epigenetic modifier, splice, and transcription factor mutations.
- AML and MDS with inv(3)/t(3;3) display similar mutational and gene expression profiles and should be considered a single molecular entity.

Introduction

Acute myeloid leukemia (AML) with inv(3)/t(3;3) or t(3;3)(q21;q26.2) [inv(3)/t(3;3)] is a distinct disease entity in the current World Health Organization classification.1 High therapy resistance is the common feature of myeloid malignancies, particularly AML with 3q21/3q26 aberrations, manifesting as low rates of complete remission and subsequent failure of current treatment strategies.2,3 Appearance of the characteristic 3q aberrations also indicates disease progression and portends adverse outcome in myelodysplastic syndrome (MDS) and chronic myeloid leukemia (CML).5,6 Therapy resistance in this subtype of malignancies is linked to the inappropriate activation of the proto-oncogene ectopic viral integration-1 (EVI1) as a consequence of the chromosome 3 rearrangements. EVI1 is a hematopoietic stemness factor and transcription factor with chromatin-remodeling activity.8,10 EVI1 is also overexpressed in approximately 11% of all AML cases in the absence of 3q aberrations and represents an independent adverse prognostic factor in these patients.11 We and others have shown that, as a consequence of inv(3)/t(3;3) rearrangements, EVI1 becomes activated via structural repositioning of a distal GATA2 enhancer from 3q21 to the EVI1 locus at 3q26.12,13 Relocation of the enhancer additionally confers reduced and monoallelic GATA2 expression in this AML subtype.

Myeloid malignancies bearing chromosomal inv(3)/t(3;3) abnormalities are among the most therapy-resistant leukemias. Deregulated expression of EVI1 is the molecular hallmark of this disease; however, the genome-wide spectrum of cooperating mutations in this disease subset has not been systematically elucidated. Here, we show that 98% of inv(3)/t(3;3) myeloid malignancies harbor mutations in genes activating RAS/receptor tyrosine kinase (RTK) signaling pathways. In addition, hemizygous mutations in GATA2, as well as heterozygous alterations in RUNX1, SF3B1, and genes encoding epigenetic modifiers, frequently co-occur with the inv(3)/t(3;3) aberration. Notably, neither mutational patterns nor gene expression profiles differ across inv(3)/t(3;3) acute myeloid leukemia, chronic myeloid leukemia, and myelodysplastic syndrome cases, suggesting recognition of inv(3)/t(3;3) myeloid malignancies as a single disease entity irrespective of blast count. The high incidence of activating RAS/RTK signaling mutations may provide a target for a rational treatment strategy in this high-risk patient group. (Blood. 2015;125(1):133-139)
Group we selected 32 AML (including 2 cell lines MUTZ-3 and UCSD-AML1), 4 CML-B.C (including 2 cell lines HNT-34 and MOL-M1), and 5 MDS cases for NGS analysis. Included patients harbored an inv(3)(t(3;3)) aberration on chromosome banding analysis (supplemental Table 1 available on the Blood Web site) subsequently confirmed by NGS analysis. Cultured CD3^+ T cells from diagnostic bone marrow served as whole-exome sequencing (WES, n = 10) germline control. Written informed consent was obtained from all individuals in accordance with the Declaration of Helsinki. All trials were approved by the institutional review boards of Erasmus University Medical Center and the University of Ulm. All samples were sequenced on the Illumina HiSeq 2500 system and processed as described previously.12

3q-capture sequencing

From the collected patient material, the genomic DNA was sheared with the Covaris S2 device (Covaris) with default settings. Subsequently, the sample libraries were prepared using the TruSeq DNA Sample Preparation Guide (Illumina). The target chromosomal regions 3q21.1-3q26.2 (~40 Mb) were captured by using custom in-solution oligonucleotide baits (Nimblegen SeqCap EZ Choice XL). The final sample libraries were subjected to paired-end sequencing (2 × 100 bp) and were aligned against the human genome hg19 (hg19) using the Burrows Wheeler Aligner (BWA) with default settings. Exact breakpoint positions were determined with Breakdancer v1.1.8 Exact breakpoint sequences were resolved by extracting proximal reads supporting or spanning the breakpoint using the identified breakpoint positions and an algorithm able to extract the relevant reads from BAM files by using the Samtools API.19 Relevant reads were identified by their discordant distance to the paired mate read as a result of the inv(3)(t(3;3)) aberration (supporting reads) or being a member of a cluster of truncated reads with the same clipping position (spanning reads). The extracted reads were subsequently used as input for the de novo assembler Velvet v1.0.1721 with default settings, and the assembled region was validated with UCSC Blat.21 If resolved, the breakpoint sequences of 3q21 and 3q26 were used for the estimation of the variant allele frequency (VAF) to infer the cellular prevalence of the inv(3)/t(3;3) aberration. All 3q-capture sequencing (3q-Seq) reads were aligned against the resolved breakpoint sequences of 3q21 and 3q26 and their respective native wild-type sequences. The VAF was estimated by comparing the total number of reads aligning on the breakpoint sequence to the total number of reads aligning to the respective native wild-type sequence.

RNA-Seq and whole-exome sequencing

From the collected patient material, total RNA was extracted with phenol-chloroform and subsequently transcribed by using Superscript II RT (Invitrogen). Shearing of the cDNA was performed with the Covaris S2 device (Covaris) with default settings and was further constructed according to the TruSeq RNA Sample Preparation v2 Guide (Illumina). The sample libraries were subjected to paired-end sequencing (2 × 75 bp) and aligned against hg19 using TopHat v2.22 Genomic DNA from patients and in vitro cultured control CD3^+ T cells were processed similar to 3q-Seq protocols and captured by exome bead capture (SeqCap EZ Human Exome Library v3.0). The sample libraries were paired-end sequenced (2 × 100 bp) and subsequently aligned against hg19 using BWA with default settings.17

Overall, we performed whole-transcriptome sequencing (RNA-Seq) on 41 and WES on 10 out of these 41 inv(3)(t(3;3)) myeloid malignancies. Read and alignment statistics for RNA-Seq and WES data are found in supplemental Figure 1A-C and supplemental Table 4. On average, we observed a medium to high coverage for the targeted exome in WES data (~62x), sufficient to detect mutations with a VAF of 10% or more. Reads generated for RNA-Seq analyses predominately fell within transcribed regions (~52%) (ie, ribosomal genes, coding sequence, and UTRs), according to the RefSeq Transcriptome database, and, on average, 91% of the reads could be aligned to hg19. Gene expression profiles (GEP) for 24 inv(3)(t(3;3)) patients were constructed for differential expression, cluster, and principle component analyses with the DESeq2 package.23 Copy number variation (CNV) profiles from the WES data were calculated by CNVnsvd (M.A.S., R.H, and P.J.M.V., manuscript in preparation; supplemental Figure 2). Brieﬂy, per patient the total number of fragments was determined for each exon or determined from consecutive 500 nucleotide-wide windows for large exons.

Clonality analysis

The VAFs of the acquired mutations were estimated from the 10 paired inv(3)/t(3;3) malignancies characterized with WES. The VAF of the inv(3)/t(3;3) aberration was estimated from the 3q-Seq data unless the breakpoints could not be resolved or no 3q-Seq data were available. In these cases, the cytogenetically determined inv(3)(t(3;3)) positive metaphases were used. The VAFs were corrected by the local CNV, determined by CNVnsvd, and possible loss-of-heterozygosity ascertainment by determining the loss of proximal heterozygous SNVs with respect to the control WES data. The clonal architecture was illustrated in violin plots. In brief, the density of mutations with a similar VAF was determined by a kernel-density approach and is represented by the width of the graph. These plots were generated by the R package “easyGgplot2.”30

Mutant disease allele categorization

We first assigned mutations to mutational categories to discern patterns of mutations within inv(3)(t(3;3)) myeloid disease (Figure 1A).28 All

Results

Variant detection

RNA-Seq data were preprocessed for variant detection by splitting the exon boundary spanning reads using the Genome Analysis Toolkit.22 Subsequently, the variants were determined with the Samtools API and MuTect for RNA-Seq and WES data.19,25 The detected variants were annotated with Annovar26 and further characterized by multiple read statistics determined by an in-house developed algorithm. In brief, the algorithm determines for each variant the VAF, local read statistics based on the alignment and base qualities, mutation likelihood given the local sequence context, recurrence given the catalog of somatic mutations in cancer, recurrence determined from population-based sequencing efforts (1000 genomes project), and, when available, the likelihood of the mutation given the same set of read statistics in a control sample. The validity of our approach combining WES data with RNA-Seq data to infer variants is substantiated by the observation that nonsense-obsolete decay was negligible for mutant allele detection, as was demonstrated by similar VAFs of mutant disease alleles observed within cases characterized by both WES and RNA-Seq (supplemental Figure 3). Frameshift and premature stop codon–introducing mutations were selected and dichotomized on their location in the gene body. Mutations located in the terminal exon or approximately 50 bp from the exon boundary of the penultimate exon should theoretically be unaffected by nonsense-obsolete decay, whereas stop codon–introducing mutations situated in other locations of the gene body should be affected. Finally, variants were examined when they were recurrently detected in ≥2 patients or if they were previously linked to leukemogenesis or cancer pathogenesis.27,28 All listed variants were validated by Sanger sequencing, except for FLT3-ITD, which was determined by reverse-transcription polymerase chain reaction.

Allelic imbalance of GATA2

In total, 30 inv(3)(t(3;3)) cases accommodated informative heterozygous single-nucleotide variants (SNVs) in the GATA2 locus according to the 3q-Seq data. We previously showed that the inv(3)(t(3;3)) causes monoallelic expression of GATA2 from the nonarranged allele.12 Subsequently, we determined the allelic contribution of the genotypes of the heterozygous SNV in the matched RNA-Seq case. The average of the allelic contribution was taken when multiple heterozygous SNVs were accommodated in the GATA2 locus. The polar histogram was constructed with the R package “phenotypicForest.”29

References

1. GROSCHEL et al. Blood, 1 January 2015. Volume 125, Number 1

From www.bloodjournal.org by guest on April 15, 2017. For personal use only.
identified mutations were confirmed to be somatic in samples with available paired T-cell control (10/41 cases). In addition to the “hardwired” deregulated expression of EVI1 and GATA2, all 41 samples contained at least one additional mutation in one of the categories relevant for leukemia pathogenesis (average 2.3 category mutations per sample [Figure 1A and supplemental Tables 2 and 3]). Notably, all AML and CML-BC, as well as 4 out of 5 MDS samples contained mutations in genes activating RAS/RTK signaling, amounting to an incidence of 98% of all malignancies with an inv(3)/t(3;3). Furthermore, mutations were frequently found in myeloid transcription factor genes (32%), splice factor–encoding genes (29%), epigenetic modifier genes (29%), tumor-suppressor genes (10%), DNA-methylation genes (10%), and cohesin-complex genes (5%) (Figure 1A).

Complementing previous reports on the high incidence of NRAS mutations in inv(3)/t(3;3) AML, we found on aggregate 47% of all samples containing mutations directly affecting RAS, that is, NRAS (27%), KRAS (11%), and NFI (9%) (Figure 1B). These mutations were mutually exclusive and also largely nonoverlapping with any other mutation affecting signaling pathways involving RAS (ie, PTPN11 [20%], FLT3 [13%], CBL [7%], KIT [2%], and BCR-ABL1 [12%]) (Figure 1B). GATA2 was the most commonly mutated transcription factor in inv(3)/t(3;3) myeloid malignancies (15%; 5 AML and 1 MDS patient) and occurred in all cases in 1 of the 2 GATA2 zinc-finger domains. RUNXI1 mutations were present in 12% and did not coincide with GATA2 mutations; however, mutations in the splice factor–encoding gene SF3B1 (27%) were enriched in GATA2-mutated samples. Mutations in GATA2, SF3B1, and RUNXI1 were established to be somatic in all cases with control material available. Interestingly, we detected novel truncating mutations and CNVs, resulting in the loss of 1 copy of the transcription factor FOXP1 on 3p14.1 (10%), which is recurrently involved in chromosomal aberrations within lymphoma, but its association with AML pathogenesis is unknown. The predominant monosomal karyotype within inv(3)/t(3;3) myeloid malignancies, mainly conferred by
monosomy 7 (68%), is contrasted by the low incidence of TP53 mutations (5%) (Figure 1A), which had been suggested to be involved in the etiology of complex and of monosomal karyotype AML.32,33

No mutational pattern alluded to the high coincidence of the loss of chromosome 7 in inv(3)/t(3;3) myeloid disease (Figure 1 and supplemental Tables 2 and 3). However, previous reports have indicated that haploinsufficiency for CUX1 (located on 7q22.1), a gene strongly downregulated in our cohort of inv(3)/t(3;3)/-7 patients (supplemental Table 5), activates phosphoinositide-3 kinase (PI3K) signaling by transcriptional downregulation of the PI3K inhibitor PI3KIP1,34 and could therefore be an important cooperating lesion in inv(3)/t(3;3)/monosomy 7 myeloid syndromes.35

To date, no independent prognostic factor within the inv(3)/t(3;3) AML subset has been identified as a result of its low incidence and the extremely short median survival of inv(3)/t(3;3) AML patients (10 months).3 Baseline patient characteristics and clinical outcome data were available in 21 individuals with inv(3)/t(3;3) AML. The high frequency of RAS/RTK pathway mutations allowed us to perform an exploratory analysis within this small patient cohort. There were no statistically significant differences in patient characteristics, nor overall survival (OS) and event-free survival in cases with RAS mutations (NRASmut, KRASmut, NF1mut) compared with cases with other mutations activating signaling pathways (supplemental Figure 4). The median OS of RASmut patients was 9.8 months vs 8.9 months of other RTKmut patients (median OS 9.8 months).

Clonality analysis

To address the question of whether the highly overrepresented RAS/RTK pathway mutations and other recurrent somatic alterations in inv(3)/t(3;3) AML co-occurred in the same dominant clone, we assessed the allelic ratios of the EVII-rearranged and mutant-candidate disease alleles (Figure 2). WES analysis in conjunction with germline T-cell control was available from 10 AML patients. Cytogenetic evaluation of blast percentage and NGS read count estimation of the percentage of the 3q21q26.2 fusion (allele frequency) were concordant. In 2 cases (AML 20908 and 29656) without available 3q-seq data, cytogenetics served to estimate the percentage of the inv(3) allele. The inv(3)/t(3;3) aberrations were detectable in the majority of cases (7/10 cases) in as much as 100% of the cells (ie, resulting in an allelic ratio of the heterozygous 3q21q26.2 fusion allele of approximately 0.5), reflecting high blast percentage in these cases. The RAS and RTK mutations were mainly found in the dominant EVII-rearranged clone, and a similar pattern is found for all other identified alterations (eg, in transcription factor, splice factor, epigenetic modifier genes), which mostly co-occur at a similar frequency as the RAS/RTK mutations. However, in AML 12383 (PTPN11 mutation) and AMLs 29656 and 30309 (both NF1-mutated), the 3q-rearrangement was found in the major clone, whereas the RAS pathway mutations were present in only about half of these cells. In 2 cases (AML 20613 and 20908), the inv(3)/t(3;3) aberrations were less frequent than other concomitant mutations. In the inv(3) MDS case 28382 without any detected activating signaling mutation, the allelic ratio of the inv(3) was about 0.25, suggesting that both dysplastic-appearing cells as well as myeloblasts (blast percentage as per cytologic evaluation <20%) carried both the inv(3) aberration and coincident gene mutations (SF3B1, TP53, DNMT3A; see Figure 1A). Together, these data suggest that the inv(3) or t(3;3) aberration is the primary genetic hit in this subset of malignancies, with high proportion of clones harboring concurrent activating signaling mutations. Owing to the very short survival of these patients and general failure to achieve complete remission, no time-course monitoring could be performed to reveal clonal evolution.

Expression of mutant GATA2

The inv(3)/t(3;3) chromosomal rearrangements separate an upstream GATA2 enhancer from 3q21 and fuse it to the 3q26.2/EVII locus, thereby acquiring features of a monoallelic super-enhancer on the rearranged 3q allele.12,13,36,37 Integrative analysis of RNA-seq with 3q-capture DNA-seq data using informative, heterozygous SNVs (single-nucleotide polymorphisms plus somatic mutations) revealed almost exclusive monoallelic expression of the mutant GATA2 alleles (Figure 3), as shown in the polar plot by the contribution of the rearranged 3q and nonrearranged 3q allele read counts for GATA2 in 30 inv(3)/t(3;3) cases, including cell lines available for analysis. This observation indicates that the remaining active, non-rearranged GATA2 allele acquired the mutation, whereas the nonmutated GATA2 allele was silenced as a result of the chromosomal rearrangement. Thus, in our inv(3)/t(3;3) AML cohort, heterozygous GATA2 mutations were “functionally” hemizygous as a result of monoallelic GATA2 silencing.
Gene expression and mutation patterns in AML and MDS

It is a matter of debate whether MDS with the distinct inv(3)(q21q26.2) or t(3:3)(q21.q26.2) should be regarded as AML, irrespective of blast percentage in the bone marrow, similar to the current World Health Organization (WHO) guidelines applied in the diagnosis of core-binding factor AML with inv(16)(t16;16) or t(8;21) and of acute promyelocytic leukemia with t(15;17). In an effort to discriminate MDS and AML with inv(3)(t3:3) based on gene expression programs and the spectrum of coincident gene mutations, we performed cluster and principle component analyses (Figure 4A-B). No cluster formation emerged, neither based on the MDS/AML dichotomy nor any other unsubstantiated group within our data set. Furthermore, we performed a differential expression analysis to infer genes that could differentiate between MDS and AML. In summary, after Benjamini-Hochberg correction for multiple testing, we could only detect 2 differentially expressed genes (C11orf45: P = .0009, CILP: P = .04) without a documented role in leukemogenesis. In addition, we observed that MDS patients with inv(3)(t3:3) are as equally therapy resistant as their AML counterparts in a small set of cases analyzed (data not shown). In conclusion, we were unable to detect cluster formation, indicating the strong homogeneity of inv(3)(t3:3) myeloid malignancies based on GEPs and the pattern of cooperating genetic lesions.

Discussion

Collectively, we present data that suggest a common genetic background of myeloid malignancies harboring inv(3) or t(3:3) and show that RAS alterations and activating RTK mutations are more frequent in this disease subset than has been previously reported. The spectrum of secondary genetic lesions is generally found in the same ETV1-rearranged dominant clone. No unique cluster within inv(3)(t3:3) myeloid malignancies could be identified, neither by gene expression or mutation profiling nor by analysis of patient characteristics or clinical outcome. Thus, our data further support the notion that inv(3)(t3:3) myeloid disorders could be categorized as AML, irrespective of blast count, similar to WHO AML categories t(8;21), inv(16)(t16;16), or t(15;17), which is also suggested by the molecular pathobiology common to all inv(3)(t3:3) myeloid malignancies.

Reclassification of the currently annotated WHO AML subtype inv(3)(t3:3); RPN1-EVI1–rearranged as inv(3)(t3:3); GATA2-EVI1–rearranged AML is supported by the observation that GATA2 allelic imbalances and monoallelic expression of heterozygous GATA2 mutations occur because of the distinct chromosomal rearrangements. Whether this and other myeloid transcription factor alterations contribute to disease biology and the highly adverse clinical phenotype of inv(3)(t3:3) patients remains to be shown, although GATA2 and other transcription factor disturbances have been described to be preleukemic lesions. Of note, myeloid malignancies with inv(3) or t(3:3) define yet another subset of AML with high enrichment of GATA2 mutations next to CEBPA-mutated AML.

We included CML cases in blast crisis with an inv(3)(t3:3) under the assumption that CML-BC closely resembles AML biology. The BCR-ABL1 fusion is an RTK mutant that in itself activates RAS pathways and is the first event in transformation of myeloid precursors, as opposed to MDS and AML cells first acquiring inv(3)(t3:3). Despite the difference of the biology and the etiology of CML, the mutational spectrum of inv(3)(t3:3) CML-BC cells appeared to be same, as was further suggested by transcriptome analysis, which showed that GEP of the one CML-BC case did not differ from that of AML and MDS cases. However, the small number of inv(3)(t3:3) MDS and CML cases in our study preclude conclusive assessment of the role of inv(3)(t3:3) with regard to clinical phenotype. In summary, inv(3)(t3:3) myeloid malignancies harbor a common set of molecular alterations (ie, ETV1 and GATA2 deregulation coupled with mutations activating key signaling pathways). The dependence on constitutive RAS/RTK signaling activity of inv(3)(t3:3)-transformed AML cells might be the molecular correlate of the observed high white blood cell counts in this disease subset. Also, in view of the negative impact of GATA2 deficiencies on proliferation and regeneration of myeloid progenitors, these activated signaling mutations may be indispensable for survival and propagation of inv(3)(t3:3)-transformed myeloid progenitors. The high mutational burden of inv(3)(t3:3) cells compared with other AML subtypes (supplemental Table 2) could also provide clues about why inv(3)(t3:3) malignancies invariably associate with an extremely poor prognosis. Because these rare inv(3)(t3:3) myeloid malignancies form a highly...
unmet medical need, novel therapeutic approaches could be derived from the observation of constitutive activation of the MAPK pathway in almost 100% of these tumors. Exploiting signaling pathways therapeutically by using FLT3- or PI3K-inhibitors or hypothetically by interfering with RAS-signaling, possibly in combination with BET-inhibitors, may serve as valuable adjuncts to the scarce armamentarium of chemotherapeutic drugs effective in this subset of malignancies.

Acknowledgments

This work was supported by grants from the Deutsche Forschungsgemeinschaft (GR3955/1-1) (S.G.), the Lady Tata Memorial Trust (S.G.), the Center for Translational Molecular Medicine (GR03O-102) (M.A.S.), an EHA Research Fellowship (S.G.), and the Worldwide Cancer Research (formerly AICR) (12-1309) (E.B.).

Authorship

Contribution: S.G., M.A.S., R.D., and P.J.M.V. designed research, performed experiments, analyzed and interpreted data, and wrote the manuscript; S.G., A.Z., M.H., R.H., E.M.J.B., and C.E. generated NGS libraries and performed Sanger and Illumina sequencing; H.B.B., B.L., K.D., H.D., and P.J.M.V. collected specimens and clinical data; H.B.B., K.D., H.D., and P.J.M.V. performed cytogenetic and molecular analyses of leukemia samples.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Peter J. M. Valk, Department of Hematology Erasmus University Medical Center, Wytemaweg 80, Room Nc806, 3015 CN Rotterdam, The Netherlands; e-mail: p.valk@erasmusmc.nl.
References


6. Haferlach C, Bachner U, Haferlach T, et al. The inv(3)(q21;26)/t(3;3) is frequently accompanied by alterations of the RUNX1, KRAS and NRAS and NFI genes and mediates adverse prognosis both in MDS and in AML: a study in 39 cases of MDS or AML. Leukemia. 2011;25(5):874-877.


Mutational spectrum of myeloid malignancies with inv(3)/t(3;3) reveals a predominant involvement of RAS/RTK signaling pathways

Stefan Gröschel, Mathijs A. Sanders, Remco Hoogenboezem, Annelieke Zeilemaker, Marije Havermans, Claudia Erpelinck, Eric M. J. Bindels, H. Berna Beverloo, Hartmut Döhner, Bob Löwenberg, Konstanze Döhner, Ruud Delwel and Peter J. M. Valk