Supplement for:

Impact of *TP53* on outcome of patients with myelofibrosis undergoing hematopoietic stem cell transplantation

Nico Gagelmann, Anita Badbaran, Rachel B Salit, Thomas Schroeder, Carmelo Gurnari, Simona Pagliuca, Victoria Panagiota, Christina Rautenberg, Bruno Cassinat, Felicitas Thol, Christine Wolschke, Marie Robin, Michael Heuser, Marie-Thérèse Rubio, Jaroslaw P Maciejewski, Hans Christian Reinhardt, Bart L Scott, Nicolaus Kröger

Results

Additional methods

Results

Supplemental Table 1. Univariate analysis on survival.

Factor	Hazard ratio	95% CI	P
***TP53* status**			
Wild-type	Reference		
Single-hit	1.34	0.65-2.75	0.43
Multi-hit	2.64	1.65-4.23	<0.001
Karnofsky performance status			
90-100%	Reference		
<90%	1.64	1.16-2.31	0.005
Age			
<57 years	Reference		
≥57 years	1.86	1.29-2.69	0.001
***ASXL1* mutation**	1.54	1.07-2.22	0.02
Driver mutation genotype			
CALR	Reference		
JAK2	2.15	1.28-3.62	0.004
MPL	0.54	0.12-2.33	0.41
Triple negative	2.74	1.50-5.00	0.001
Platelets <150 x 10^9/L	2.04	1.42-2.94	<0.001
Leukocytes >25 x 10^9/L	1.47	0.97-2.23	0.07
Hemoglobin level (continuous)	0.94	0.85-1.03	0.19
Circulating blasts (continuous)	1.03	0.99-1.07	0.12
Donor relation			
Matched related	Reference		
Matched unrelated	1.18	0.73-1.91	0.49
Mismatched related	1.49	0.52-4.30	0.46
Mismatched unrelated	2.55	1.55-4.20	<0.001
High molecular risk[1]	1.33	0.91-1.50	0.15
High intensity conditioning	1.40	0.93-2.11	0.10
Use of antithymocyte globulin	0.74	0.52-1.07	0.11
Secondary myelofibrosis	1.08	0.74-1.56	0.70
Ruxolitinib before transplant	0.94	0.65-1.35	0.73
Female patient	0.94	0.64-1.38	0.76
Transfusion dependence	1.15	0.80-1.64	0.46

Supplemental Table 2. Multivariate analysis on survival.

Factor	Hazard ratio	95% CI	P
***TP53* status**			
Wild-type	Reference		
Single-hit	1.59	0.76-3.03	0.21
Multi-hit	2.15	1.32-3.53	0.002
Karnofsky performance status			
90-100%	Reference		
<90%	1.48	1.03-2.13	0.03
Age			
<57 years	Reference		
≥57 years	1.58	1.07-2.34	0.02
***ASXL1* mutation**	1.53	1.05-2.22	0.03
Non-*CALR/MPL* driver mutation genotype	1.67	1.00-2.85	0.05
Platelets <150 x 10^9/L	2.08	1.40-3.09	<0.001
Leukocytes >25 x 10^9/L	1.51	0.96-2.38	0.07
Mismatched unrelated donor	2.14	1.46-3.13	<0.001

Methods

Bone marrow or peripheral blood samples were obtained before transplantation and start of conditioning. Genes studied in all patients were: *JAK2, CALR, MPL, ASXL1, IDH1/2, CBL, DNMT3A, TET2, SF3B1, SRSF2, U2AF1, EZH2, TP53, NRAS, KRAS, RUNX1, IKZF1, SH2B3,* and *FLT3*; and the detection threshold for all methods was 2%.

Most patients (n=247) were sequenced in the Hamburg University Medical Center.[2,3] In brief, mononuclear cells from bone marrow or peripheral blood were enriched by Ficoll density gradient centrifugation and stored in liquid nitrogen until further use. Genomic DNA was extracted from frozen samples using the All Prep DNA/RNA Kit (Qiagen, Hilden, Germany), according to the manufacturer's recommendations. The mutation hotspots of the following genes were amplified by PCR and sequenced by Sanger sequencing: *JAK2* (exons 12 and 14), *CALR* (exon 9), *MPL* (exon 10), *ASXL1* (exon 12), *IDH1* (exon 4), *IDH2* (exon 4), *CBL* (exons 8 and 9), *DNMT3A* (exon 23), *TET2* (exons 3 to 11), *SF3B1* (exons 13 to 16), *SRSF2* (exons 1 to 3), and *U2AF1* (exons 1 to 8). Sequences were analyzed using Mutation Surveyor software (SoftGenetics, State College, PA).

In addition, patients were sequenced to the above-mentioned genes by SOLiD sequencing for *EZH2* exons 2 to 20, *DNMT3A* exons 2 to 23, *ETV6* exons 2 to 8, and *PTPN11* exons 3 and 13. The sequencing protocol has been described in detail before. PCR amplicons were used to generate DNA libraries with patient-specific barcodes. Libraries were sequenced on the SOLiD system (Life Technologies, Darmstadt, Germany) according to the manufacturer's protocol. Individual reads were 75 base pairs long. Reads were assigned to their patient-specific barcode, and sequences were analyzed twice separately using the DNAnexus software (San Francisco, CA) and the following pipeline of bioinformatics software. The color-space reads were aligned with NovoAlignCS and genotyped with GATK's Unified Genotyper. Single-nucleotide variants (SNVs) and indel discovery was performed across all samples using standard parameters; a maximum coverage of 10,000 was considered in the analysis.

From the resulting list of candidate SNVs we retained only SNVs that were located within the exons of interest using the core functions of R. First, mutations outside the coding region were excluded. Second, known single nucleotide polymorphisms (dbSNP version 137) were removed. Third, mutations with a VCFv4.1 quality score (https://samtools.github.io/hts-specs/VCFv4.1.pdf) of less than 8000 or an allele frequency of less than 15% were excluded (mean coverage of SNVs was 1232 reads, minimal read depth was 199 reads). To verify the results, all detected somatic mutations were sequenced again using an Ion Torrent Sequencer (PGM; Ion Torrent—Life Technologies GmbH/Thermo Fisher, Darmstadt, Germany) for the following genes: *DNMT3A, FLT3, RUNX1, NRAS, KRAS, MPL, ASXL1, TET2, EZH2, IKAROS, JAK2-V617F, JAK2-EXON12, CBL, LNK, IDH1, IDH2, SF3B1, SRSF2*. Changes in nucleic acid sequence were annotated using the IGV-Data bank as well as the Ion Reporter software (Life Technologies GmbH/Thermo Fisher). In addition, all mutations were validated by Sanger sequencing.

Analysis of mutations in genes of the splicing machinery of patients from Hannover Medical School (n=7) was done as reported previously.[2,4,5] Genomic DNA was amplified using a whole-genome amplification kit following the manufacturer's

recommendations (Sigma-Aldrich). All 11 exons of *ZRSR2* and the mutational hotspots of *U2AF1* (exon 2 and exon 6) and *SRSF2* (exon 1) were sequenced in all patients. Mutational hotspots of *U2AF1* and *SRSF2* were also sequenced in 50 healthy volunteers. Patient DNA was amplified using the following PCR conditions: 95°C for 10 minutes and 39 cycles of 95°C for 1 minute, 56°C for 1 minute, and 72°C for 70 seconds, followed by 1 cycle of 72°C for 10 minutes. Sequencing of *SRSF2* required 2 rounds of amplification using nested primers with the same cycling conditions as above for 35 cycles with each primer pair (supplemental Table 1). Purified PCR fragments were sequenced directly using forward or reverse primers. The sequences were compared with the reference sequence available from Ensembl (transcripts *ZRSR2* ENST00000307771, *U2AF1* ENST00000291552, and *SRSF2* ENST00000392485). All mutations were confirmed in nonamplified genomic DNA in an independent experiment. The somatic or germline status of mutations in *U2AF1* and *SRSF2* was established by evaluating T cells (CD3+CD11b−CD14−CD33−) purified from diagnostic samples by flow cytometry.

Next-generation sequencing for patients from Paris (n=35) was as follows:[6] final quality was checked with the Agilent 2100 Bioanalyzer and Agilent High Sensitivity DNA Assay, and the quantity by using the QubitTM dsDNA HS Assay Kit. Libraries were pooled for multiplex sequencing on a NextSeq500, Illumina, with High or Mid Output Kit v2 300 cycles to ensure a theoretical minimum depth of 4000X. Bioinformatic pipeline was developed to analyze sequencing data in order to control each step of analysis (SMPHD 2.8).
First, demultiplexing step with bcl2fastq (version v2.20.0.422) was done. Then, the quality of FASTQ files was checked using Fastqc tool (version 0.11.8). For each paired of FASTQ files, alignment against reference genome hg19 (2013) was done with BWA (bwa 0.7.17-r1188). A BAM file was generated using samtools 1.9 and was intersected to the BED file of the sequenced panel using bedtools v2.27 to keep reads in the region of the panel. After this step, duplicate reads were tagged but not removed using Mark duplicates (dupmark 2.18.23-SNAPSHOT. Finally, coverage was analyzed for each panel gene with bedtools coverage (v2.27.1-65-gc2af1e7-dirty -sorted).
An R script allowed to generate and report coverage quality in a html file. This file provided median and mean of coverage, percent of bases upper to 200X and exons with a coverage below 500X in the panel. Four different tools (GATK with methods of Haplotype caller, Mutect2 (provided by GATK to detect somatic mutation), Varscan (general) and Pindel (only for long deletions or insertions) were used for variant calling in order to detect mutations with optimal accuracy. Details of tools and parameters are as follows: GATK Haplotype caller (4.1.2.0) was used following the GATK best practices. A base recalibration with dbSNP138 was performed. Haplotype caller was launched with parameters --min-base-quality-score 30 and --dont-use-soft-clipped-bases true. Mutect2 (4.1.2.0) is more precise than GATK haplotype caller to detect somatic mutation and was used with the same parameters than for GATK HC. Varscan (2.4.3). Primary, we used samtools mpileup with default value parameter (-Q 13 -q 0 -A -B -d 100000). Then, Varscan calling was performed with the following parameters: min coverage 50, min-reads2 8, min-avg-qual 30, min var-freq 0.02 --p-value 0.1 --strand-filter 0. Pindel (0.2.5) used a primary bam file with unalignement reads. As parameter of length of variant, we used a 500 pb-max insertion or deletion. Software Annovar (version 2018-04-16) was used to annotate all called variants. Annovar allows

to download main database of annotation. For annotation, COSMIC90 (Release 5 September 2019) Catalogue Of Somatic Mutations In Cancer COSMIC 89 (Release) to validate and support new version of cosmic 90, gnomad 2.1.1 Exomes (Update, March 6, 2019): to know variant frequency in world population, and dbnsfp35a (20180921): Score columns like (whole-exome SIFT, PolyPhen2 HDIV, PolyPhen2 HVAR, LRT, MutationTaster, MutationAssessor, FATHMM, PROVEAN, MetaSVM, MetaLR, VEST, M-CAP, CADD, GERP++, DANN, fathmm-MKL, Eigen, GenoCanyon, fitCons, PhyloP and SiPhy scores from dbNSFP) to classify impact of mutation, clinvar (20190305) Clinical Data, cytoBand to know position in chromosome, IARC *TP53* Database July 2019, a WHO database of *TP53* mutations were used. Then, a vcf for each variant calling tool was used.

Finally, pandas (a module of python3.6.9) to merge the four files was used, to calculate the variant allele frequency (mean of VAF for each caller) and to filter mutations. The filters applied were as follows:, intronic and synonymous mutations were removed, variants with VAF < 2% were removed, variants with a minor allele frequency (MAF) ≥ 1% in a gnomad population were removed, variants known as recurring artifact (manually curated and stocked in a local database) were removed.

For cohorts from Cleveland and Nancy (n=53), DNA was extracted from whole blood or mononuclear cells and subjected to targeted sequencing using various next generation sequencing platforms as previously described[7] and according to manufacturer's protocols. Variants with a minimum variant allelic frequency (VAF) of 2%, an average of >500X coverage, and >98% of targeted regions showing over 100X coverage were annotated using Annovar, and their somatic status was called using an in-house bio-analytic pipeline, by comparison with public sequences of healthy subjects and databases: dbSNP 138, 1000 Genomes Project, NHLBI Exome Sequencing Project 6500, Exome Aggregation Consortium (ExAC) and genome aggregation database (gnomAD). VAF were adjusted according to zygosity and copy number based on conventional metaphase karyotyping and/or single nucleotide polymorphism array results.

Finally, retained variants were reviewed by molecular biologists from each center. Cases of discrepancies were discussed in concertation meetings.

References

1. Guglielmelli P, Lasho TL, Rotunno G, et al. MIPSS70: Mutation-Enhanced International Prognostic Score System for Transplantation-Age Patients With Primary Myelofibrosis. J Clin Oncol 2018;36:310–8.

2. Kröger N, Panagiota V, Badbaran A, et al. Impact of Molecular Genetics on Outcome in Myelofibrosis Patients after Allogeneic Stem Cell Transplantation. Biol Blood Marrow Transplant 2017;23:1095–101.

3. Woo J, Choi DR, Storer BE, et al. Impact of clinical, cytogenetic, and molecular profiles on long-term survival after transplantation in patients with chronic myelomonocytic leukemia. Haematologica 2020;105:652–60.

4. Thol F, Bollin R, Gehlhaar M, et al. Mutations in the cohesin complex in acute myeloid leukemia: clinical and prognostic implications. Blood 2014;123:914–20.

5. Panagiota V, Thol F, Markus B, et al. Prognostic effect of calreticulin mutations in patients with myelofibrosis after allogeneic hematopoietic stem cell transplantation. Leukemia 2014;28:1552–5.
6. Luque Paz D, Riou J, Verger E, et al. Genomic analysis of primary and secondary myelofibrosis redefines the prognostic impact of ASXL1 mutations: a FIM study. Blood Advances 2021;5:1442–51.
7. Gurnari C, Pagliuca S, Prata PH, et al. Clinical and Molecular Determinants of Clonal Evolution in Aplastic Anemia and Paroxysmal Nocturnal Hemoglobinuria. Journal of Clinical Oncology 2023;41:132–42.