Mutated *ASXL1* and number of somatic mutations as possible indicators of progression to chronic myelomonocytic leukemia of myelodysplastic syndromes with single or multilineage dysplasia

Ana Valencia-Martinez,[1] Alessandro Sanna,[1] Erico Masala,[1] Elisa Contini,[2] Alice Brogi,[1,3] Antonella Gozzini[1] and Valeria Santini[1]

[1]Department of Hematology, Università degli Studi di Firenze, AOU Careggi, Florence; [2]SOD Diagnostica Genetica and [3]Department of Biotechnologies, University of Siena, Italy

Correspondence: valeria.santini@unifi.it
doi:10.3324/haematol.2017.166124

Supplementary data for methods

The SureDesing of Agilent Technologies was used to design a custom enrichment of the candidate genes: *ASXL1*, *EZH2*, *TP53*, *SF3B1*, *U2AF1*, *SRSF2*, *TET2*, *DNMT3A*, *ETV6*, *RUNX1*, *NPM1*, *FLT3*, *CBL*, *SETBP1*, *CSF3R*, *CEBPA*, *IDH1*, *IDH2*, *JAK2*, *MPL*, *CARL*, *NRAS*. Genes were selected after a manually curated literature screening of the most commonly mutated genes in MDS and CMML. Library preparation was performed using the HaloPlex target enrichment protocol (Agilent technologies, Santa Clara, USA). The genomic DNA input for amplicon library preparation was 225 ng for each sample according to manufacturer's instructions. All sample libraries were equimolarly pooled and sequenced on the Illumina MiSeq Sequencer (Illumina, San Diego, CA, USA) with a default 150 bp paired-end reads protocol, according to the manufacturer's instructions.

Data analysis and processing

Reads quality was checked with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and reads were aligned to the reference human genome hg19 with BWA mem [Li, Heng. "Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM." arXiv preprint arXiv:1303.3997 (2013)].Genome Analysis Toolkit (GATK) version 3.3 was used to recalibrate base qualities and realign aligned reads around indels. Coverage statistics was performed by DepthOfCoverage utility of GATK. BASH and R custom scripts were used to obtain the list of low coverage (≤100X) regions per sample.

For the detection of low abundance variants: SNVs were identified using MuTect version 1.1.4 with standard parameters and GATK IndelGenotyperV2 was used to detect InDels. Genomic and functional annotation of detected variants was made by Annotate Variation Software (ANNOVAR, version 14 Dec, 2015). Detected variants were distilled on the basis of their exonic function, allele frequency, the presence in variants databases, such as the current release of dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/), 1000 Genomes Project (http://ftp.ncbi.nih.gov/), Exome Sequencing Project (http://evs.gs.washington.edu/EVS/) and several prediction scores (SIFT, MutationTaster, PolyPhen2, Gerp and PhyloP). Allelic depths for the reference and altered alleles were used to calculate the variant allele frequency (VAF) of all variants. All variants showing a VAF >5% were extracted and annotated with ANNOVAR. Candidate somatic variants were filtered using the following criteria:

1. Removal of all synonymous variants and noncoding variants more than 6 bases from splice junctions were discarded;

2. Removal of missense SNVs which were positioned in coding regions and filtered as polymorphisms when corresponding to known SNPs in public and private databases, including dbSNP144, ESP6500 (version march 2015), the 1000 genomes project (version 1 000g2015aug), and our in-house database at a population frequency of 1% or more;

3. Removal of variants present within regions prone to sequence context specific artifacts, including regions enriched for reads of low mapping quality that harbor multiple mismatches;

4. Removal of variants generated by incorrect calling after a manually revision using The Integrative Genomics Viewer (IGV)

5. Retention of variants present in the COSMIC (v70), ClinVar (v 20150629) and in-house databases as known somatic variation in myeloid and other common malignancies.

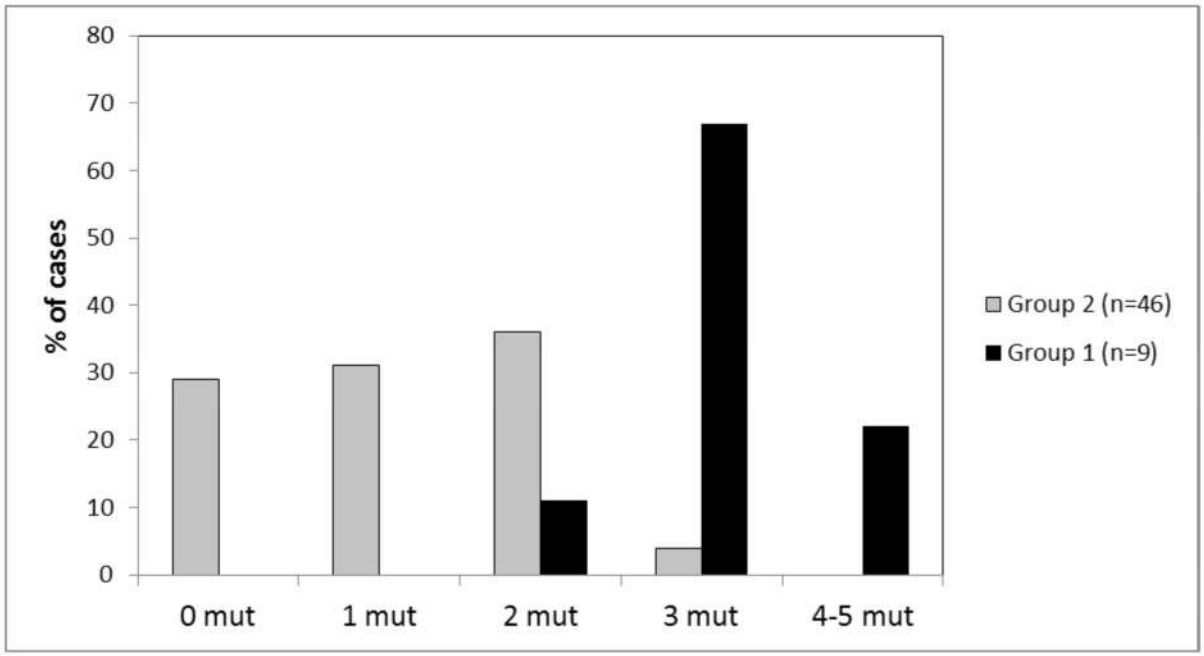
6. Remaining variants were the somatic origin was confirmed after and accurate and extensive research of the PubMed database (www.pubmed.gov).

Supplementary Table 1. Clinical and cytogenetic characteristics of patients of Group 2

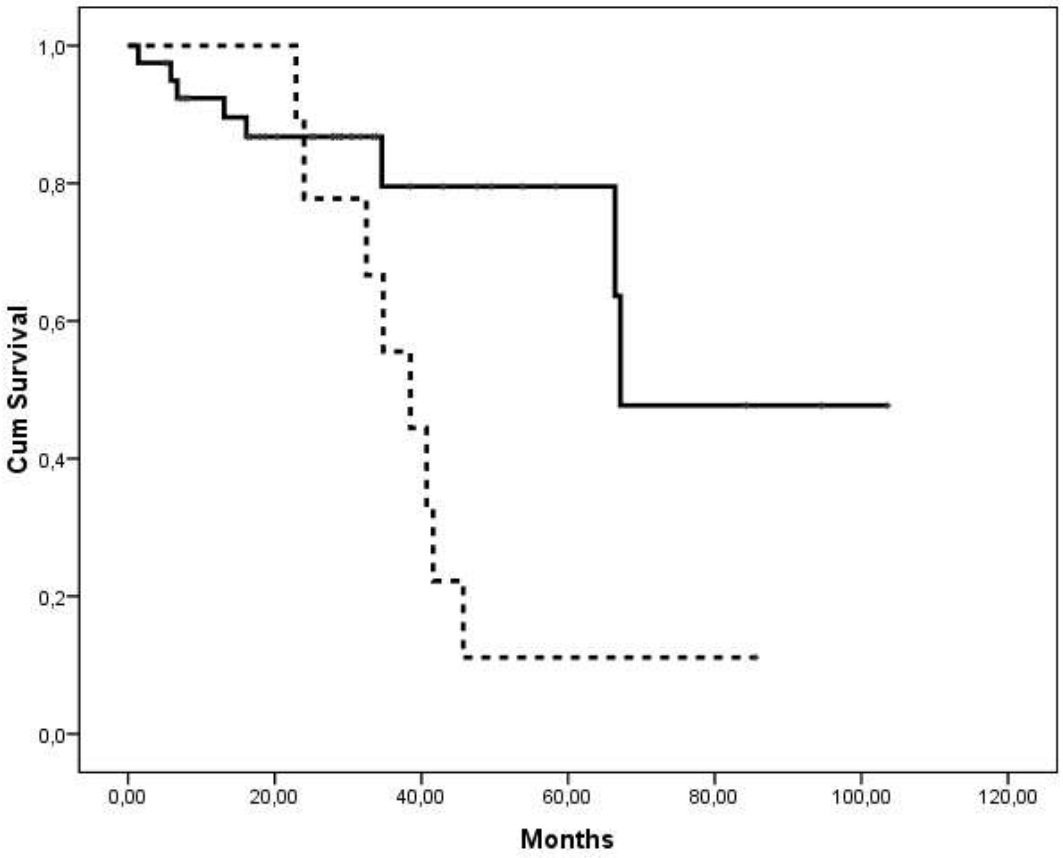
Characteristics	*Median (range)*	*n = 46*	*%*
Age, *years*	65 (43-88)		
Sex			
Male		26	56
Female		20	44
WBC count *(x10^9/L)*	4.15 (2.7-15.7)		
ANC count *(x10^9/L)*	2.3 (0.8-14.8)		
AMC count *(x10^9/L)*	0.4 (0.02-0.70)		
Haemoglobin *(g/dL)*	10.25 (8.1-13.11)		
Platelets *(x109/L)*	140 (17-722)		
WHO Classification			
MDS-SLD		17	37
MDS-RS-SLD		2	4
MDS-MLD		24	52
del(5q)		3	7
Cytogenetics			
Normal		32	69
Aberrant		14	31
IPSS risk group			
Low		28	61
Intermediate-1		18	39
IPSS-R risk group			
Very Low		17	36
Low		20	42
Intermediate		9	19

Abbreviations: AMC, absolute monocyte count; WHO, world health organization; MDS-SLD, myelodyplastic syndrome with single lineage dysplasia; MDS-RS-SLD, myelodyplastic syndrome-ring syderoblasts with single lineage dysplasia; MDS-MLD, myelodyplastic syndrome with multilineage lineage dysplasia; IPSS, international prognostic score system; IPSS-R, revised international prognostic score system

Supplementary Figure 1. Distribution of the number of mutations at diagnosis identified in Group 1 (black bars) and 2 (grey bars). The median number of mutations in Group 1 and 2 was 3 mutations (range 2-5) and 1 mutation (range 1-3), respectively ($P<0.001$).

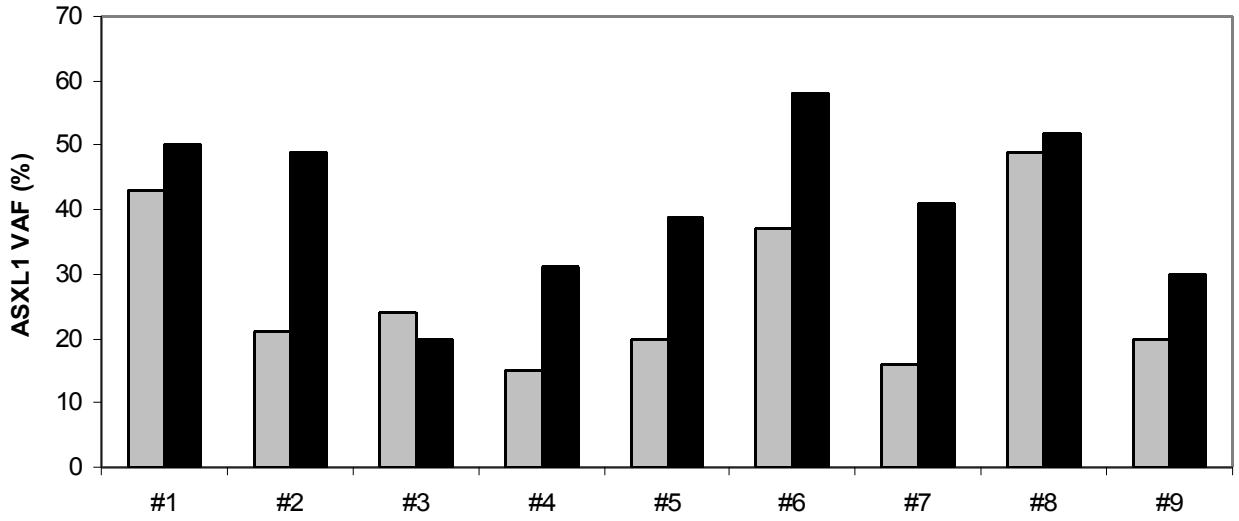


Supplementary Figure 2. Overall survival curves for Group 1 (dashed line) and 2 (solid line): median OS values, 38 months *vs.* 67 months, respectively; P=0.015.



Supplementary Figure 3. (A) VAF value of *ASXL1* and (B) *TET2* mutations in case #1 to #9 of Group 1 of patients in the MDS phase (grey bars) and CMML phase (black bars). (A) The median VAF value of *ASXL1* mutations was significantly higher in CMML than in the corresponding MDS phase (42% *vs.* 29%, respectively; *P*= 0.039) while the median VAF value of *TET2* mutations (B) did not significantly varied in both phases (46% *vs.* 55%; *P*=0.38).

A



B