SUPPLEMENTARY INFORMATION

MOLECULAR AND CYTOGENETIC CHARACTERIZATION OF MYELODYSPLASTIC SYNDROMES IN CELL-FREE DNA

Nieves García-Gisbert¹,², Sara García-Ávila¹,³, Brayan Merchán¹, Marta Salido⁴,⁵, Concepción Fernández-Rodríguez C¹,⁵, Joan Gibert¹, Lierni Fernández-Ibarrondo¹,², Laura Camacho¹,², Marta Lafuente¹,², Raquel Longarón¹,⁵, Blanca Espinet⁴,⁵, Patricía Vélez¹,³, Ramon M Pujo², Marcio Andrade-Campos¹, Leonor Arenillas⁴,⁵, Antonio Salar¹,³, Xavier Calvo⁴,⁵, Carles Besses¹, Beatriz Bellosillo¹,³

¹Group of Applied Clinical Research in Hematology, IMIM-Hospital del Mar, Barcelona, Spain ²Pompeu Fabra University, Barcelona, Spain ³Department of Hematology, Hospital del Mar, Barcelona, Spain ⁴Group of Translational Research on Hematological Neoplasms, IMIM-Hospital del Mar, Barcelona, Spain ⁵Department of Pathology, Hospital del Mar, Barcelona, Spain ⁶Department of Dermatology, Hospital del Mar, Barcelona, Spain

Supplementary Table 1. Classification of AML patients included in the study.

| PATIENT | WHO 2017 CLASSIFICATION (Arber et al, Blood, 2016) |
|---------|---------------------------------------------------|
| 1       | AML with minimal differentiation                  |
| 2       | AML with myelodysplasia-related changes           |
| 3       | AML with myelodysplasia-related changes           |
| 4       | AML with recurrent genetic abnormalities (biallelic mutations of CEBPA) |
| 5       | AML with recurrent genetic abnormalities (NPM1 mutated) |
| 6       | AML with recurrent genetic abnormalities (NPM1 mutated) |
| 7       | AML with recurrent genetic abnormalities (NPM1 mutated) |
| 8       | AML with recurrent genetic abnormalities (NPM1 mutated) |
| 9       | AML with recurrent genetic abnormalities (NPM1 mutated) |
| 10      | AML with recurrent genetic abnormalities, t(8;21)(q22;q22.1);RUNX1-RUNX1T1 |
| 11      | AML, NOS                                          |
| 12      | AML, NOS, Pure erythroid leukemia                 |
| 13      | AML, NOS, with maturation                         |
| 14      | AML, NOS, without maturation                      |
| 15      | AML, NOS, without maturation                      |
| 16      | Therapy-related myeloid neoplasm                  |
| 17      | Therapy-related myeloid neoplasm                  |
| 18      | Therapy-related myeloid neoplasm                  |
**Supplementary Table 2.** Genes and genomic regions covered by the NGS panel design (hg19).

| GENES   | CHROMOSOME LOCATION | COVERED REGION      |
|---------|---------------------|---------------------|
| ASXL1   | 20q11.21            | Full exonic region  |
| ATM     | 11q22.3             | Full exonic region  |
| BCOR    | Xp11.4              | Full exonic region  |
| BCORL1  | Xq26.1              | Full exonic region  |
| CALR    | 19p13.13            | Exon 9              |
| CBL     | 11q23.3             | Full exonic region  |
| CEBPA   | 19q13.11            | Full exonic region  |
| CHEK2   | 22q12.1             | Full exonic region  |
| CSF3R   | 1p34.3              | Full exonic region  |
| CSNK1A1 | 5q32                | Full exonic region  |
| CUX1    | 7q22.1              | Full exonic region  |
| DDX41   | 5q35.3              | Full exonic region  |
| DLEU7   | 13q14.3             | Full exonic region  |
| DNM1A   | 2q23.3              | Full exonic region  |
| EGR1    | 5q31.2              | Full exonic region  |
| ETV6    | 12p13.2             | Full exonic region  |
| EZH2    | 7q36.1              | Full exonic region  |
| FLT3    | 13q12.2             | Full exonic region  |
| GATA2   | 3q21.3              | Full exonic region  |
| IDH1    | 2q34                | Exon 4              |
| IDH2    | 15q26.1             | Exon 4              |
| JAK2    | 9p24.1              | Full exonic region  |
| KIT     | 4q12                | Exon 17             |
| KMT2A   | 11q23.3             | Full exonic region  |
| KRAS    | 12p12.1             | Full exonic region  |
| MPL     | 1p34.2              | Full exonic region  |
| NF1     | 17q11.2             | Full exonic region  |
| NPM1    | 5q35.1              | Full exonic region  |
| NRAS    | 1p13.2              | Full exonic region  |
| PHF6    | Xq26.2              | Full exonic region  |
| PPM1D   | 17q23.2             | Full exonic region  |
| PRPF8   | 17p13.3             | Full exonic region  |
| PTPN11  | 12q24.13            | Full exonic region  |
| RAD21   | 8q24.11             | Full exonic region  |
| RUNX1   | 21q22.12            | Full exonic region  |
| SETBP1  | 18q12.3             | Full exonic region  |
| SF3B1   | 2q33.1              | Exons 14,15,16     |
| SH2B3   | 12q24.12            | Full exonic region  |
| SRSF2   | 17q25.1             | Full exonic region  |
| STAG2   | Xq25                | Full exonic region  |
| TET2    | 4q24                | Full exonic region  |
| TNFSF11 | 13q14.11            | Full exonic region  |
| TP53    | 17p13.1             | Full exonic region  |
| TP53RK  | 20q13.12            | Full exonic region  |
| TP53TG5 | 20q13.12            | Full exonic region  |
| U2AF1   | 21q22.3             | Full exonic region  |
| WT1     | 11p13               | Full exonic region  |
| ZRSR2   | Xp22.2              | Full exonic region  |

**Polymorphic region close to EGR1**
- 5q31.2  chr5:137805574-137805662

**Polymorphic region in locus D7S486**
- (1) 7q3.1  chr7:115814732-115815082
- (2) 7q3.1  chr7:115825276-115825301
- (3) 7q3.1  chr7:115900252-115900830
- (4) 7q3.1  chr7:115948439-115949024
- (5) 7q3.1  chr7:115953508-115953582
**Supplementary Fig 1.** Sample workflow for DNA extraction and mutational analysis.

**Supplementary Fig 2.** Correlation between percentage of ring sideroblasts in bone marrow and VAFs of SF3B1 mutations in BM and cfDNA.

**Supplementary Fig 3.** CNV results by NGS in a patient with 20q- y 5q- alterations. Results of the coverage analysis of EGR1 (chr5) and TP53TG5 (chr20) genes are shown, which were included in the design of the NGS gene panel to cover chr5 and chr20 chromosomal aberrations. Each dot in the plot represents a genomic region covered by the gene panel. The green line shows the normal values (two copies of the genomic region). Dots above 2 indicate a potential gain of genetic material and dots below 2 indicate a potential loss of genetic material. A) CNV analysis of EGR1 and TP53TG5 in BM DNA. B) CNV analysis of EGR1 and TP53TG5 in cfDNA C) CNV analysis of a patient with normal karyotype. D)CMA results confirming the 5q- and 20q- in the patient.
**Supplementary methods.** R code used in R 3.6.2 to create the figures. Required files to generate the figures and full list of variants identified is provided in Supplemental Data 1-3.

```r
#Imports ####
library(readr)
library(readxl)
library(tidyverse)
library(maftools)

#Mutation Data
MUTS_SMDs_BxLIQUIDA_TODOS <- read_delim("SupplementaryData1_Variants.tsv", ",", escape_double = FALSE, trim_ws = TRUE)

# Patient conditions
Listado_paciente_muestraSMD <- read_delim("SupplementaryData2_PatientList.csv", ";", escape_double = FALSE, col_types = cols(DNA_MO = col_character(), NHC = col_character(), cfDNA = col_character()), trim_ws = TRUE)
Listado_paciente_muestraSMD <- pivot_longer(Listado_paciente_muestraSMD, cols = cfDNA:DNA_MO)

#Sample name preprocessing
MUTS_SMDs_BxLIQUIDA_TODOS$Sample <- gsub("_[0-9]+.smCounter.anno", "", MUTS_SMDs_BxLIQUIDA_TODOS$Sample)
MUTS_SMDs_BxLIQUIDA_TODOS$Sample <- gsub("^[0-9]+\."", "", MUTS_SMDs_BxLIQUIDA_TODOS$Sample)

#Join Databases
MUTS_SMDs_BxLIQUIDA_TODOS <- left_join(MUTS_SMDs_BxLIQUIDA_TODOS, Listado_paciente_muestraSMD, by=c("Sample"="value"))
setdiff(Listado_paciente_muestraSMD$value, MUTS_SMDs_BxLIQUIDA_TODOS$Sample)
setdiff(MUTS_SMDs_BxLIQUIDA_TODOS$Sample, Listado_paciente_muestraSMD$value)

# Upload pathways for the oncoplot###
Genes_and_pathways <- read_delim("SupplementaryData3_Genes_and_pathways.csv", ";", escape_double = FALSE, col_names = FALSE, trim_ws = TRUE)
colnames(Genes_and_pathways) <- c("Genes", "Pathways")
Genes_and_pathways$PATHWAY <- as.factor(Genes_and_pathways$Pathways)
genes <- Genes_and_pathways$Genes

#Preprocessing to fin maftools import
MUTS_SMDs_BxLIQUIDA_TODOS$Consequence[grepl("missense_variant", MUTS_SMDs_BxLIQUIDA_TODOS$Consequence)] <- "Missense_Mutation"
MUTS_SMDs_BxLIQUIDA_TODOS$Consequence[grepl("start_lost", MUTS_SMDs_BxLIQUIDA_TODOS$Consequence)] <- "Translation_Start_Site"
MUTS_SMDs_BxLIQUIDA_TODOS$Consequence[grepl("stop_gained", MUTS_SMDs_BxLIQUIDA_TODOS$Consequence)] <- "Nonsense_Mutation"
MUTS_SMDs_BxLIQUIDA_TODOS$Consequence[grepl("stop_lost", MUTS_SMDs_BxLIQUIDA_TODOS$Consequence)] <- "Nonstop_Mutation"
MUTS_SMDs_BxLIQUIDA_TODOS$Consequence[grepl("splice_donor_variant", MUTS_SMDs_BxLIQUIDA_TODOS$Consequence)] <- "Splice_Site"
MUTS_SMDs_BxLIQUIDA_TODOS$Consequence[grepl("splice_acceptor_variant", MUTS_SMDs_BxLIQUIDA_TODOS$Consequence)] <- "Splice_Site"
MUTS_SMDs_BxLIQUIDA_TODOS$Consequence[grepl("splice_region_variant,intron_variant", MUTS_SMDs_BxLIQUIDA_TODOS$Consequence)] <- "Splice_Site"
MUTS_SMDs_BxLIQUIDA_TODOS$Consequence[grepl("splice_donor_variant,coding_sequence_variant", MUTS_SMDs_BxLIQUIDA_TODOS$Consequence)] <- "Splice_Site"
MUTS_SMDs_BxLIQUIDA_TODOS$Consequence[grepl("frameshift_variant", MUTS_SMDs_BxLIQUIDA_TODOS$Consequence)] <- "Frame_Shift_Inc"
MUTS_SMDs_BxLIQUIDA_TODOS$Consequence[grepl("splice_donor_variant", MUTS_SMDs_BxLIQUIDA_TODOS$Consequence)] <- "Splice_Site"
MUTS_SMDs_BxLIQUIDA_TODOS$Consequence[grepl("frameshift_variant", MUTS_SMDs_BxLIQUIDA_TODOS$Consequence)] <- "Frame_Shift_Del"
MUTS_SMDs_BxLIQUIDA_TODOS$Consequence[grepl("splice_donor_variant", MUTS_SMDs_BxLIQUIDA_TODOS$Consequence)] <- "In_Frame_Del"
MUTS_SMDs_BxLIQUIDA_TODOS$Consequence[grepl("splice_donor_variant", MUTS_SMDs_BxLIQUIDA_TODOS$Consequence)] <- "In_Frame_Del"

unique(MUTS_SMDs_BxLIQUIDA_TODOS$Consequence)
colnames(MUTS_SMDs_BxLIQUIDA_TODOS$Consequence)[1] <- "Variant_Type"
colnames(MUTS_SMDs_BxLIQUIDA_TODOS[5]) <- "Start_Position"
colnames(MUTS_SMDs_BxLIQUIDA_TODOS$End_Position) <- "Start_Position"
colnames(MUTS_SMDs_BxLIQUIDA_TODOS$Reference_Allele) <- gsub("^>.*\\/", "", MUTS_SMDs_BxLIQUIDA_TODOS$Variant)
colnames(MUTS_SMDs_BxLIQUIDA_TODOS$Tumor_Seq_Allele2) <- gsub("^>.*\\/", "", MUTS_SMDs_BxLIQUIDA_TODOS$Variant)
colnames(MUTS_SMDs_BxLIQUIDA_TODOS$Type) <- "Type"
colnames(MUTS_SMDs_BxLIQUIDA_TODOS$Type) <- "Type"
colnames(MUTS_SMDs_BxLIQUIDA_TODOS$Type) <- "Type"
colnames(MUTS_SMDs_BxLIQUIDA_TODOS$Type) <- "Type"
# Check samples for each condition
SMD_MO <- MUTS_SMDs_BxLIQUIDA_TODOS[MUTS_SMDs_BxLIQUIDA_TODOS$name == "DNA_MO",]
unique(SMD_MO$NHC)
unique(SMD_MO$Sample)

SMD_cfDNA <- MUTS_SMDs_BxLIQUIDA_TODOS[MUTS_SMDs_BxLIQUIDA_TODOS$name == "cfDNA",]
unique(SMD_cfDNA$NHC)
unique(SMD_cfDNA$Sample)

# Check if mutation is only tissue / only plasma or in both
only_tissue <- anti_join(SMD_MO, SMD_cfDNA, by=c("NHC"="NHC", "Hugo_Symbol"="Hugo_Symbol",  
"Start_Position"="Start_Position", "HGVSc"="HGVSc"))
only_tissue$Role <- "Tissue"
only_plasma <- anti_join(SMD_cfDNA, SMD_MO, by=c("NHC"="NHC", "Hugo_Symbol"="Hugo_Symbol",  
"Start_Position"="Start_Position", "HGVSc"="HGVSc"))
only_plasma$Role <- "Plasma"
paired_moving_muts <- rbind(only_tissue, only_plasma)

paired_staying <- anti_join(MUTS_SMDs_BxLIQUIDA_TODOS, paired_moving_muts, by=c("Sample"="Sample",  
"Hugo_Symbol"="Hugo_Symbol", "Start_Position"="Start_Position", "HGVSc"="HGVSc"))
paired_staying$Role <- "Stay"

# Merge and change ID to shared tissue / cfDNA identifier
MUTS_SMDs_BxLIQUIDA_TODOS <- rbind(paired_moving_muts, paired_staying)
colnames(MUTS_SMDs_BxLIQUIDA_TODOS)[1] <- "Sample"
colnames(MUTS_SMDs_BxLIQUIDA_TODOS)[80] <- "Tumor_Sample_Barcode"
MUTS_SMDs_BxLIQUIDA_TODOS$Hugo_Symbol[MUTS_SMDs_BxLIQUIDA_TODOS$HGVSp ==  
"NP_004963.1:p.Val617Phe"] <- "JAK2_p.V617F"

# Upload to maftools and check everything looks nice
SMDs <- read.maf(MUTS_SMDs_BxLIQUIDA_TODOS, isTCGA = F, clinicalData = Listado_paciente_muestraSMD)  
check <- SMDs@data
Sample_type_colors <- RColorBrewer::brewer.pal(n = 4, name = "ReDS")[c(2,4)]
CL_colors <- c(Sample_type_colors)
names(CL_colors) <- c("DNA_MO", "cfDNA")
CL_colors <- list(name = CL_colors)

vc_cols <- c("#A6CEE3","darkolivegreen3","cornflowerblue","#33A02C","coral","#A36D90","#FDBF6F","#A36D90",  
"firebrick3")
names(vc_cols) <- c(  
'Frame_Shift_Del',  
'Missense_Mutation',  
'Nonsense_Mutation',  
'Multi_Hit',  
'Frame_Shift_Ins',  
'In_Frame_Ins',  
'Splice_Site',  
'In_Frame_Del',  
'Translation_Start_Site'  
)
# Plot oncoplot
oncoplot(SMDs, colors = vc_cols, annotationColor = CL_colors, removeNonMutated = FALSE, pathways =  
Genes_and_pathways,  
top = 1000, fontSize = 0.7, sortByAnnotation = TRUE, anno_height = 0.5, SampleNameFontSize = 0.6,  
additionalFeature = list(c("Role", "Tissue"), c("Role", "Plasma")), annoBorderCol = 'black',  
additionalFeatureCol = c("black", "black"), additionalFeatureCex = 0.7, gene_mar = 9,  
additionalFeaturePch = c(0,15), showTumorSampleBarcodes = TRUE)
dev.off()