Supplementary materials

1. Supplementary methods

BMSC cultures

Mononuclear cells separated by Ficoll sedimentation, either from thawed cryopreserved bone marrow (BM) (n=32) or from fresh BM samples at follow-up (n=10), were seeded in T-25 cm^2 flasks with DMEM medium and 20% fetal bovine serum and cultured in a humidified atmosphere at 37°C containing 5% carbon dioxide. Forty-eight hours later supernatant was discarded and fresh medium replaced every 48-72h. When BMSC cultures formed colonies with a confluence of 80-100% they were transferred to a T-75 cm^2 flask using trypsin. BMSC were collected when they again reached 80-100% confluence.

DNA extraction, library preparation and sequencing

DNA was extracted with QIAgen mini-kit. DNA was quantified by fluorimetric Qubit technology. Quality was assessed with Fragment Analyzer® (Agilent) with a threshold of 10.000bp. Exome sequencing was performed at the Centro Nacional de Análisis Genómico (CNAG-CRG, Barcelona, Spain). Paired-end multiplex libraries were prepared according to manufacturer's instructions and enriched with the Kapa HyperExome DNA genome design exome kit. Libraries were loaded to Illumina flowcells for cluster generation prior to producing 100 base read pairs on a NovaSeq6000 instrument following the Illumina protocol. Image analysis, base calling and quality scoring of the run were processed using the manufacturer's software Real Time Analysis and followed by generation of FASTQ sequence files.

Data analysis

Reads were mapped to human build GRCh38.hl with BWA-MEM 0.7.17[1]. Alignment files containing only properly paired, uniquely mapping reads without duplicates were processed using Picard 2.20 [http://broadinstitute.github.io/picard/] to add read groups and to remove duplicates. The Genome Analysis Tool Kit (GATK 4.1.8.0)[2] was used for local realignment and base quality score recalibration. Variant calling was done using HaplotypeCaller from GATK. Functional annotations were added using SnpEff v.4.3e with the GRCh38.86 database[3]. Variants were annotated with SnpSift v4.3e[4] using

population frequencies, conservation scores and deleteriousness predictions from dbNSFPv4.1a[5]. Other sources of annotations, such as gnomAD, CADD and Clinvar were also used[6,7].

Somatic variant calling was performed with GATK v4.1.3 Mutect2 and processed following GATK best practices (https://software.broadinstitute.org/gatk/documentation/article?id=11136). In parallel, somatic calling was performed also with Strelka2 v2.9.9[8], feed with candidate small indels predicted by Manta v1.6.0[9]. Both somatic variant calling from Mutect2 and Strelka2 were functionally annotated with SnpEff v4.3.e and variants were annotated with SnpSift, as described before. Two purity estimates were obtained from Facets v0.5.6 [https://github.com/mskcc/facets] and Canvas[10]. Tumor contamination of germline samples was assessed by GATK CalculateContamination tool (version 4.1.4.0). Estimation of the oncogenic effect of variants in driver genes was predicted with OncodriveMUT[11].

Chimerism analysis

Chimerism analysis was performed in post-alloHCT BMSC samples using the Mentype® Chimera® PCR Amplification Kit, that analyzes 12 polymorphic autosomal Short Tandem Repeat (STR) loci. The results were compared to those recorded for both recipient and donor.

List of 288 interrogated genes (modified from[12])

ACD, ANKRD26, ATG2B, BLM, BRCA1, BRCA2, BRIP1, CBL, CEBPA, CTC1, DDX41, DKC1, DNAJC21, EFL1, ELANE, EPCAM, ERCC4, ERCC6L2 , ETV6, FANCA, FANCB, FANCC, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCL, FANCM, G6PC3, GATA1, GATA2, GFI1, GSKIP, HAX1, KRAS, MAD2L2, MLH1, MSH2, MSH6, NAF1, NBN, NF1, NHP2, PALB2, PARN, PMS2, PTPN11, RPL11, RPL15, RPL23, RPL27, RPL31, RPL35A, RPL36, RPL5, RPS10, RPS15, RPS17, RPS19, RPS24, RPS26, RPS27, RPS27A, RPS28, RPS29, RPS7, RTEL1, RUNX1, SAMD9, SAMD9L, SBDS, SRP72, TERC, TERT, TINF2, TP53, TSR2, UBE2T, USB1, WRAP53, XRCC2, ATM, DOCK8, FAS, GBA, ITK, LIG4, NOP10, PAX5, POT1, PRF1, RAD51, RAD51C, RMRP, SH2B3, SH2D1A, SLX4, STAT3, STN1, STX11, STXBP2, UNC13D, WAS, CASP10, CHEK2, JAK2, MECOM, MPL, PPM1D, RBBP6, ATRX, BTK, BUB1B, CD27, CDKN2A, CSF3R, GRHL2, HCLS1, LAPTM5, MAGT1, MAP2K1, MAP2K2, NPAT, NSD1, PIK3CD, RAF1, RIT1, SHOC2, SOS1, SPRED1, ABL1, AKT1, ARID1A, ARID1B, ASXL2, BCL2, BCL6, BCOR, BCORL1, BIRC3, BIRC6, BRAF, BRCC3, BRD4, BRINP3, CALR, CARD11, CBLB, CBLC, CCND1, CCND3, CCR4, CD40LG,

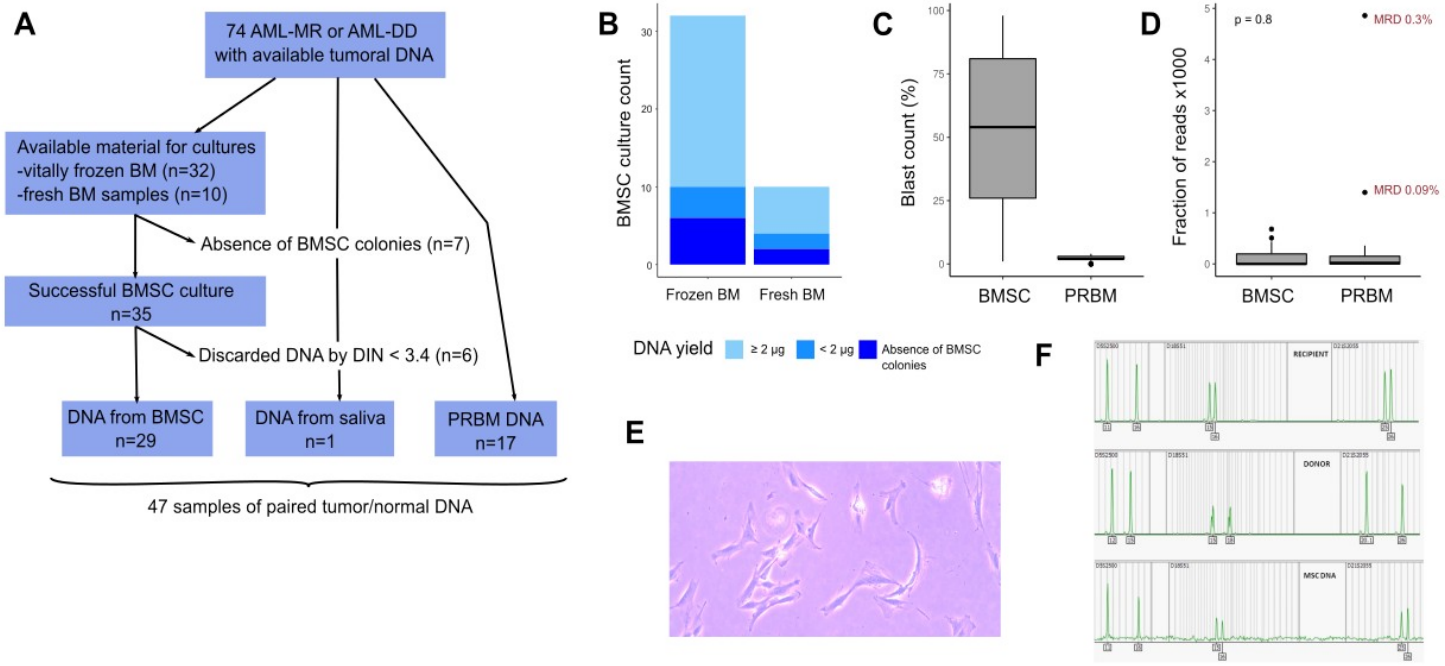
CD79A, CD79B, CDKN2C, CHD2, CREBBP, CRLF2, CSF1R, CTCF, CTLA4, CUX1, CXCR2, CXCR4, DAXX, DDX54, DHX15, DHX29, DIS3, DNAH5, DNAH9, DNM2, DNMT1, DTX1, EED, EGFR, EP300, ERBB4, ETNK1, EZH2, FAM47A, FAT1, FAT4, FBXO11, FBXW7, FLT3, FOXO1, FYN, GATA3, GNA13, GNAS, GNB1, H1-4, HNRNPK, HRAS, HVCN1, ID3, IDH1, IDH2, IGLL5, IKZF1, IKZF3, IL7R, IRF4, JAK1, JAK3, KDM6A, KIT, KLF2, KLHL6, KMT2A, KMT2C, KMT2D, LLGL2, LRRC4, LUC7L2, MAML1, MED12, MEF2B, MGA, MRTFA, MYC, MYD88, NFKBIE, NLRP2, NOTCH1, NOTCH2, NPM1, NRAS, NSD2, NT5C2, NXF1, PCLO, PDGFRA, PHF6, PIGA, PIM1, PLCG1, PLCG2, PRDM1, PRKCB, PRPF40B, PRPF8, PRPS1, PSMB5, PTCH1, PTEN, RAD21, RB1, RELN, RHOA, SAMHD1, SETBP1, SETD2, SETDB1, SF1, SF3A1, SF3B1, SMARCA2, SMARCB1, SMC1A, SMC3, SOCS1, SPEN, SPI1, SRSF2, STAG2, STAT5B, SUZ12, SYK, SYNE1, TBL1XR1, TCF3, TCF4, TNFAIP3, TNFRSF14, TPP1, TRAF3, U2AF1, U2AF2, UBR5, VAV1, WT1, XPO1, ZBTB7A, ZRSR2.

List of 42 recurrently mutated genes used for somatic mutation identification

ABL1, ASXL1, ATM, BCOR, BRAF, CALR, CBL, CEBPA, CSF3R, DNMT3A, DDX41, ETV6, EZH2, FLT3, GATA2, HRAS, IDH1, IDH2, IKZF1, JAK2, KIT, KRAS, MPL, MYD88, NF1, NPM1, NRAS, PHF6, PRPF8, PTPN11, RB1, RUNX1, SETBP1, SF3B1, SH2B3, SRSF2, STAG2, TET2, TP53, U2AF1, WT1, ZRSR2.

2. Supplementary Figures

Figure S1: A) Selection of cases according to normal DNA availability; B) DNA yield of bone marrow stroma cells (BMSC) cultures; C) Blast percentage according to morphological examination of the sample used to collect germline DNA divided by source type; D) Cross-sample fraction of reads x1000 estimated for BMSC and post-remission bone marrow (PRBM). The two outliers found in PRBM correspond to the only two patients with positive MRD as detected by flow cytometry. E) BMSC colony 7 days after trypsinization. F) Representative small tandem repeat microsatellite pattern of recipient, donor and BMSC DNA in a case where BMSC culture was derived from a follow-up bone marrow sample after allogeneic hematopoietic cell transplantation. The pattern is consistent with recipient-belonging BMSC.



Abbreviations: AML-MR: acute myeloid leukemia, myelodysplasia-related; AML-DD: acute myeloid leukemia defined by differentiation; BM: bone marrow; BMSC: bone marrow stromal cells; DIN: DNA integrity index; PRBM: post-remission bone marrow

Figure S2: Variant allele frequencies (VAF) of recurrently mutated genes in bone marrow at diagnosis and after remission from five cases with detection of any of the mutations at diagnosis in post-remission bone marrow (PRBM).

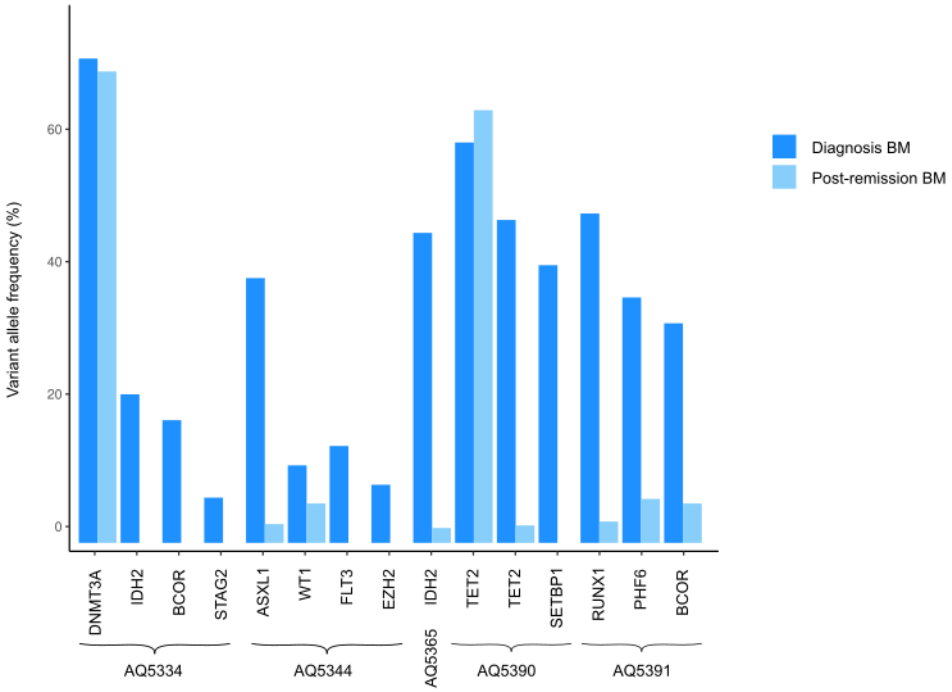
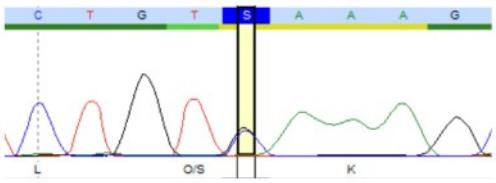


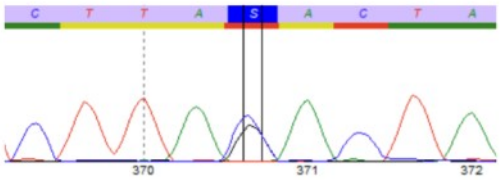
Figure S3: Sanger validation of *ATM* pathogenic variants



c.2672C>G; p.Ser891Ter

Primer code: Hs00473308_CE

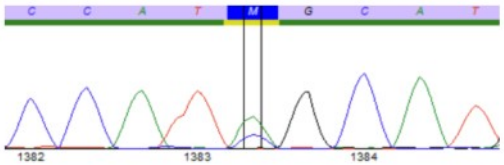
Primer position:
Chr.11: 108268355-108268546



c.1110C>G; p.Tyr370Ter

Primer code: Hs00525075_CE

Primer position:
Chr.11: 108248941-108249082



c.4148C>A; p.Ser1383Ter

Primer code: Hs00521465_CE

Primer position:
Chr.11: 108288987-108289120

3. Supplementary Tables

Table S4 (included separately as a xlsx file): Manually curated germline variants

Table S5: Comparison of clinical characteristics from germline variants carriers (n=11) vs non-carriers (n=36)

	Germline variant carriers (n=11)		**Absence of germline variants (n=36)**		
	n	**%**	**n**	**%**	**p-value**
Median age, years (range)	58		54		0.54
Sex (female/male)	6 / 5	55 / 45	20 / 16	56 / 44	1
AML-MR (WHO2022 / ICC 2022)	6 / 7	55 / 64	17 / 21	47 / 58	1 / 1
Previous myelodysplastic syndrome	0	0	5	14	0.32
Multilineage dysplasia	3	27	12	33	1
Normal karyotype	7	30	25	69	0.93
WBC (10^9/L), median (range)	3 (1.4-95)		6.2 (0.5-171)		0.59
BM blast count (%), median (range)	69 (10-86)		62 (20-98)		0.55
ELN 2022 - Adverse risk	9	82	26	72	0.7
Number of somatic mutations, mean (range)	2.6 (0-5)		2.6 (0-6)		0.92
Disease response after induction (n=44)					
Complete response	8	80	29	85	0.58

Abbreviations: AML-MR, acute myeloid leukemia myelodysplasia-related; AML-DD, acute myeloid leukemia defined by differentiation

References

1. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. **00**, 1–3 (2013).
2. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
3. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin).* **6**, 80–92 (2012).
4. Cingolani, P. *et al.* Using Drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Front. Genet.* **3**, 1–9 (2012).
5. Liu, X., Jian, X. & Boerwinkle, E. dbNSFP v2.0: A database of human non-synonymous SNVs and their functional predictions and annotations. *Hum. Mutat.* **34**, 2393–2402 (2013).
6. Landrum, M. J. *et al.* ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).
7. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
8. Kim, S. *et al.* Strelka2: fast and accurate calling of germline and somatic variants. *Nat. Methods* **15**, 591–594 (2018).
9. Chen, X. *et al.* Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).
10. Roller, E., Ivakhno, S., Lee, S., Royce, T. & Tanner, S. Canvas: Versatile and scalable detection of copy number variants. *Bioinformatics* **32**, 2375–2377 (2016).
11. Tamborero, D. *et al.* Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med.* **10**, 1–8 (2018).
12. Yang, F. *et al.* Identification and prioritization of myeloid malignancy germline variants in a large cohort of adult patients with AML. *Blood* **139**, 1208–1221 (2022).
13. Mujahed, H., Jansson, M., Bengtzén, S. & Lehamnn, S. Bone marrow stroma cells derived from mononuclear cells at diagnosis as a source of germline control DNA for determination of somatic mutations in acute myeloid leukemia. *Blood Cancer J.* **7**, 10–12 (2017).