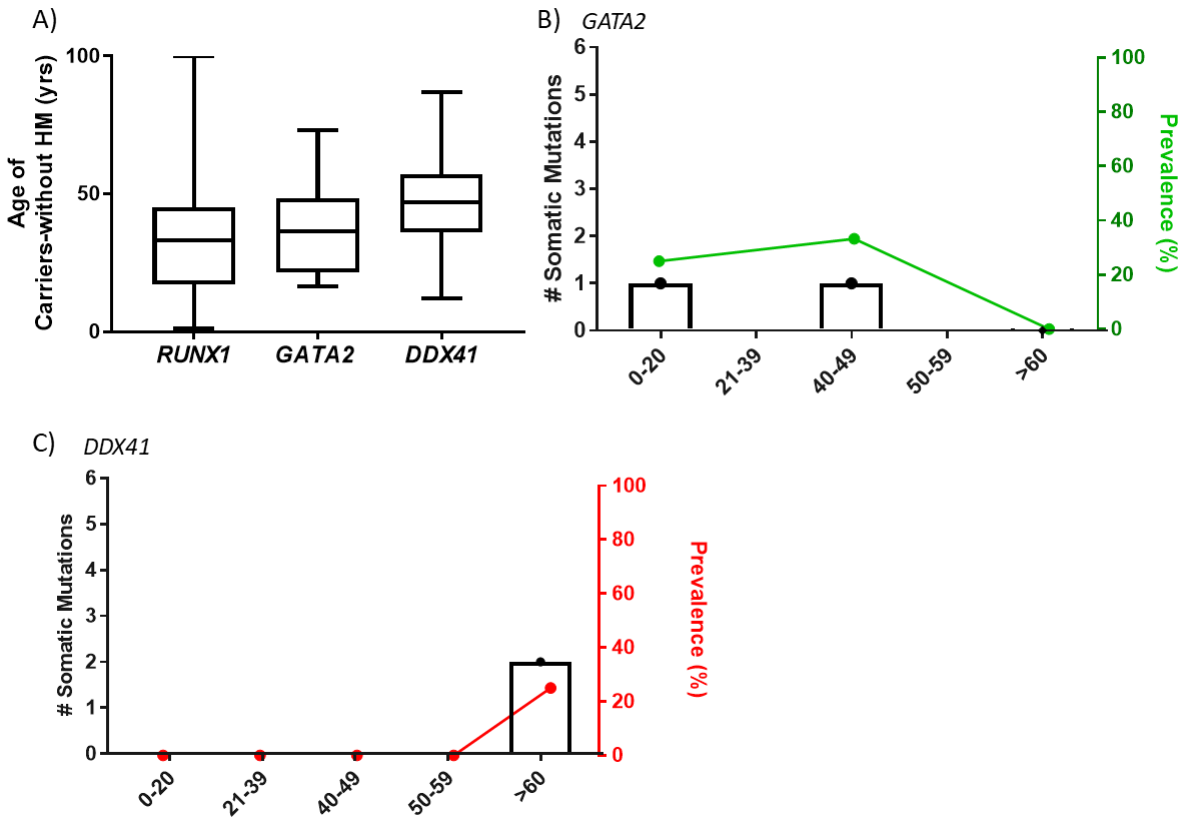
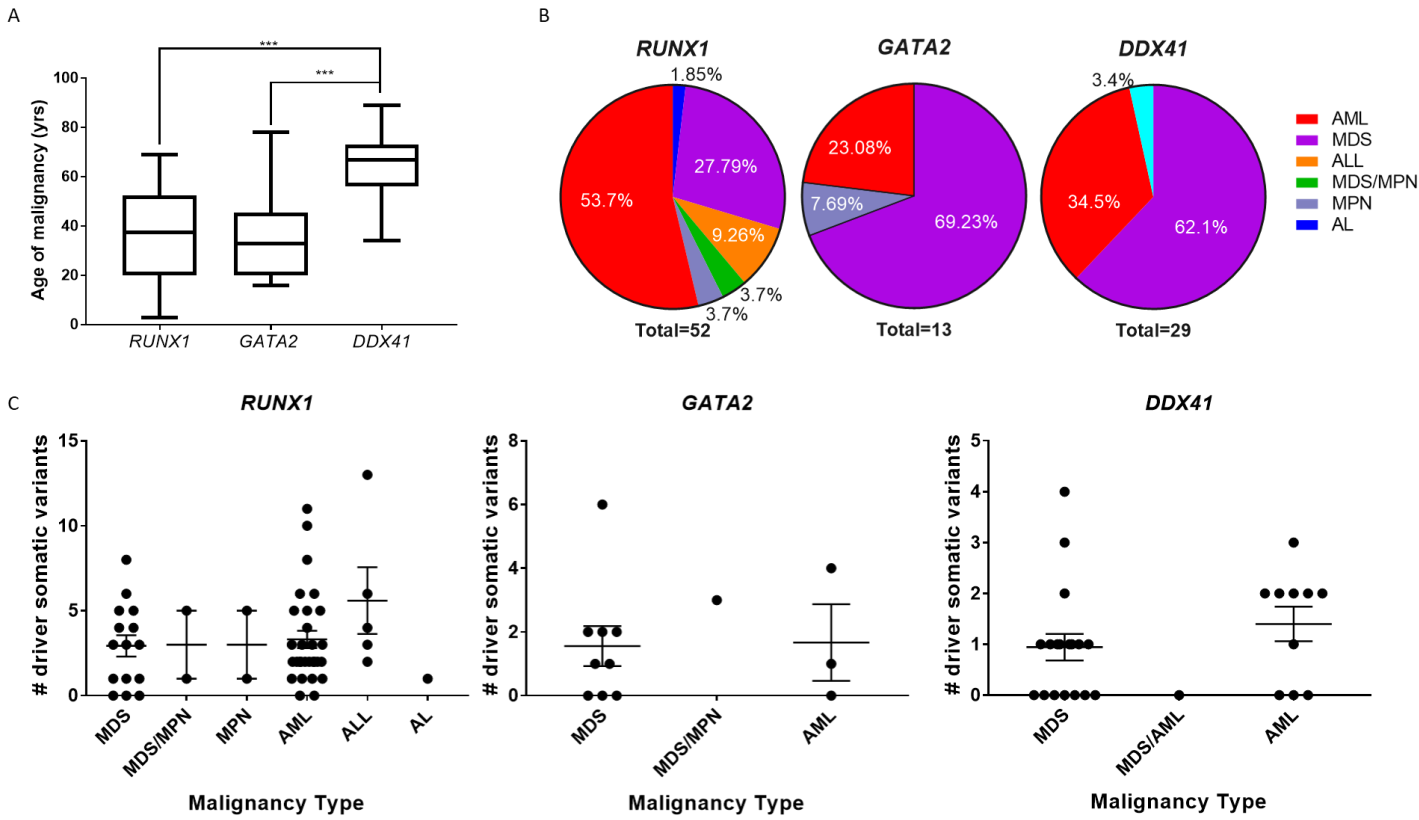
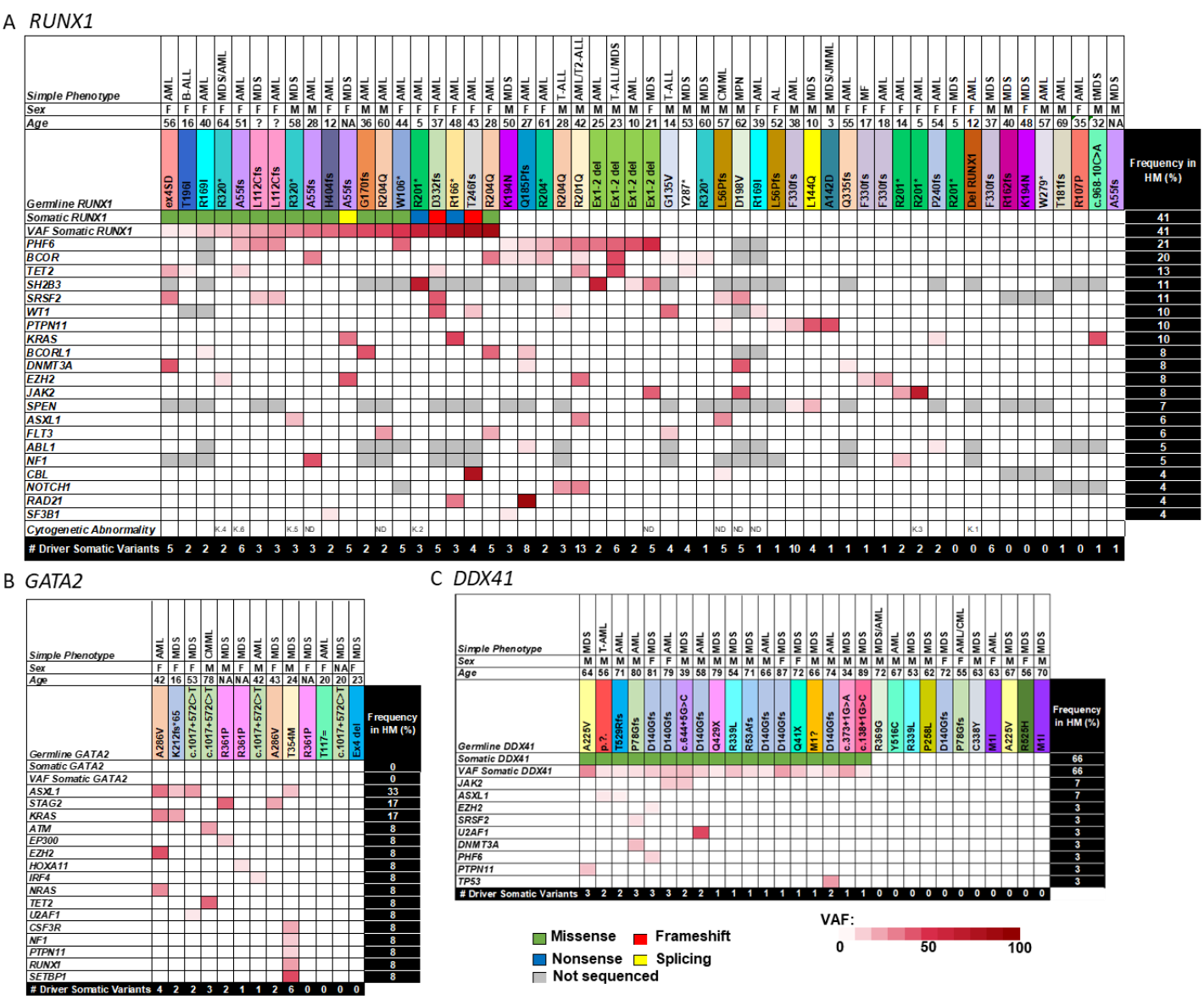


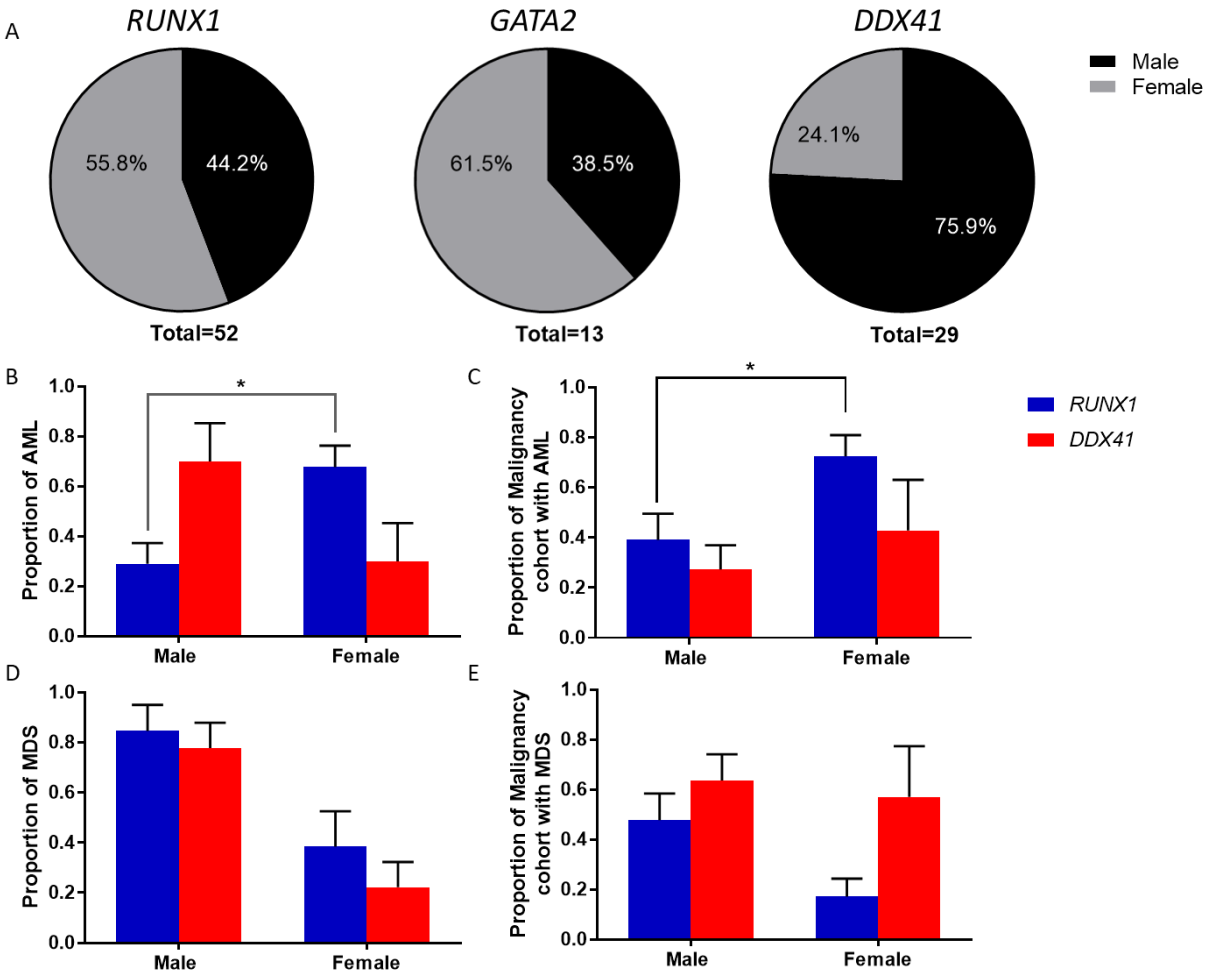
Supplementary Figure 1: Defining the spectrum of clonal hematopoiesis in germline *GATA2, RUNX1* and *DDX41* carrier-without HM cohorts. A) Age of individuals in the carrier-without HM cohorts. Graph shows box plot with the median age of individuals, error bars show the min and max value. Demographics of individuals with CH variants in the *GATA2* and *DDX41* germline carrier-without HM cohorts. B) germline *GATA2*; C) germline *DDX41*; carrier-without HM cohorts. Column graph shows the number of somatic variants identified in individuals with CH. Error bars show the SEM. Line graphs show the prevalence of CH in the carrier-without HM germline cohorts in different age groups, CH=clonal hematopoiesis



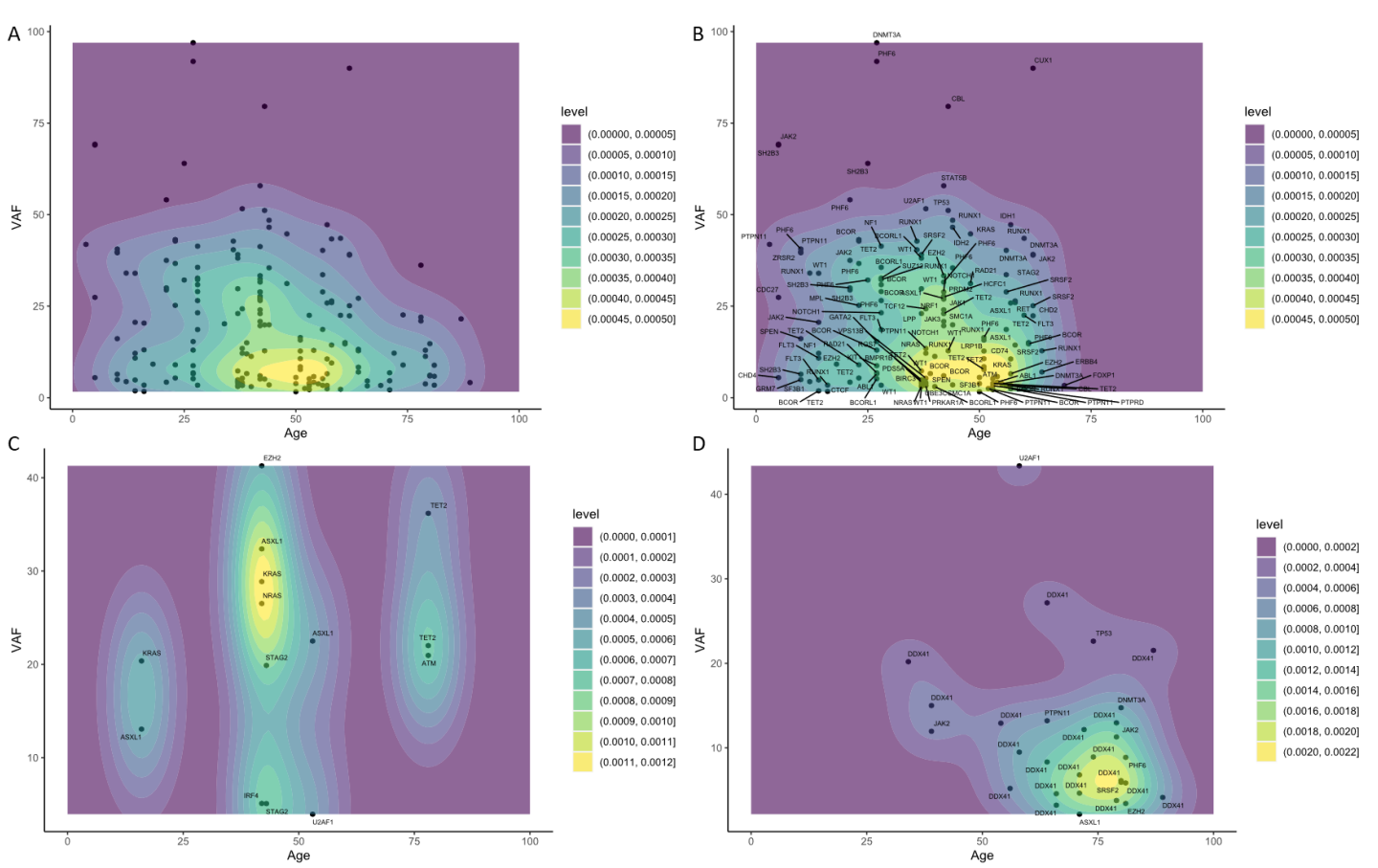
Supplementary Figure 2: Germline *RUNX1, GATA2* and *DDX41* carrier-with HM cohort demographics. A) Age of individuals with malignancy in the cohort. Graph shows box plot with the median age of malignancy, error bars show the min and max value. B) Pie charts showing the percentage of individuals with each malignancy presentation in the three germline cohorts (Broad HM categories are used, some individuals are represented twice in different disease stages). C) Breakdown of malignancy samples and the number of somatic mutations. Scatter plot with the mean and SEM displayed by the error bars. ***P<0.001 One-way ANOVA with Tukey's multiple comparison.



Supplementary Figure 3. Oncoplots showing clinically relevant somatic variants identified in the germline carrier-with HM cohorts. Distribution of the clinically relevant somatic variants "driver somatic variants" identified in the carrier-with HM cohorts. VAF of the driver somatic variant in the sample as represented on a sliding scale (darker=high VAF, lighter= low VAF). A) germline *RUNX1* carrier-with HM cohort. Only shown are the genes that are identified as somatically mutated in two or more individuals. B) germline *GATA2* carrier-with HM cohort, showing all driver somatic variants and C) germline *DDX41* carrier-with HM cohort showing all driver somatic variants. VAF= variant allele frequency, AML= acute myeloid leukemia, MDS= myeloid dysplastic syndrome, CML= chronic myeloid leukemia, CMML=Chronic myelomonocytic leukemia, B-ALL= B-cell acute lymphoblastic leukemia, T-ALL= T-cell acute lymphoblastic leukemia, MPN= Myeloproliferative neoplasms, AL= Acute leukemia, JMML= Juvenile myelomonocytic leukemia. Karyotypes are as follows: K.1- 48,XXXc,+21[28]/47,XXXc[3], K.2-21q gain (46,XX,der(21).ish amp(RUNX1)[15]/46,XX[5]), K.3 - 9q deletion (45,XX,-7[2]/45,XX,idem,del(9)(q21q31)[17]/46,XX[1]), K.4: Monosomy 7, K.5-der(1;7)(q10;p10), K.6- Monosomy 5

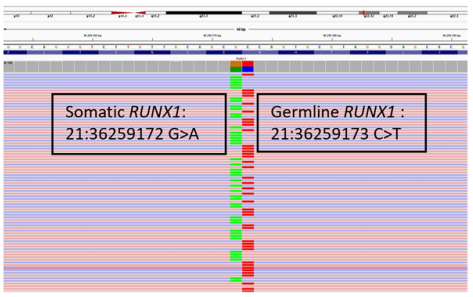


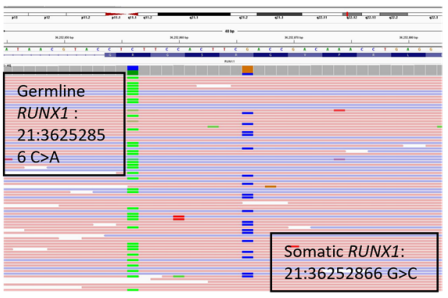
Supplementary Figure 4 : Gender difference associated with common haematological malignancy presentations in the germline *RUNX1, GATA2* and *DDX41* carrier-with HM cohorts. A) Gender distribution of the germline *RUNX1, GATA2* and *DDX41* carrier-with HM cohorts. B) Proportion of male and females presenting with AML. C) Proportion of AML presentation in the germline carrier-with HM cohorts. D) Proportion of males and females presenting with MDS. E) Proportion of MDS presentation in the germline carrier-with HM cohort. Blue= germline RUNX1, Red=germline DDX41. *P<0.01 fisher's exact test.

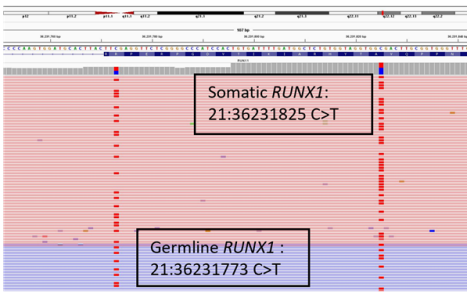


Supplementary Figure 5: Density plots showing the distribution of the age and VAF of somatic clinically relevant variants in the germline carrier-with HM cohorts. A) combined cohorts B) *RUNX1* carrier-with HM cohort C) *GATA2* carrier-with HM cohort and D) *DDX41* carrier-with HM cohort. Level is the 2D density estimate as calculated by the ggplot2 rpackage, geom_density_2D.

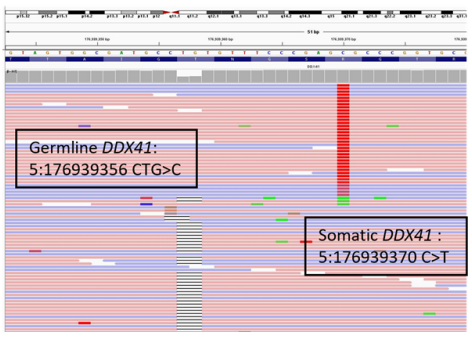
A



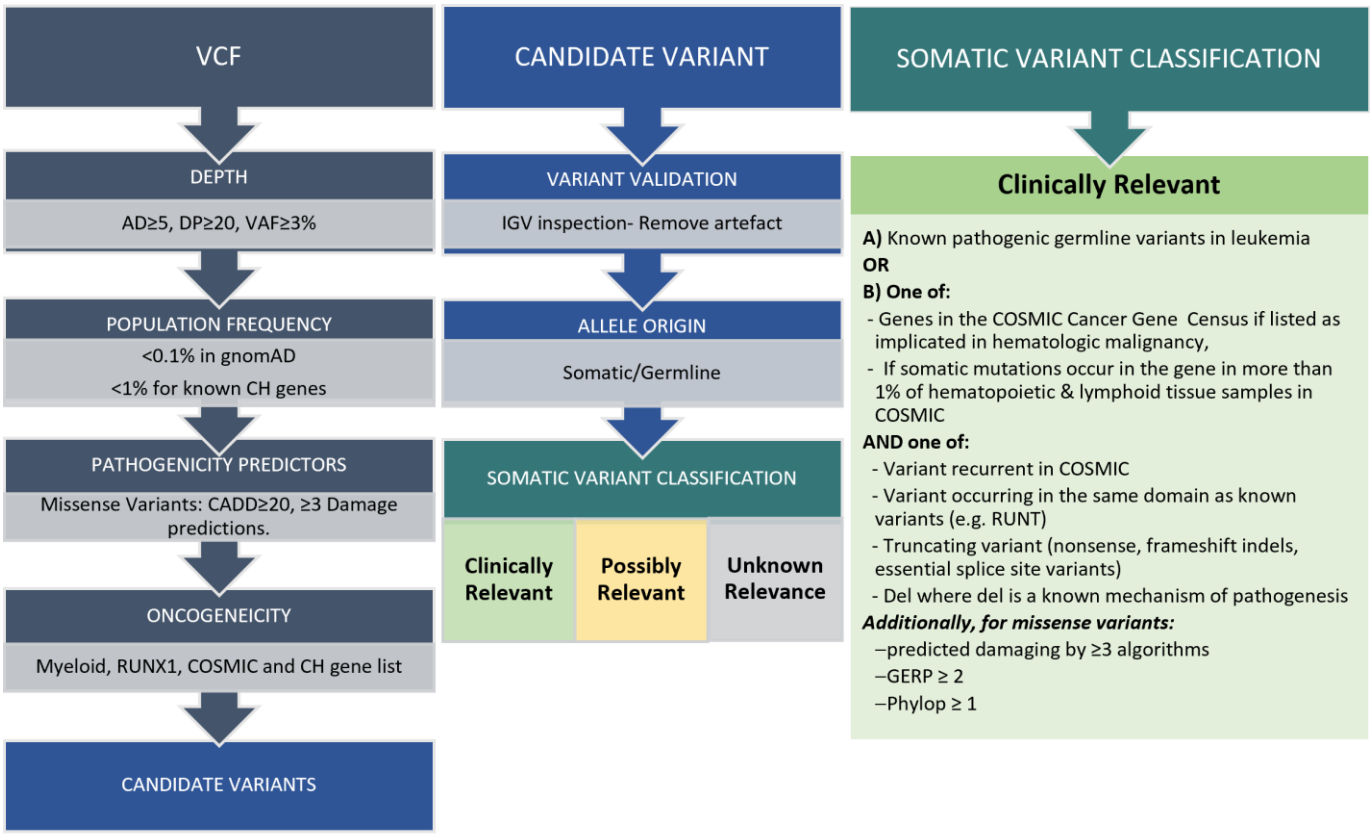




B



Supplementary Figure 6. Acquired *DDX41* or *RUNX1* variants are in trans with the germline variant. A) Interactive genome viewer (IGV) capture of panel sequencing from three germline *RUNX1* individuals where the germline and somatic variants are located in close proximity to be sequenced in the same DNA-read. Reads show that the germline *RUNX1* variant is in trans with the somatic *RUNX1* variant in all three individuals. B) IGV capture of panel sequencing from one germline *DDX41* individual where the germline and somatic variants are located in close proximity to be sequenced in the same DNA-read. Reads show that the germline *DDX41* variant is in trans with the somatic *DDX41* mutation.



Supplementary Figure 7. Somatic variant filtering and curation pipeline. The same curation approach was used as designed for the RUNX1 database[20]. Oncogenicity filter gene lists are included in supplementary table 2.

Supplementary Methods:

Inclusion criteria for germline cohorts Individuals with germline *RUNX1, GATA2* and *DDX41* variants were included in the cohorts based on the germline variant being classified according to ACMG and Myeloid Malignancy Variant Curation Expert Panel Criteria[59,60]as either Pathogenic, Likely Pathogenic, or a Variant of Uncertain Significance (VUS) with clinical, familial and/or phenotypic data which strongly suggests causation for the HHM syndrome. Germline status was ascertained with matched germline tissue and/or familial samples or as determined for the RUNX1db[22]. Our dataset included germline *RUNX1* patients previously published.[13,19,61–71]

Inclusion criteria for VUS 3A: MM-VCEP rules for classifying germline *RUNX1* variants are continually evolving, and the following VUS have been included as they are highly suspicious VUS that would be considered for further clinical follow-up in our diagnostic laboratories. Details in addition to the applied ACMG criteria (**Supplementary Table 1**) are as follows:

NP_001745.2:p.Gly135Val and NP_001745.2:p.Gly135Ser: Both variants are located in amino acid 135 of RUNT domain (3 probands meeting FPD-MM phenotypic criteria). Somatic variants have been observed in the COSMIC database 4 times including in HM and Breast cancer. p.Gly135Val HM samples, has a germline *RUNX1* mutational signature including a *BCOR* somatic variant.

NM_001754.4:c.98-1G>A: This variant is an acceptor loss variant with a SpliceAI score of 0.75 and only affects isoform c (PVS1). Frequent germline deletions associated with a FPD-MM phenotype are predicted to affect only the RUNX1c isoform suggesting that this is the predominant "oncogenic" isoform (ClinVar:650703).

NP_001745.2:p.Ala142Asp: Located within the RUNT domain, between two DNA binding molecules. Protein modelling predicts decrease of molecule flexibility(ΔΔSVib ENCoM: -0.482 kcal.mol-1.K-1) and Protein de-stabilization.

carrier-without HM status was assigned to individuals that at the time of sample collection did not meet clinical criteria for hematological neoplasms as defined by The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms (Leukemia. 2022).

Somatic variant curation

VariantGrid analysis software and somatic variant curation pipeline was utilised as in the RUNX1 database[22]. The somatic variant filtering pipeline is as follows: 1) Sample Filter: AD≥5, DP≥20, VAF≥3%. 2) Population Filter: Max population frequency of 0.1% in gnomAD (selected populations: African/African American, East Asian, Latino/Mixed Amerindian, non-Finnish European, South Asian), 1.0% for known clonal hematopoiesis genes. 3) Damage Filter: Impact minimum=moderate, CADD score ≥20, Minimum 3 damage predictions (missense variants), allow null (frameshift considered damaging) and keep splice variants. 4) Oncogenicity Filters: 1. Myeloid, RUNX1 or COSMIC Gene Census gene lists. 2. Clonal hematopoiesis gene lists. 3. COSMIC variants, with gene observed at >3% in hematopoietic and/or lymphoid tissue. Variants which passed all filtering criteria were subsequently manually curated and confirmed as real with visual IGV inspection. Somatic variants were identified with a matched familial sample or a germline sample. If none was available variant was classified somatic if VAF <30% or if VAF >30%, then <0.001% gnomAD population frequency and/or somatic in COSMIC in >2 samples. Variants were then curated as somatic clinically relevant ("driver somatic variants") (**Supplementary Figure 7**).

Curation of somatic *DDX41* and *RUNX1* variants Due to the frequency of somatic *RUNX1* and *DDX41* variants in the RUNX1 and DDX41 carrier-with HM cohorts, respectively. The BAM files for each sequencing sample were manually reviewed using IGV for sequencing coverage of the gene (*RUNX1*) and low-frequency somatic variants (DDX41).