Identification and prioritization of candidate myeloid malignancy germline variants in a large cohort of adult AML patients

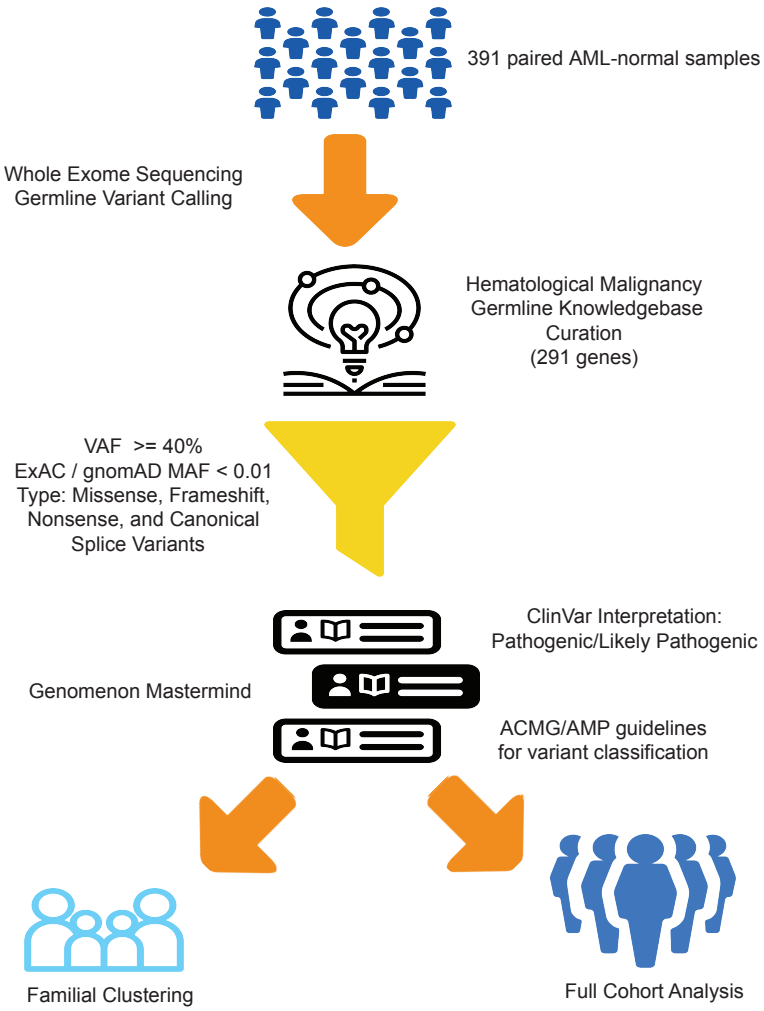
INVENTORY FOR SUPPLEMENTAL INFORMATION

SUPPLEMENTAL FIGURES

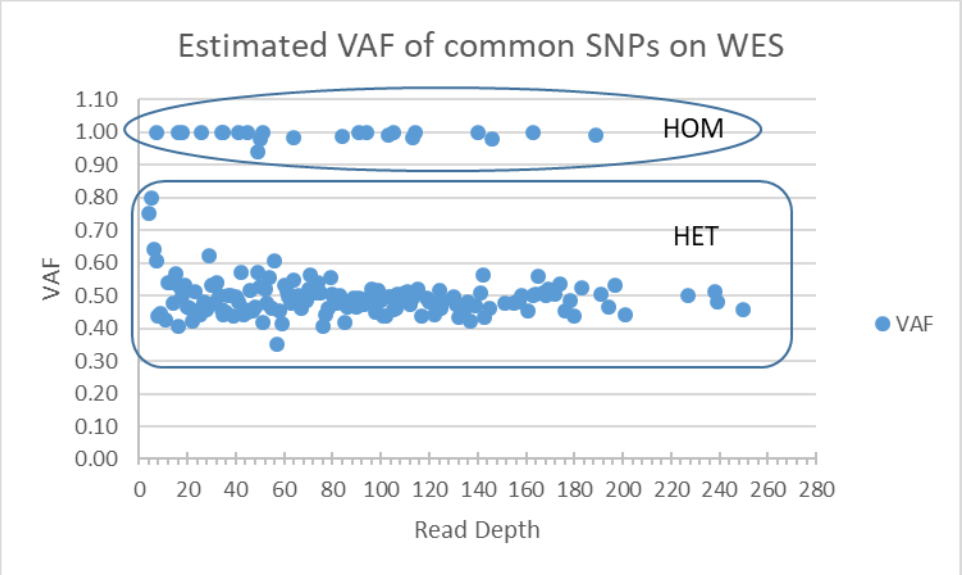
SUPPLEMENTAL TABLES

SUPPLEMENTAL METHODS

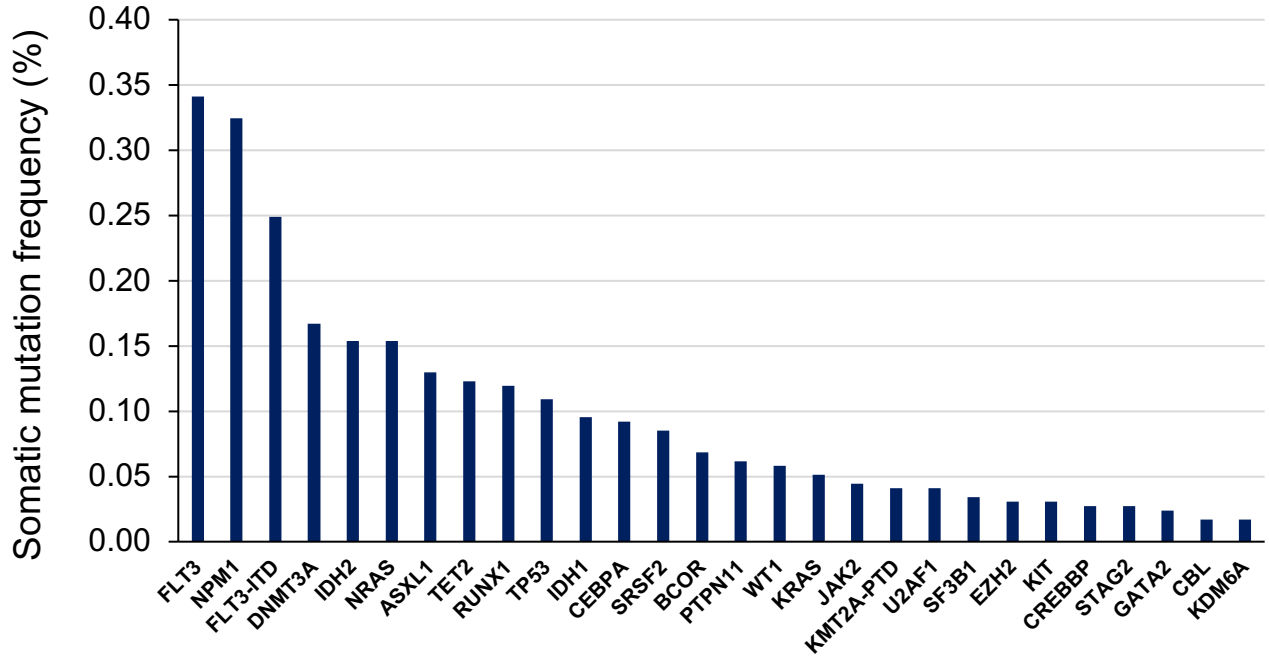
SUPPLEMENTAL FIGURES



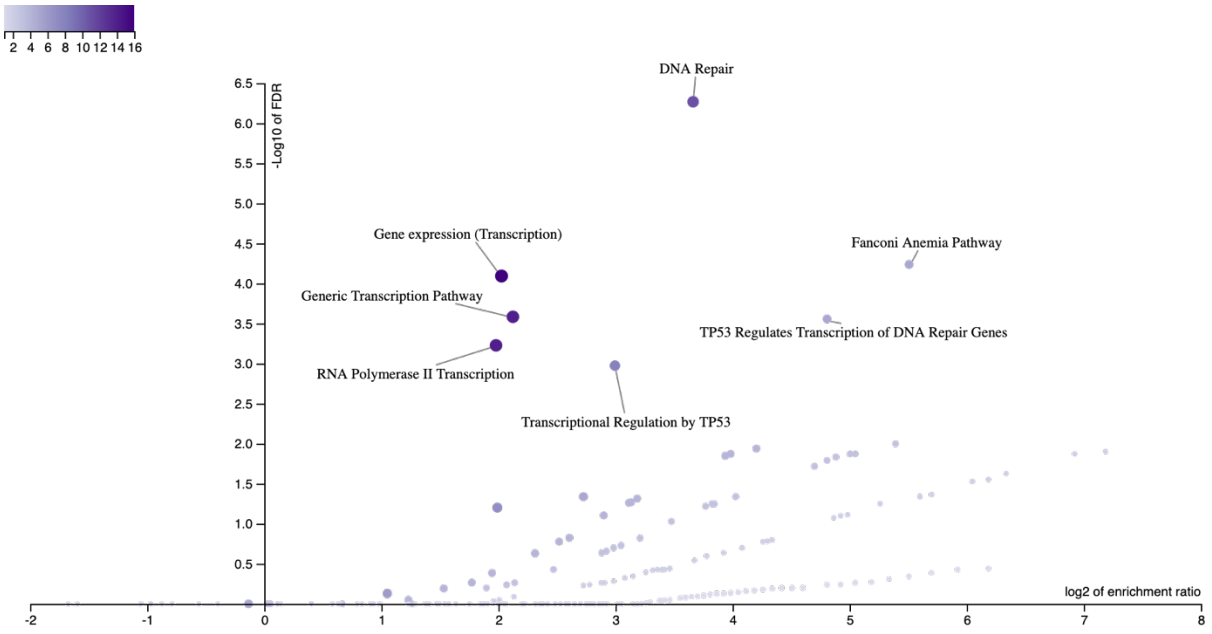
Supplementary Figure 1s. Work flow for curating germline variants in genes associated with hematological malignancies. Flow diagram showing step-wise filters used to prioritize germline variants for interpretive curation using 2015 ACMG/AMP guidelines for 391 AML patients with paired tumor and skin biopsy samples.



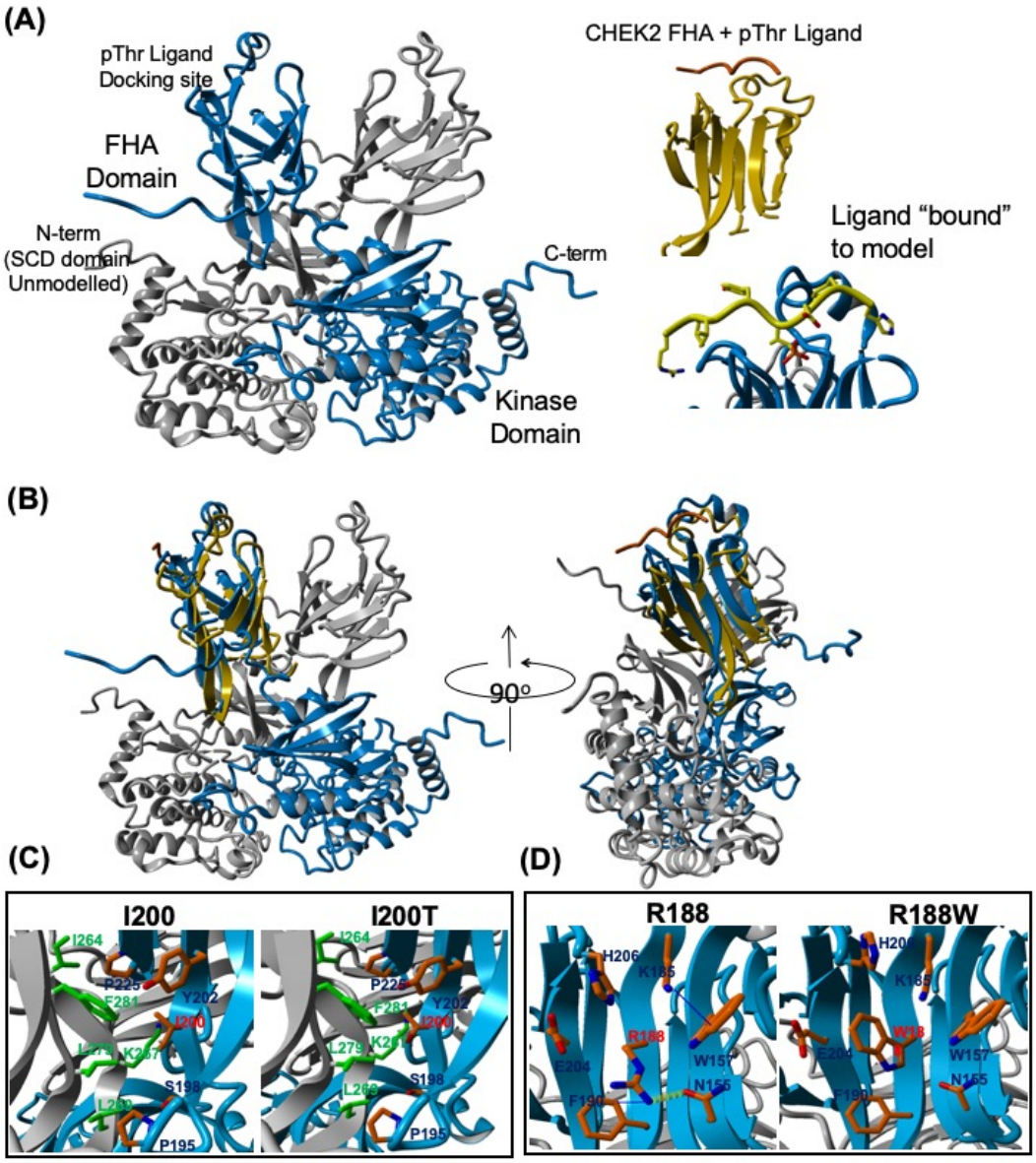
Supplemental Figure 2s. Variant allele frequency of heterozygous or homozygous common SNPs estimated by WES in the function of read depth. The range of VAF is 0.94 – 1.00 for homozygous SNPs, and 0.40 – 0.60 for heterozygous SNPs when the read depth is above 8.



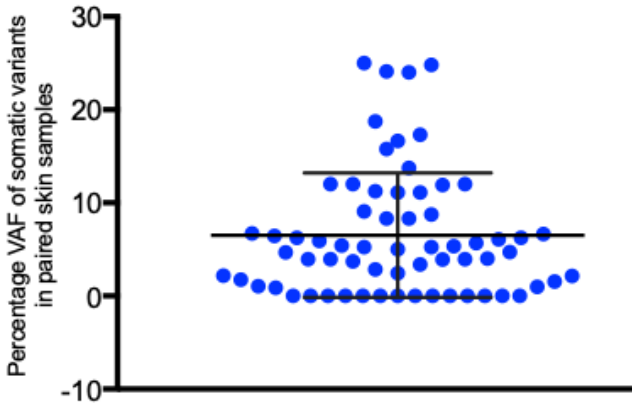
Supplemental Figure 3s (Related to Table 1). Somatic mutation frequencies of genes in 391 AML patients who has a diagnostic molecular profiling. Genes that recurrently mutated by somatic events, including SNV and small indels, in more than 1% of the cohort are displayed in the bar chart.



Supplemental Figure 4s. Pathway enrichment analysis for the genes with pathogenic and likely pathogenic germline variants described in Supplemental Table 4. The most significantly enriched pathway was the Reactome DNA repair pathway (11 genes, FDR adjusted p-value =5.3939e-7). Circle size indicates number of P/LP genes in pathway and color refers to the level of enrichment (enrichment ratio).



Supplemental Figure 5s (Related to Figure 4). 3D protein modeling on CHEK2 protein and the effect of missense variants on the protein structure. (A) CHEK2 dimer model with labelled domains, associated FHA crystal structure for ligand orientation, and final pThr ligand orientation with FHA domain of CHEK2 model. See methods for model construction details. (B) pThr ligand orientation was determined via structural alignment of model with FHA-ligand structure 1GXC (see methods). (C) CHEK2 I200 exists within the dimer interface of the FHA domains of CHEK2. A hydrophobic patch has been proposed to be critical for dimerization and is supported by our model. An I200T variant would disrupt this interaction through the insertion of a polar hydroxyl at the center of the hydrophobic patch (D) The R188W variant in the FHA domain induces a repulsion of the surrounding polar residues, resulting in a loss of three interactions (blue lines and green dash line) and likely destabilizing the structure of the FHA domain. CHEK2 R188 is a charged, solvent exposed polar residue that does not directly engage in inter-molecular interactions. However, our model suggests the R188W variant reduces FHA domain stability. A tryptophan substitution at this position abrogates three important intra-molecular interactions: two intra-beta sheet contacts—a Cation-Pi interaction between K185 and W157, and an H-bond between R188 and N155—and one inter beta-sheet cation-Pi interaction—between R199 and F190—are lost with this alteration.



Descriptive statistics (Quantitative data)	
Statistic	VAF
Number of observations	65
Minimum	0%
Maximum	25.0%
1st Quartile	1.0%
Median	5.0%
3rd Quartile	10.1%
Mean	6.5%
Standard deviation (n)	6.6%

Supplemental Figure 6s: Method to assess infiltration of hematopoietic cells in the skin sample. To assess the degree of infiltration of hematopoietic cells in the skin sample, we analyzed the variant allele frequency (VAF) of co-occuring somatic mutations with high clonal burden in the neoplastic samples ((≥40% VAF) observed in the paired skin samples (Supplementary Table 7s). The scatter plot represents mean ± SD and PRISM software was used to carry out the statistical analysis. The mean VAF of the somatic variants observed in the skin samples is 6.5% ± 6.6%. For 21% (14/65) variants, somatic variants were absent in the skin samples. These data suggested that the infiltration of hematopoietic cells in the skin sample is relatively low, if present.

SUPPLEMENTARY METHODS

Sequencing of tumor and constitutional material using skin biopsy samples

The paired tumor and matched constitutional skin specimens were assessed using whole exome sequencing. In brief particular, ficoll gradient centrifugation was used to isolate mononuclear cells from bone marrow aspirates or peripheral blood or leukapheresis samples. The site of Jamshidi needle insertion for bone marrow biopsies was the location for the skin punch biopsy. Skin biopsies were PBS washed to remove most of the infiltrating hematopoietic cells before DNA extraction and sequencing. Genomic DNA was extracted from the fresh frozen mononuclear cell pellets and the matched skin biopsies with the DNeasy Blood & Tissue Kit (Qiagen, DNeasy Blood & Tissue Kit) for whole exome sequencing.

Cancer predisposition gene selection

A total of 291 genes were chosen for analysis, informed by a review of the professional guidelines[1-3], expert reviews[4-7], and current literature (Supplementary Table 2). These genes were divided into five tiers, with genes in the first 4 tiers associated with cancer predisposition to hematological malignancies (leukemia and lymphomas) or inherited hematological disorders, based on the strength of the supporting evidence: (1) professional guidelines of 2016 WHO, 2017 ELN, and 2019 NCCN; (2) multiple review papers; (3) multiple small cohort studies; (4) one small cohort study or at least one case report. Genes in the fifth tier are associated with somatic alterations in HM or germline predisposition to non-HM cancer.

Germline variant detection by whole exome sequencing (WES)

The paired tumor and skin specimens were assessed using whole exome sequencing as previously described[8] High confidence germline variants were called by VarScan2. The following parameters were utilized in VarScan2: min-var-count = 3, min-var-count-lc= 1, min-strandedness=0, min-var-basequal= 30, min-ref-readpos= 0.20, min-ref-dist3=0.20, min-var-readpos=0.15, min-var-dist3=0.15, max-rl-diff=0.05, max-mapqual-diff=10, min-ref-mapqual=20.

Assess infiltration of hematopoietic cells in the skin sample

To assess the degree of infiltration of hematopoietic cells in the skin sample, we analyzed the variant allele frequency (VAF) of co-occuring somatic mutations with high clonal burden in the neoplastic samples ((≥40% VAF) observed in the paired skin samples (Supplementary Table 7s). PRISM software was used to carry out the statistical analysis.

Variant interpretation and classification

The prioritized germline variants were interpreted according to the 2015 ACMG (American Society of Medical Genetics and Genomics) recommendations[9] by two geneticists/pathologists curators independently. Annotations from the ClinVar database and the Genomenon Mastermind Variants database (via UCSC Genome Browser[10]) had been merged in the prioritized variants dataset to facilitate the interpretation. To assess computational prediction, we developed a workflow that utilized the tool for assessment and prioritization in exome studies (TAPES)[11] which was extended to allow annotation both by VEP (Variant Effect Predictor) and ANNOVAR (ANNOtate VARiation) tools. The categorical ACMG criteria are transformed into a continuous

probability, allowing for a more accurate classification of pathogenic or benign ariants.[12] The predictions were compared to the manually curated set to determine sensitivity and specificity.

Correlation of the deleterious germline variant detection rate and monosomy 7

The patients with monosomy 7 are presented in the supplementary table 5s. The 2x2 contingency table on patients with or without monosomy 7 vs. those with or without at least one deleterious germline variant(s) detected are shown below. The XLSTAT software (2020.5.1.1041) was used to carry out the Fisher's exact test.

	Patients with monosomy 7	**Patients without monosomy 7**
Patients with P/LP germline variant(s)	5	41
Patients without P/LP germline variant(s)	11	264

Fisher's exact test:

p-value (Two-tailed)	0.062
Alpha	0.05

Test interpretation:
H0: The rows and the columns of the table are independent.
Ha: There is a link between the rows and the columns of the table.

Sensitivity and specificity of the automatic prediction tool (TAPES) in variant classification using the 2015 ACMG/AMP guidelines

The 2x2 contingency table on the benign/likely benign (B/LB) calls vs. non-B/LB calls and the pathogenic/likely pathogenic (P/LP) calls vs. non-P/LP calls, respectively, are shown below. The

(A) represents the TAPES automated curation, and the (M) represents the manual curation. The XLSTAT software (2020.5.1.1041) was used to calculate the sensitivity and specificity.

	B/LB(M)	non B/LB(M)
B/LB(A)	653	40
non B/LB(A)	235	588

Statistic	Value	Lower bound (95%)	Upper bound (95%)
Correct classification	81.86%	79.92%	83.80%
Misclassification	18.14%	16.20%	20.08%
Sensitivity	**73.54%**	**70.53%**	**76.33%**
Specificity	**93.63%**	**91.41%**	**95.30%**
False positive rate	6.37%	4.47%	8.27%
False negative rate	26.46%	23.57%	29.36%
Prevalence	58.58%	56.10%	61.05%
PPV (Positive Predictive Value)	94.23%	92.49%	95.96%
NPV (Negative Predictive Value)	71.45%	68.36%	74.53%

	P/LP(M)	non P/LP(M)
P/LP(A)	30	13
non P/LP(A)	19	1454

Statistic	Value	Lower bound (95%)	Upper bound (95%)
Correct classification	97.89%	97.17%	98.61%
Misclassification	2.11%	1.39%	2.83%
Sensitivity	**61.22%**	**47.21%**	**73.55%**
Specificity	**99.11%**	**98.47%**	**99.49%**
False positive rate	0.89%	0.41%	1.37%
False negative rate	38.78%	25.66%	51.89%
Prevalence	3.23%	2.34%	4.12%
PPV (Positive Predictive Value)	69.77%	56.04%	83.49%
NPV (Negative Predictive Value)	98.71%	98.13%	99.29%

Pathway enrichment analysis

To assess the pathway enrichment, overrepresentation enrichment analysis was performed using a hypergeometric test per pathway. All p-values were False Discovery Rate (FDR) adjusted. The Reactome Knowledgebase (www.reactome.org) database was utilized for pathway annotation.

Three-dimensional structural modeling of CHEK2 protein

A 3D structural model of CHEK2 was built using the YASARA Suite to facilitate hypothesis generation on the effects of missense variants on protein structure, function, stability, and ligand association. The partial crystal structure of CHEK2 (3I6U) was used as the starting template, yielding a complete dimer complex that includes kinase domains, FHA domains, unstructured loops, and a docked pThr ligand. Models were subsequently evaluated using the What-If, ProCheck, Verify3D, and ProSA packages. To determine ligand orientation, the X-ray structure of an isolated FHA-ligand complex (1GXC) was aligned to the CHEK2 model, the FHA domain of 1GXC was then deleted, leaving only the ligand and model which were subsequently minimized. This approach allowed us to observe biologically relevant ligand docking orientations and amino acid interactions.

References

1. Swerdlow SH, International Agency for Research on C, World Health O. WHO classification of tumours of haematopoietic and lymphoid tissues. Lyon: International Agency for Research on Cancer; 2017.
2. Döhner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*. 2017;129(4):424-447.
3. Network NCC. Myelodysplastic Syndromes (Version 2.2019).
4. Martin AR, Williams E, Foulger RE, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat Genet*. 2019.
5. University of Chicago Hematopoietic Malignancies Cancer Risk T. How I diagnose and manage individuals at risk for inherited myeloid malignancies. *Blood*. 2016;128(14):1800-1813.
6. Zhang J, Walsh MF, Wu G, et al. Germline Mutations in Predisposition Genes in Pediatric Cancer. *N Engl J Med*. 2015;373(24):2336-2346.
7. Guidugli L, Johnson AK, Alkorta-Aranburu G, et al. Clinical utility of gene panel-based testing for hereditary myelodysplastic syndrome/acute leukemia predisposition syndromes. *Leukemia*. 2017;31(5):1226-1229.
8. Tyner JW, Tognon CE, Bottomly D, et al. Functional genomic landscape of acute myeloid leukaemia. *Nature*. 2018;562(7728):526-531.
9. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-424.
10. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res*. 2002;12(6):996-1006.
11. Xavier A, Scott RJ, Talseth-Palmer BA. TAPES: A tool for assessment and prioritisation in exome studies. *PLoS Comput Biol*. 2019;15(10):e1007453.
12. Tavtigian SV, Greenblatt MS, Harrison SM, et al. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet Med*. 2018;20(9):1054-1060.