ARTICLE OPEN

Check for updates

Genomic structural variations contribute to inform prognosis in patients with cytogenetically normal acute myeloid leukemia

Niccolò Bartalucci[1]✉, Francesco Mannelli[1,2], Danilo Tarantino[1], Alessio Enderti[1], Simone Romagnoli[1], Daniele Colazzo[1], Barbara Scappini[2], Giacomo Gianfaldoni[2], Matteo Piccini[2], Leonardo Signori[1], Fiorenza Irushani[1], Silvia Salmoiraghi[3], Tiziana Ottone[4], Alfonso Piciocchi[5], Margherita Vannucchi[6], Maria Chiara Siciliano[6], Roberto Boccacci[6], Silvia Orsi[7,8], Alessia Civini[3], Paola Fazi[5], Raffaella Santi[9], Marco Vignetti[5], Stefania Bortoluzzi[8], Alberto Bosi[10], Adriano Venditti[4], Alessandro Rambaldi[11], Maria Teresa Voso[4], Paola Guglielmelli[1,2,12] and Alessandro M. Vannucchi[1,2,12]✉

Cytogenetic and genomic profiling of acute myeloid leukemia (AML) guides personalized treatment according to ELN2022 recommendations. However, marked outcome variability persists among cytogenetically normal (CN-) patients, representing an unmet clinical need. We used long-read whole-genome sequencing to interrogate the prognostic significance of structural variations (SVs) in a prospective cohort of 162 intensively treated CN-AML patients. After stringent filtering, we identified 5 somatic SVs associated with shorter overall survival (OS) (HR:4.18, $p < 0.001$) and event-free survival (EFS) (HR:3.59, $p < 0.001$) in 13% of the patients. Results were validated in a real-world cohort of 149 CN-AML, using target assays. These high-risk SVs (HRVs) operationally defined a "very high-risk" category in the framework of ELN2022, overall resulting in more accurate OS prediction. HRVs were independent of most frequent mutations, particularly *FLT3*^{ITD} and *NPM1*^{mut}. Among the latter patients, HRVs independently predicted shorter OS (8.2 months versus not-reached; $p < 0.001$), EFS (3.5 versus 25.7 months; $p < 0.001$), and lower complete response rates (66.7% versus 90.1%; $p < 0.005$). Finally, we provided evidence of transcriptional deregulation of SV-related genes in primary samples and engineered cell models. Current findings support the value of SVs for refining risk stratification in CN-AML, by identifying patients at exceedingly dismal outcome who might benefit from personalized approaches.

Blood Cancer Journal (2026)16:37 ; https://doi.org/10.1038/s41408-026-01465-3

INTRODUCTION

Acute myeloid leukemia (AML), the commonest form of leukemia in adults, is characterized by accumulation of leukemic blasts, blocked in their differentiation process to terminally mature cells. AML is very heterogeneous, posing challenges to diagnosis, risk categorization and treatment [1, 2]. Criteria for AML diagnosis were updated in 2022 by the 5th World Health Organization (WHO) Classification [3] and the International Consensus Conference (ICC) [4] that, with minor differences, rely on combination of cytogenetic and molecular criteria to distinguish specific sub-entities within genetically-defined disease categories [5]. Karyotyping and mutation profiling were included also in the 2022-revised European Leukemia Net (ELN) recommendations for AML risk stratification and management [6]. Accordingly, patients are stratified into favorable, intermediate and adverse risk category, which correspond to overall complete response (CR) rates after intensive chemotherapy of 73%, 66% and 45%, respectively, and 5-year overall survival (OS) rates of 55%, 34% and 15%. Patients older than 60 years have worse outcomes [7]. In the context of intensive chemotherapy, risk categorization at diagnosis, prospectively adjusted by prospective assessment of measurable residual disease (MRD) [8], drives therapeutic strategies aimed to prevent relapse [6].

More than hundred recurrent numeric and structural cytogenetic aberrations are described in AML, with some clustering with morphologically distinct entities [9]. Cytogenetic aberrations not only represent diagnostic markers of unique subtypes, but contribute also one of the most powerful independent predictors of CR achievement, relapse and OS [10, 11]. However, approximately 45% of patients do not present chromosome abnormalities and are labelled as citogenetically normal (CN-) AML. Patients with CN-AML are included in ELN2022 intermediate risk category [12]; however, the co-mutational environment, particularly *CEBPA*, *FLT3*, and *NPM1* mutations, overall accounting for 60% of cases [13], and other less frequent mutations [6], may ultimately determine patients' allocation to low or adverse category [14]. In spite of that,

[1]Department of Experimental and Clinical Medicine, Centro di Ricerca e Innovazione Malattie Mieloproliferative (CRIMM), AOU Careggi, University of Florence, Florence, Italy. [2]Hematology Unit, Azienda Ospedaliero-Universitaria Careggi, Florence, Italy. [3]FROM ResearchFoundation, Azienda Socio-Sanitaria Territoriale Papa Giovanni XXIII, Bergamo, Italy. [4]Department of Biomedicine and Prevention, University of Rome Tor Vergata, Rome, Italy. [5]GIMEMA Foundation, Rome, Italy. [6]Department of Medical Biotechnology, Section of Pathology, University of Siena, Siena, Italy. [7]Department of Medicine, University of Padova, Padova, Italy. [8]Department of Surgery, Oncology and Gastroenterology, University of Padova, Padova, Italy. [9]Department of Health Sciences, Pathology Section, University of Florence, Florence, Italy. [10]Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy. [11]Department of Oncology and Hematology, University of Milan and Azienda Socio-Sanitaria Territoriale Papa Giovanni XXIII, Bergamo, Italy. [12]These authors contributed equally: Paola Guglielmelli, Alessandro M. Vannucchi. ✉email: niccolo.bartalucci@unifi.it; a.vannucchi@unifi.it

Received: 16 December 2025 Revised: 15 January 2026 Accepted: 25 February 2026
Published online: 06 March 2026

the course and outcome of CN-AML patients remain highly heterogeneous and hard to predict [15, 16], with reported 5-year survival rates from 24% to 42% [12, 17–19]. Such variability makes treatment decisions after induction therapy arduous, in particular concerning allogeneic stem cell transplantation (ASCT) in first remission, or the intensiveness of post-remission consolidation therapy [20]. Therefore, it is urgent to identify novel prognostic variables that might contribute to improve current risk stratification by identifying subgroups of CN-AML patients with the most unfavorable characteristics, who might benefit from personalized approaches, including ASCT immediately after induction therapy or experimental therapies. To this end, we focused on submicroscopical chromosomal abnormalities, collectively known as structural variations (SVs) [21], and present evidence that specific SVs have the potential to identify CN-AML patients at shortened survival.

MATERIALS/SUBJECTS AND METHODS

Patients and sample collection

The study was conducted in accordance with the Declaration of Helsinki and employed samples from 311 CN-AML patients, diagnosed and risk-stratified according to ICC2022 and ELN2022. All patients provided written informed consent for genomic sequencing, under protocol #14560 (MYNERVA project #21267, 03/2019) approved by local ethic committee.

Whole genome long-read sequencing. For each patient, 0.7-1µg of BM high molecular-weight genomic DNA (gDNA) was used for whole-genome long-read libraries according to Oxford nanopore technology (ONT).

SVs calling. Fastq files were basecalled by Guppy (ONT, v.5.1.13). Pass reads (Q-score >7) were aligned to GRCh38 by Minimap2 (v2.17) [22]. SVs-calling was independently performed by Sniffles (v. 1.0.12) [23], Sniffles2 [24], SVIM [25] and cuteSV [26] (v. 1.0.10). Only SVs confirmed by all 4 callers were retained for downstream analysis. SVs were filtered against a pool of healthy individuals for technical false filtering. Population SVs (allele frequency ≥0.01) from public database were filtered-out as well.

Selection of prognostically informative SVs. SVs associated with patients' OS were selected by elastic net regularization and independent univariate analysis, then included in multivariable models by Cox Proportional Hazards regression, along with mutations known to impact survival by ELN2022 criteria.

Survival analysis and prognostic model comparison. OS and event-free survival (EFS) curves were prepared by the Kaplan–Meier method and analyzed by log-rank test. To assess the discriminative ability of prognostic models we used Harrel's C-index, Akaike Information Criterion (AIC), and time-dependent Receiver Operating Characteristic (ROC) analysis [27], as appropriate.

Characterization of selected SVs. Regions comprising SVs of interest were technically validated by both long-read and Sanger method. Variant-specific, High-Resolution Melting (HRM)- or Capillary Electrophoresis (CE)-based, assays were developed using custom primers. To evaluate gene expression levels of SVs of interest we used qRT-PCR. The TCGA and GTEx datasets were interrogated regarding target gene expression in the AML dataset (LAML)to GEPIA2 [28]. For immunohistochemistry analysis, formalin-fixed, paraffin-embedded (FFPE) diagnostic BM biopsies were probed with specific antibodies.

Cellular models of selected SVs. HEK293 cells (ATCC CRL-1573) were transfected using Lipofectamine 3000 with either plasmid of interest and EGFP-expressing control, or siRNAs. Transfection efficiency was assessed via EGFP/GFP signal by flow cytometry, with >75% transfection efficiency. Target gene expression was assessed by qRT-PCR. Growth rate was assessed with WST-1 assay (Roche). Propidium iodide (PI) staining was used for cell cycle distribution analysis, while apoptosis was measured by Annexin-V positivity.

Detailed description of methods and statistical methods is provided in Supplementary material.

RESULTS

Patients' cohorts

The study included 2 independent cohorts of CN-AML patients, that served as training (TC; n = 162) and validation (VC; n = 149) cohort. The TC was comprised of CN-AML patients prospectively enrolled in two academic, multicenter, interventional trials: GIMEMA AML1310 (n = 67), a risk-adapted, MRD-oriented, phase 2 study, in intensively treated patients (NCT01452646) [29, 30]; NILG02/06 (n = 95), a randomized phase 2 study that compared CR rate between patients receiving sequential high-dose versus conventional intensive chemotherapy (NCT00495287) [31]. The VC was retrospective, comprised of intensively treated patients, consecutively retrieved from database of Hematology Unit in Florence (period 2012-2019), for whom stored, diagnostic sample for sequencing were available. Main clinical characteristics, molecular profile and outcomes of the 2 series were comparable (Table 1**)**. Median age was 53.0 *vs* 56.0 years in the TC *vs* VC cohort, median follow-up (FU) 52.4 *vs* 52.9 months; a *FLT3* ITD mutation was found in 25.9% *vs* 25.5% and *NPM1* mutation in 58.0% *vs* 47.6%. According to ELN2022, 46.9% *vs* 41.6% of the patients were favorable, 36.4% *vs* 35.5% intermediate, and 16.7% *vs* 22.8% adverse risk.

Identification of genomic structural variations by long-read WGS

The overall research outline is shown in Fig. 1a. We leveraged low-pass, long-read, sequencing using ONT to generate a whole genome sequencing dataset of CN-AML blasts from patients of TC (Fig. 1a, step 1). Sequencing fastq files were generated locally by the ONT proprietary GPU-enhanced basecaller Guppy (v.5.1.13) and pass reads (quality score >7) were aligned to the human reference genome GRCh38 by Minimap2 (v2.17) [22]. Sequencing data were normalized according to local coverage [32, 33] and regions with z-score (calculated every 500 bp) >1.5 were flagged [34] and filtered-out by using a R custom script (available upon request). Average coverage depth was 5.74X (95%CI: 3.52-8.14), read length 5.2 kbp, N50 read length 15.8 Kbp (Fig. 1b); median read quality 14.4% and percent identity 96%. A quality >Q10 and >Q15 characterized 89.2% (650654.6 Mbp) and 42.4% (355036.6 Mb) of the reads (Supplementary Table 1). SVs calling was independently performed using 4 algorithms (Sniffles [23], Sniffles2 [24], cuteSV [26] and SVIM [25]); only SVs concurrently called by all of the 4 methods (n = 2030) were retained for further analysis. After filtering for technical false and SVs detected among local healthy subjects (n = 50), we further excluded SVs with AF ≥ 0.01 in public databases (1000 Genomes, gnomAD v1-4SVs), as well as those with recurrence rate of <1.5% among TC patients. Manual curation [35, 36] combined with high-depth ($>3\times10^4\times$) ONT targeted resequencing using custom-designed primers enabled the refinement and selection of a final high-confidence set of 118 SVs. The latter were annotated according to dbSNP, Database of Genomic Variants (DGV, nstd174) and NCBI Curated Common Structural Variants (nstd186). An Ensembl human gene identifier (ENSG) was matched for 84/118 (71.2%) of SVs. SVs affecting intronic regions were 50%, followed by intergenic (24.6%), gene-flanking (16.9%), regulatory (5.9%), and coding (2.5%) regions (Supplementary Table 2); no inter-chromosomal translocations were included (for details of analysis, see

Table 1. Characteristics of study cohorts.

Characteristic	Training Cohort	Validation Cohort	Full Cohort
Patients	162	149	311
Age at diagnosis, median (95%CI)	53.0 (28.1–68.6)	56 (32.0–74.3)	54 (31.3–71.2)
WBCs, median (95%CI)	17495 (1712–150295)	20400 (1215–171500)	19300 (1500–164800)
FLT3 ITD^{mut}, n (%)	42 (25.9)	38 (25.5)	80 (25.7)
NPM1^{mut}, n (%)	94 (58.0)	71 (47.6)	165 (53.0)
CEBPA^{mut}, n (%)	14 (8.6)	20 (13.4)	34 (10.9)
RUNX1^{mut}, n (%)	13 (8.0)	13 (8.6)	26 (8.4)
ASXL1^{mut}, n (%)	8 (8.0)	15 (10.1)	23 (7.4)
TP53^{mut}, n (%)	0 (0)	5 (3.3)	5 (1.6)
ELN2022 Favorable, n (%)	76 (46.9)	62 (41.6)	138 (44.4)
ELN2022 Intermediate, n (%)	59 (36.4)	53 (35.6)	112 (36.0)
ELN2022 Adverse, n (%)	27 (16.7)	34 (22.8)	61 (19.6)
ASCT, n (%)	42 (25.9)	36 (24.2)	78 (25.1)
CR, n (%)	134 (82.7)	115 (77.2)	249 (80.1)
Relapse, n (%)	61 (37.6)	58 (38.9)	168 (54.0)
Refractory, n (%)	28 (17.3)	34 (20.9)	62 (19.9)
Overall Survival, months, median (95%CI)	47.63 (nr-nr)	35.03 (21.8–48.2)	38.83 (24.3–53.4)
Overall Survival ASCT censored, months, median (95%CI)	31.6 (11.7–51.5)	36.5 (20.9–52.0)	34.1 (22.5–45.6)
Event-free Survival, months, median (95%CI)	16.1 (7.3–24.9)	13.93 (10.31–17.5)	14.53 (10.23–18.8)
Event-free Survival ASCT censored, months, median (95%CI)	13.6 (9.3–17.9)	13.3 (9.1–17.5)	13.6 (10.6–16.6)
Follow-up, months, median (95%CI)	52.4 (47.4–57.3)	52.9 (46.2–59.6)	52.9 (49.1–56.7)

None of the variables showed statistically significant difference in the comparison among cohorts.

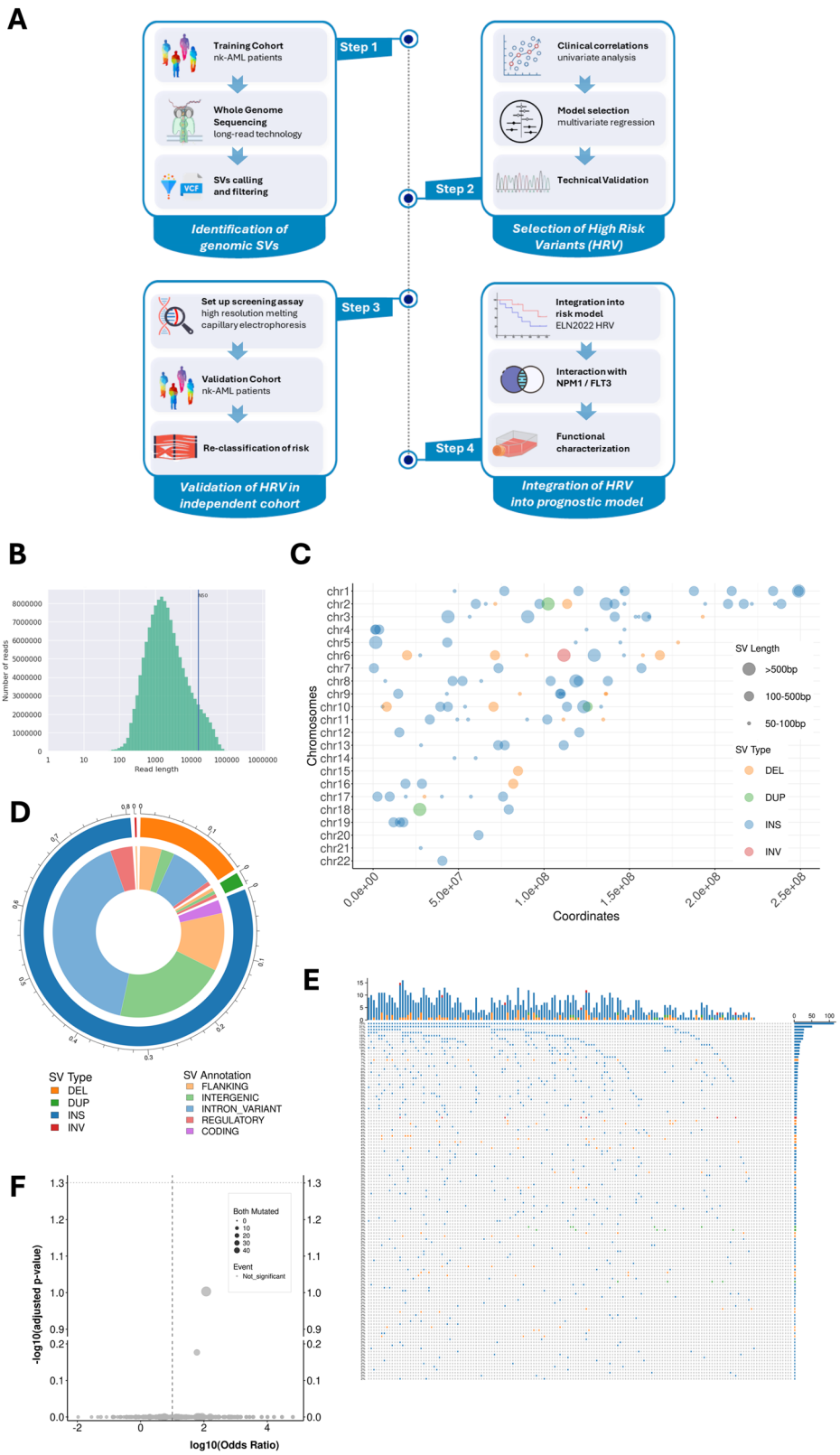
Supplemental Methods). The annotation of genomic regions affected by SVs revealed 24 loci overlapping (reciprocal sequence overlap >50%) with repetitive elements such as short and long interspersed nuclear elements (SINEs and LINEs, n = 10 and n = 4, respectively), long terminal repeats (LTRs, n = 1), microsatellites (n = 4) and tRNA genes (n = 1) (Supplementary Table 3). The 118 SVs of interest affected the 22 autosomes; chr 2 (n = 15 SVs, 12.71% of total), chr 1 (n = 10 SV, 8.47%) and chr 10 (n = 10 SV, 8.47%) were the most frequently involved (Fig. 1c). Most SVs were insertions (n = 95, 80.5%; size: 52-1,297 bp) and deletions (n = 19; 16.1%, size: 53-232 bp), while duplications (n = 3, 2.5%; size: 270-3,103 bp) and inversions (n = 1, 0.9%; size: 792 bp) were utmost rare (Fig. 1d). A total of 148 patients (72.8%) had ≥1 SVs, with mean SVs number per patient of 5.6 (min: 0, max: 16); 9 SVs occurred in ≥10% of patients (Fig. 1e); we found no pattern of SV co-occurrence or mutual exclusivity (Fig. 1f).

Selection of prognostically informative SVs

To identify SVs associated with OS among the 118 candidates (Fig. 1a, step 2), we performed a machine learning–based penalized regression using elastic net regularization (scikit-survival [37] and R glmnet packages). Model hyperparameters were optimized by grid search with repeated 5-fold cross-validation (10 repeats; random state = 0; error score = 0.5) to maximize model performance (maximum iterations: 100,000; 1000 alphas along the regularization path) (Supplementary Fig. 1a, b). This analysis identified 72 SVs with non-zero coefficient (Supplementary Table 4). To strengthen the likelihood of identifying SVs associated with shorter survival, we filtered out SVs with coefficients ≤0 ($n = 35$) and those not supported by permutation testing ($n = 16$) (Supplementary Fig. 1c), retaining only 21 SVs associated with shorter survival (coefficient range 0.02 – 0.59) for downstream analyses. Notably, an independent time-dependent univariate regression identified 6 SVs, all of which overlapped with those selected by the penalized model (Supplementary Table 4). The 21 SVs selected by univariate approach were subsequently included in independent multivariable Cox proportional hazards (PH) models, with backward elimination, which consistently retained 5 SVs (4.24%) for their independent, statistically significant, association with shorter OS, that were considered as "high-risk variations" (HRV). They included: chr2p13.1_del (HR:2.74, 95% CI:1.16-6.64; P = 0.026), chr3q26.1_ins (HR:7.47, 95%CI:2.24-24.91; P = 0.001), chr5p12_ins (HR:3.66, 95%CI:1.02-13.06; P = 0.046), chr9q34.3_del (HR:6.42, 95%CI:2.06-19.99; P = 0.001) and chr18q23_ins (HR:3.61, 95%CI:1.21-10.79; P = 0.022). Conversely, none of the ELN2022 mutation covariates that had a frequency of ≥3% in the cohort, namely *NPM1, FLT3* ITD, *CEBPA, RUNX1, ASXL1* mutation, retained significance in the multivariable model (Fig. 2a, Supplementary Table 5). Convergence across analytical approaches supports the relevance of the identified SVs: all five HRV selected in the adjusted multivariable model were consistently highlighted in univariate analyses, ranking within the top seven elastic net features and being among the six significant time-dependent predictors. To exclude the potential confounding effect of ASCT on the impact of HRV on OS, we assessed the interaction between ASCT and HRV using Cox PH model. As expected, ASCT was significantly associated with improved OS (HR:0.31, 95%CI:0.16-0.62, P < 0.001), however, ASCT-HRV interaction term was not statistically significant (P = 0.17) indicating that the prognostic impact of HRV was independent of transplant status. The PH assumption of the Cox models was verified using both statistical and graphical methods, including permutation-based analysis, Schoenfeld residual tests, and log–minus–log survival plots (Supplementary Fig. 1d–f).

Technical validation of HRV and design of screening assays

Next, genomic coordinates of HRV were determined using custom primers (Supplementary Table 6) generating HRV-spanning amplicons; at least 850 (mean: 37,271) supporting reads, obtained with ONT target sequencing, were analyzed for each HRV.



Resulting coordinates were: chr2p13.1_del:g.71444561_71444650-del; chr3q26.1_ins:g.161438946ins64bp; chr5p12_ins:-g.43393688ins320bp; chr9q34.3_del:g.136418329_136418389del; chr18q23_ins:g.79323263ins310bp (Supplementary Table 7). Further confirmation was obtained in each HRV^{mut} patient by Sanger sequencing (Supplementary Fig. 2a). Then, aiming to develop screening procedures for HRV to facilitate analysis of large patient cohorts, we designed target assays based on HRM (for chr2p13.1_del, chr3q26.1_ins, chr5p12_ins and chr18q23_ins, Supplementary Fig. 2b) and CE (for chr9q34.3_del, Supplementary Fig. 2c). Results from HRM and CE were further validated by Sanger sequencing of ≥3 samples for each HRV type and ≥10

Fig. 1 Study design and whole-genome long-read sequencing data. A Overview of research workflow: step 1, Training Cohort (TC) sequencing and identification of structural variations (SVs); step 2, selection of prognostic variables ("high risk variations", HRV) and their technical validation; step 3, development of target assays to interrogate an independent Validation Cohort (VC); step 4, integration of HRV into ELN2022 risk stratification system. **B** Bar plot showing the quantity (Y-axis) and length (X-axis, log scale) of reads collected through ONT sequencing. Vertical line indicates the N50 value, that is the read length at which 50% of total cumulative read length is contained in reads of equal or greater length. **C** Scatter-plot representing the across-chromosome distribution of 118 filtered SVs ("mature set") identified in TC. The Y-axis denotes autosomes. The X-axis shows genomic coordinates. Each dot corresponds to a single SV; SV type i.e. deletion, duplication, insertion, inversion, is represented by different colors; dot size is proportional to variant length (bp). **D** Donut-chart displaying the proportion of each SV class as percentage of the total mature set (outer ring) and SV functional annotation within each class according to Ensembl VEP (inner ring). **E** Oncoplot showing representation of SVs in TC cases (n = 162). Columns represent patients, rows correspond to SVs. Colors indicate mutation type, the bars summarize number of mutated cases (top) and number of each SV (right). **F** Co-occurrence plot of the 118 SVs across the TC. Axis represent $\log_{10}$(odds ratio) (x axis) and -$\log_{10}$(adjusted p-value) (y axis), respectively, derived from Fisher exact test. Grey points represent SV, none of which reached statistical significance ($\log_{10}$(adjusted p-value) >|1.3|) variant pair.

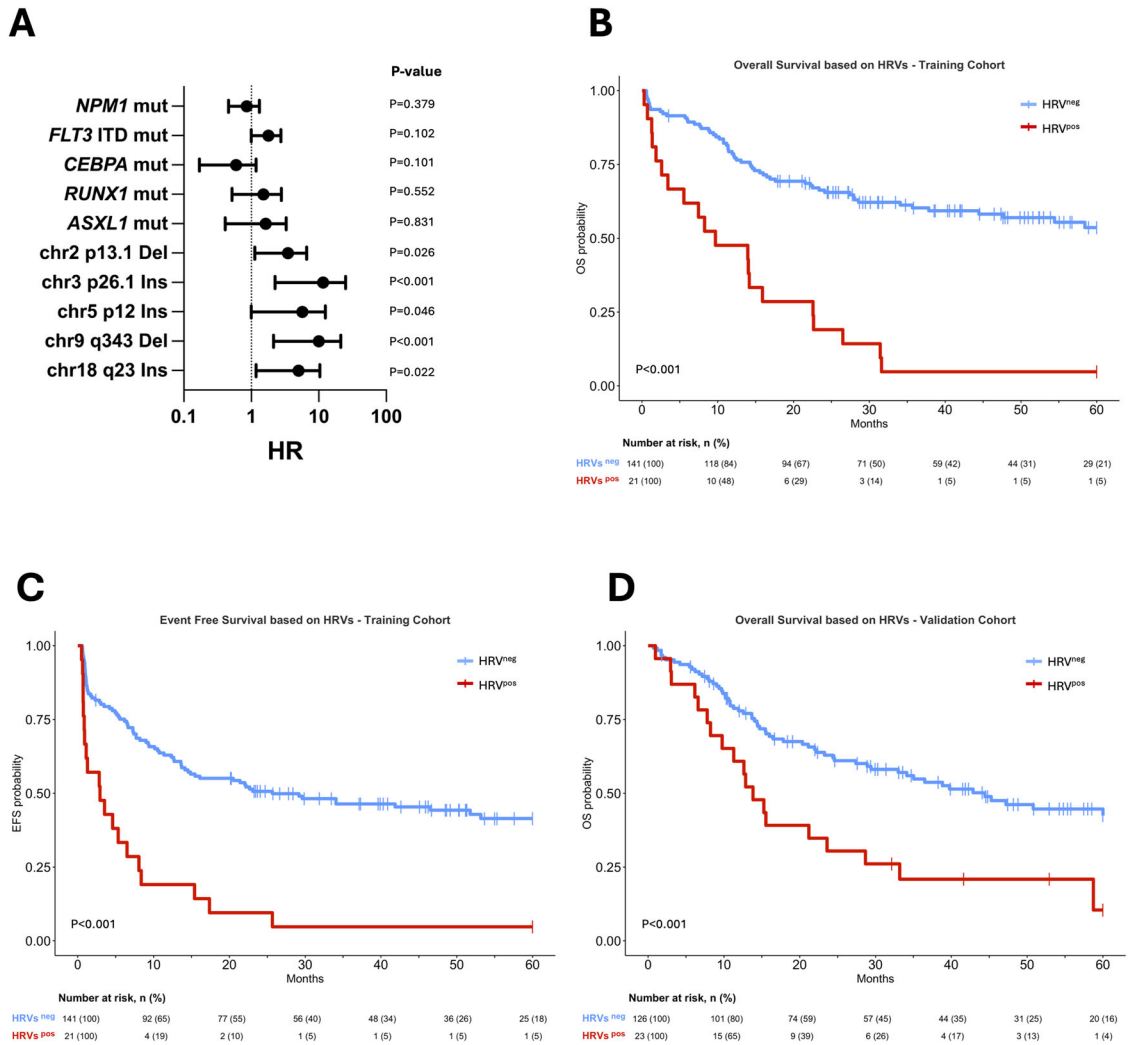


Fig. 2 Identification of prognostically adverse structural variations (HRV) in the Training Cohort. A Forest plot for OS, including prognostic variables from ELN2022 criteria and the five statistically significant SVs (High Risk Variations, HRV) identified in the TC through multivariate Cox regression with backward stepwise elimination. For each covariate, the hazard ratio (HR) and P-value is reported. Kaplan-Meier curves for OS (**B**) and EFS (**C**) of patients of the TC stratified by HRV presence. **D** Kaplan-Meier curves for OS of patients of the Validation Cohort (VC) stratified by HRV presence.

patients without HRV. Finally, we established the somatic nature of each HRV by sequencing DNA obtained from fibroblasts, cultured from BM aspirate of ≥2 patients for each variant, as germline source.

Characterization of patients harboring HRV

Based on the independent association of each HRV with OS, we operationally defined as "HRV^{pos}" the 21 patients in the TC (13%) who harbored ≥1 HRV, and compared them to patients lacking HRV (HRV^{neg}) (Table 2). The two groups were not statistically different for main clinical and molecular characteristics, except for higher leukocyte count in HRV^{pos} patients. Also, the distribution of patients in ELN2022 risk categories was similar. Conversely, significant differences resulted concerning response to treatment and outcome. Compared to HRV^{neg}, the HRV^{pos} category included more refractory patients (38.0% *vs* 14.2%; $P = 0.012$) and

Table 2. Characteristics of HRV^{pos} patients.

Characteristic	Training Cohort (TC) ($n = 162$)			Validation Cohort (VC) ($n = 149$)			Full Cohort ($n = 311$)		
	HRV^{pos}	HRV^{neg}	*P*-value	HRV^{pos}	HRV^{neg}	P-value	HRV^{pos}	HRV^{neg}	*P*-value
Patients, n (%)	21 (13.0)	141 (87.0)		23 (15.4)	126 (84.6)		44 (14.1)	267 (85.8)	
Age at diagnosis, months, median (95% CI)	48.3 (24.4–72.2)	53.08 (28.0–68.0)	0.891	58.0 (46.1–59.7)	56.0 (52.7–57.3)	0.548	52.9 (43.13–58)	54.0 (52.3–55.6)	0.597
WBCs, median (95%CI)	33800 (1708–172300)	17210 (1710–123560)	**0.006**	36900 (17824–89127)	19450 (32079–53446)	0.557	35350 (13500–49800)	19000 (13640–24860)	**0.023**
FLT3 ITD^{mut}, n (%)	9 (42.9)	33 (20.4)	0.06	9 (39.1)	29 (23)	0.121	18 (40.9)	62 (23.2)	**0.016**
NPM1^{mut}, n (%)	12 (57.1)	82 (58.2)	0.555	12 (52.2)	59 (46.8)	0.657	24 (54.5)	141 (52.8)	0.872
CEBPA^{mut}, n (%)	2 (9.5)	12 (8.5)	0.568	2 (8.7)	18 (14.3)	0.740	4 (9.1)	30 (11.2)	0.799
RUNX1^{mut}, n (%)	4 (19)	9 (6.4)	0.069	0 (0)	13 (10.3)	0.221	4 (9.1)	22 (8.2)	0.773
ASXL1^{mut}, n (%)	2 (9.5)	6 (4.3)	0.277	2 (8.7)	13 (10.3)	1	4 (9.1)	19 (7.1)	0.548
TP53^{mut}, n (%)	0 (0)	0 (0)	na	2 (8.7)	3 (2.4)	0.17	2 (4.5)	3 (1.1)	0.148
ELN2022 Favorable, n (%)	9 (42.9)	67 (47.5)	0.604	6 (26.1)	56 (44.4)	0.162	15 (34.1)	123 (46.1)	0.295
ELN2022 Intermediate, n (%)	7 (33.3)	52 (36.9)		12 (52.2)	41 (32.5)		19 (43.2)	93 (34.8)	
ELN2022 Adverse, n (%)	5 (23.8)	22 (15.6)		5 (21.7)	29 (23)		10 (22.7)	51 (19.1)	
ASCT, n (%)	3 (14.3)	39 (27.7)	0.204	16 (69.6)	99 (78.6)	0.417	8 (18.2)	70 (26.2)	0.270
CR, n (%)	13 (61.9)	121 (85.8)	**0.012**	5 (21.7)	31 (24.6)	1	29 (65.9)	220 (82.4)	**0.015**
Relapse, n (%)	19 (90.2)	70 (49.6)	**<0.001**	14 (60.9)	65 (51.6)	0.498	33 (75)	135 (50.6)	**0.003**
Refractory, n (%)	8 (38.0)	20 (14.2)	**0.012**	7 (30.4)	27 (21.4)	0.417	15 (34.1)	47 (17.6)	**0.015**
Overall Survival, months, median (95% CI)	9.7 (0 –19.42)	nr (nr–nr)	**<0.001**[a]	13.8 (9.71–17.96)	44.4 (30.21–58.65)	**<0.001**[a]	12.9 (8.24–17.49)	58.4 (nr–nr)	**<0.001**[a]
OS ASCT censored, months, median (95% CI)	9.7 (0 –20.31)	54.4 (30.49–78.37)	**<0.001**[a]	13.8 (11.75–15.91)	44.4 (32.09–56.77)	**0.006**[a]	13.8 (8.55–19.1)	47.2 (31.2–63.19)	**<0.001**[a]
Event-free Survival, months, median (95% CI)	2.9 (0–6.30)	25.7 (2.23–49.1)	**<0.001**[a]	8.1 (4.05–12.08)	14.4 (8.83–19.9)	0.324[a]	5.3 (2.08–8.58)	19.3 (12.61–25.92)	**<0.001**[a]
Event-free Survival ASCT censored, months, median (95%CI)	2.9 (0–7.01)	15.5 (6.74–24.2)	**<0.001**[a]	8.1 (4.14–12)	14.3 (9.62–19)	0.409[a]	6.2 (2.25–10.15)	14.5 (10.05–19.01)	**0.002**[a]
Follow-up, months, median (95%CI)	60 (nr–nr)	51.6 (46.22–56.98)	0.181[a]	52.9 (29.35–76.51)	52.9 (45.47–60.33)	0.428[a]	60 (nr–nr)	52.4 (47.6–57.13)	0.183[a]

P value was calculated as two-sided by Student's t-test or by [a]Log-rank.

consequently less CR (61.9% *vs* 85.8%, $P = 0.012$); relapses were increased as well (90.2% versus 49.6%; $P < 0.001$). HRV^{pos} patients experienced significantly shortened OS to 9.7 months (95%CI:0-19.4) compared to not reached in HRV^{neg} ($P < 0.001$) (Fig. 2b). The difference remained highly significant after censoring at ASCT (9.7 months; 95%CI: 0-20.3, *vs* 54.5 months; 95%CI, 30.5-78.3; $P < 0.001$) (Supplementary Fig. 3a). The HR for OS and OS censored at ASCT of HRV^{pos} patients was 4.12 (95%CI: 2.49-7; $P < 0.001$) and 3.43 (95%CI: 1.97-5.97; $P < 0.001$), respectively. Also, event-free survival (EFS) resulted significantly shorter among HRV^{pos} patients: 2.9 months (95%CI: 0–6.3) *vs* 25.7 months (95%CI: 2.2-49.1; $P < 0.001$) (Fig. 2c). The significance was maintained after censoring at ASCT (2.9 months; 95%CI: 0-7.01, *vs* 15.5 months; 95%CI: 6.74-24.02; $P < 0.001$) (Supplementary Fig. 3b).

Validation of HRV in an independent cohort

To validate findings about prognostic informativeness of HRV, we used the screening methods described above to interrogate diagnostic BM aspirates of 149 CN-AML patients, constituting a real-world validation cohort (VC) (Fig. 1a**, step 3**). The proportion of HRV^{pos} patients in the VC ($n = 23$, 15.4%) was similar to TC (12.9%) (Table 2). As regarded outcome measures, we confirmed the association of HRV^{pos} with reduced OS: 13.8 months (95%CI: 9.71–17.96) *versus* 44.4 months (95%CI: 30.21–58.65) for HRV^{neg} patients ($P < 0.001$) (Fig. 2d), accounting for a HR of 2.35 (95%CI: 1.4–3.94, $P = 0.003$). A statistically different OS remained after censoring at ASCT: 13.8 months (95%CI, 11.7–15.9) *vs* 44.4 months (95% CI, 32.1–56.7), $P = 0.006$. (Supplementary Fig. 3c). EFS tended to be shorter among HRV^{pos} patients (8.07 months, 95%CI: 4.05–12.08; *vs* 14.37 months, 95%CI: 8.83–19.9) without reaching significance (Supplementary Fig. 3d, e).

Integration of HRV^{pos} into ELN2022 risk stratification system

We then evaluated the impact of adding HRV^{pos} status as novel independent, adverse, category (hence defined, "*very adverse*") to ELN2022 risk stratification system (Fig. 1a, step 4) using the whole cohort of 311 patients, of which 44 (14.14%) were HRV^{pos} (Table 2). Compared to standard ELN2022 (Fig. 3a), the distribution of patients in the novel 4-tiers ELN2022 risk score (ELN^{HRV}) resulted from a shift to the "*very adverse*" risk category of 15 (4.82%), 19 (6.11%), and 10 (3.21%) patients, respectively, from favorable, intermediate and adverse category (Fig. 3b). According to Kaplan-Meier curves, the median OS was 12.9 months (95%CI: 8.24–17.49) for "*very adverse*" risk, 28.6 months (95%CI: 12.87–44.26) for adverse, 34.1 months (95%CI: 4.15–64.11) for intermediate, and not reached for favorable ($P < 0.001$) (Fig. 3c**;** Supplementary Fig.4a after censoring at ASCT). Respective HRs for OS were 4.79 (95% CI:3.1–7.44), 2.31 (95% CI:1.44–3.7) and 1.99 (95% CI:1.31–3) (all, $P < 0.001$), setting the favorable one as reference (Fig. 3d). ROC analysis showed significantly better performance of ELN^{HRV} compared to standard ELN2022 (Fig. 3e) over the follow-up period (time point 8 months: $P = 0.021$; 24 months: $P = 0.038$; 40 months: $P = 0.02$; Supplementary Fig. 4b). Improved accuracy of ELN^{HRV} was supported also by Harrel's C-index (bootstrapped mean C-index:0.67, AIC:1648.15, *vs* C-index:0.63, AIC:1673.22, respectively; Supplementary Fig. 4c). The ELN^{HRV} resulted more accurate also for EFS (Supplementary Fig. 4d, e), as indicated by ROC curves (Supplementary Fig. 4f).

Interaction of HRV with *FLT3* and *NPM1* mutation

We then explored in more detail possible interactions of *FLT3* and *NPM1* mutation with HRV^{pos}. In spite that *FLT3*^{ITD} mutations were enriched among HRV^{pos} patients (40.9% *vs* 23.2%; $P = 0.016$), HRV^{pos} maintained independent, negative prognostic value. In *FLT3* ^{ITD}/HRV^{pos} and *FLT3* ^{ITD}/HRV^{neg} patients, OS was 8.23 months (95%CI:4.28-12.18) and 16.9 (95%CI: 0-42.34; $P < 0.001$), corresponding to HR for death of 4.77 (95%CI: 2.82-8.07) and 1.5 (95% CI: 1.0-2.25) respectively (both $P < 0.001$), using *FLT3*^{wt}/HRV^{neg} patients as reference category.

Among *NPM1*^{mut} patients (n = 165, 53%) (Supplementary Table 8), the OS of *NPM1*^{mut}/HRV^{pos} patients (8.23 months, 95%CI:5.55-10.91) was remarkably shorter than *NPM1*^{mut}/HRV^{neg} (not reached; $P < 0.001$) (Fig. 4a), as it was EFS (3.5 months, 95% CI:0.45-6.61 *vs* 25.7 months, 95%CI:5.88-45.45; $P < 0.001$) (Supplementary Fig. 5a); significant differences were maintained after censoring at ASCT (Supplementary Fig. 5b,c). Also the rate of CR was significantly lower for HRV^{mut} patients (66.7% vs 90.1% for HRV^{neg}; $P = 0.005$) (Table 3). As conceivable, the majority of *NPM1*^{mut} patients were included in ELN2022 favorable risk category (64.8%) compared to 31.5% and 3.6% in intermediate and adverse one (Fig. 4b). However, 23 of 26 (87%) *NPM1*^{mut}/HRV^{pos} patients who were categorized as favorable or intermediate were deceased at 3 years, highlighting the unsatisfactory informativeness of ELN2022 for this patient subgroup; indeed, all 23 patients moved to the "*very adverse*" risk category since they were HRV^{mut} (Fig. 4c). According to ELN^{HRV} 4-tier stratification, median OS of *NPM1*^{mut} patients in "*very adverse*" risk category ($n = 24$) was 8.23 months (95% CI:5.55–10.91) compared to 33 months (95%CI:10.35–55.65) for adverse category ($n = 5$) and not reached for favorable ($n = 96$) and intermediate (n = 40) one ($P < 0.001$) (Fig. 4d,e). The difference remained significant for patients censored at ASCT (Supplementary Fig. 6a). The improved performance of ELN^{HRV} compared to ELN2022 in *NPM1*^{mut} patients was supported by higher values of Harrell's C-index (0.67 vs 0.61) and AIC (707.64 *vs* 727.99) (Supplementary Fig. 6b), as well as results of time-dependent ROC analysis (Fig. 4f). EFS data are reported in Supplementary Fig. 6c–e.

Functional characterization of HRV

As mentioned above, the 5 HRVs were represented by 3 insertions and 2 deletions, 4 occurring in intronic (chr3p34.3_ins at *LINC02067* region; chr5p12_ins at *CCL28* locus; chr9q34.3_del at *PMPCA* locus; chr18q23_ins at *ATP9B* locus) and 1 intergenic (chr2p13.1_del) regions. The latter mapped to a Topologically Associated Domain (TAD), as identified by SVInterpreter [38], potentially affecting the regulation of *ZNF638* and *DYSF*, located at boundaries of deletion [39]. Of note, expression levels of these 2 genes were found deregulated ($|\log_2$ fold change$| > 1$, $P < 0.01$) in the "LAML" dataset, according to TCGA and GTEx databases [28] (Supplementary Fig. 7a). Therefore, we sought to understand whether HRV could affect nearby gene expression, either directly in case of HRV occurring at intronic regions (*LINC02067*, *CCL28, PMPCA, ATP9B*), or indirectly for intergenic deletion flanked by *ZNF638* and *DYSF*. We measured the levels of gene transcripts by qRT-PCR (Supplementary Table 9), using samples of HRV^{pos} patients with ≥ 3 cases for each HRV (except for *LINC02067;* n = 2) and 12 HRV^{neg} patients as reference. We found evidence of different expression of all involved transcripts, with upregulation of *ZNF638* (average 2.33-fold change *vs* HRV^{neg}, P = 0.049), *DYSF* (2.27-fold change, $P = 0.138$) and *ATP9B* (3.26-fold change, $P = 0.018$), and downregulation of *LINC02067* (0.17-fold change, $P = 0.007$), *CCL28* (0.38-fold change, $P = 0.001$) and *PMPCA* (0.41-fold change, $P = 0.008$ Fig. 5a). Expression data were complemented by immunohistochemistry staining of BM sections (Fig. 5b), showing increased staining of ZNF638, DYSF and ATP9B and reduced levels of CCL28 and PMPCA. Finally, we exploited transfection of expression vector (for *ZNF638*) and gene silencing (for *CCL28* and *PMPCA*) in HEK293 cell line (detailed in Supplementary Table 10 and Supplementary Fig. 7b,c). As shown in Fig. 5c, overexpression of *ZNF638* (215.4±25.5-fold change vs No-Target Control (NTC) vector, $P < 0.0001$, Supplementary Fig. 7d) resulted in decreased cell viability (-41.27±4.92% compared to NTC, $P < 0.001$) associated with less cells in G0/G1 (-7.33±0.61%, $P = 0.002$) and cell accumulation in S phase (+5.23±1.79%,

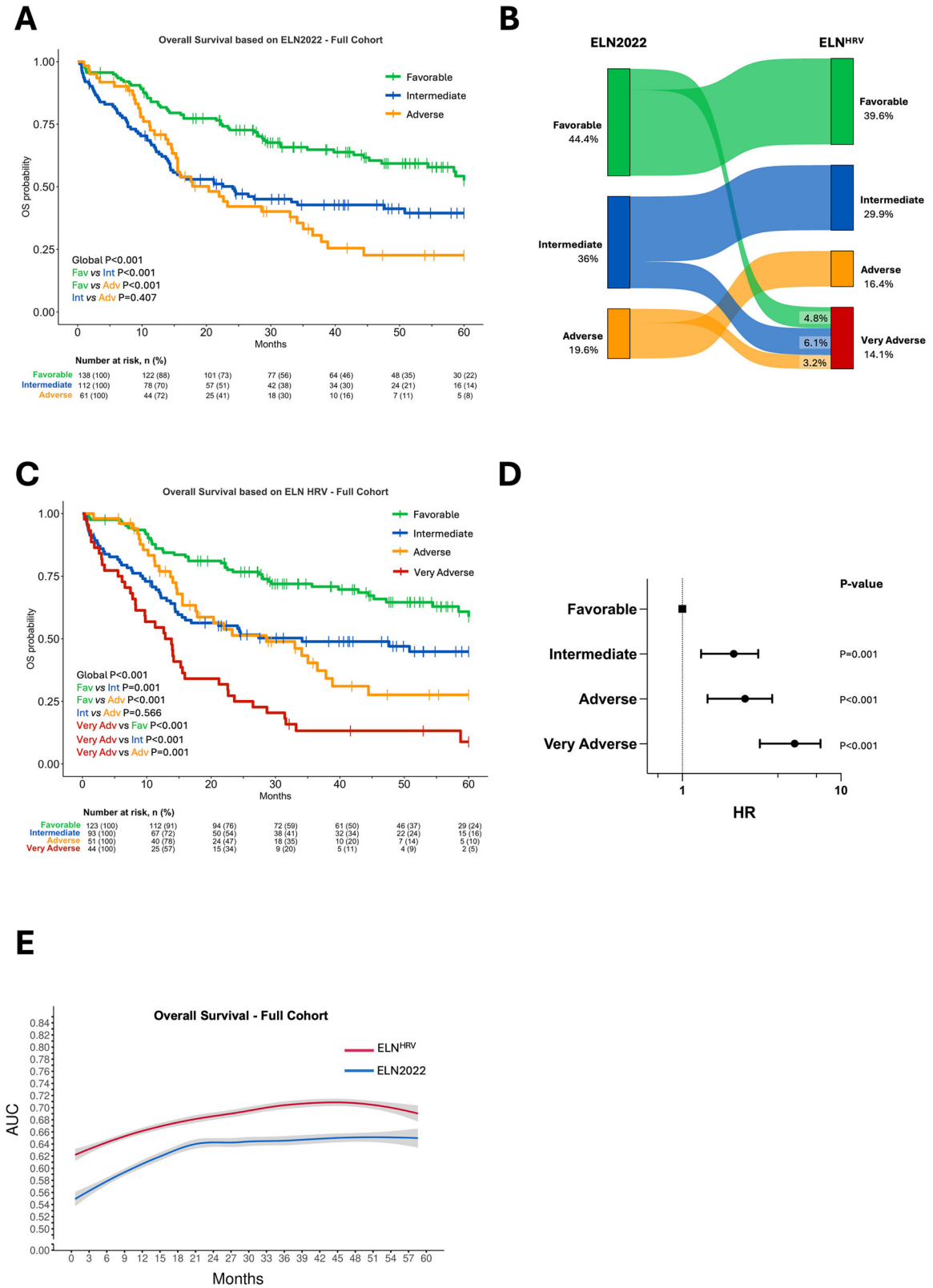
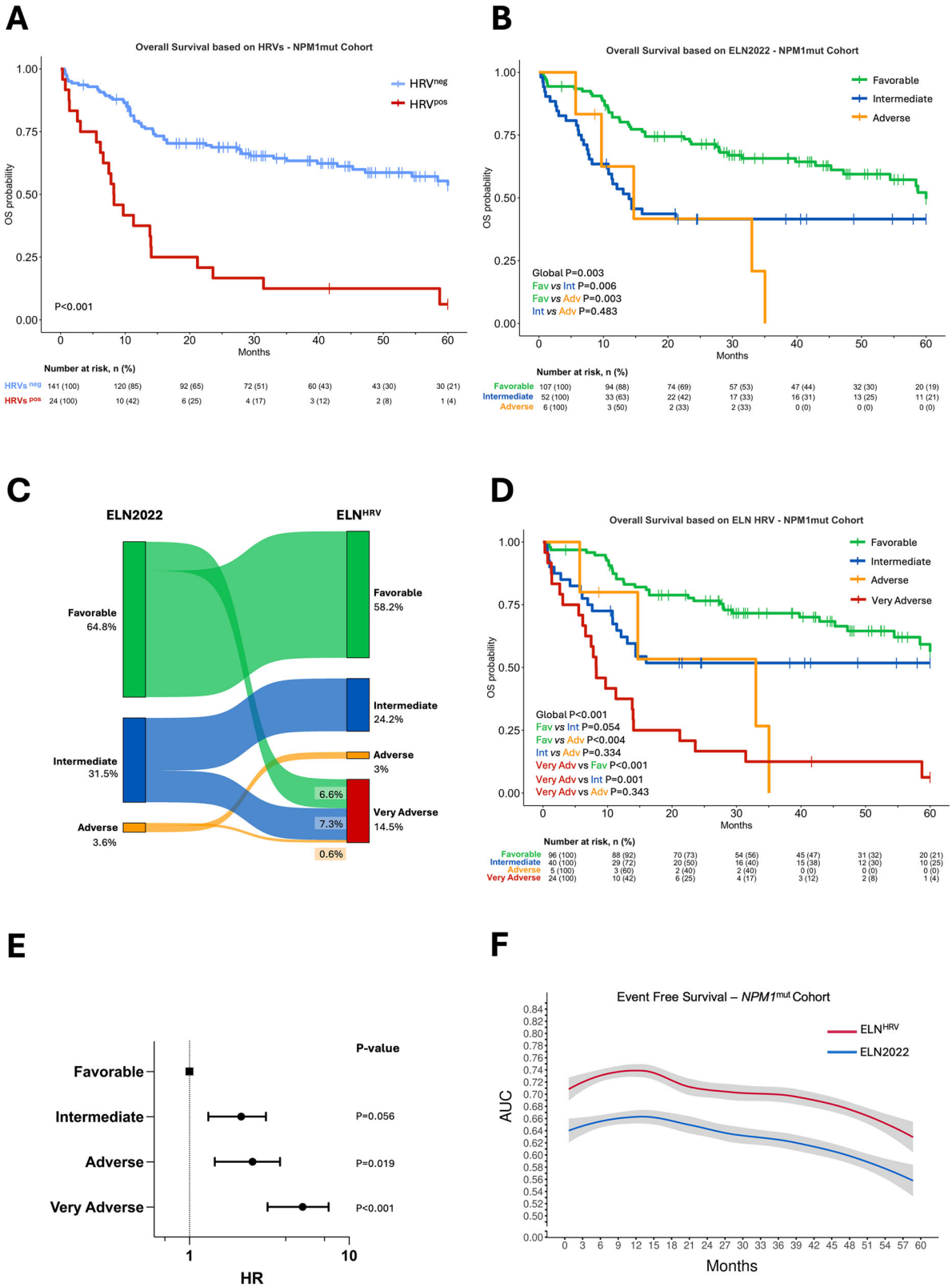


Fig. 3 Comparison of standard ELN2022 and ELN^{HRV} prognostic model in the Full Cohort. A Kaplan-Meier curves for OS of full cohort of patients (n = 311) who were risk-stratified according to ELN2022 criteria including Favorable, Intermediate and Adverse risk categories. **B** Alluvial-plot illustrating the shift of 44 (14.1%) HRV^{mut} patients from favorable ($n = 15$, 4.8%), intermediate ($n = 19$, 6.1%) and adverse ($n = 10$, 3.2%) risk category, to the novel "*very adverse*" risk category based on the presence of HRV. **C** Kaplan-Meier curves for OS of patients stratified by a 4-tiers model (ELN^{HRV}) including the *very adverse* (red) risk category. **D** Forest plot showing HR and P-value calculated in the full cohort for OS, including ELN^{HRV} risk categories (Favorable, Intermediate, Adverse, Very Adverse). **E** Area Under the Curve (AUC) values for OS calculated using time-dependent Receiver Operating Characteristic (ROC) analysis at each time point for ELN^{HRV} (red) and ELN2022 (blue) model. The solid lines represent mean values, shaded areas indicate the 95% confidence intervals.



$P = 0.042$), suggesting an impaired S-phase progression (Fig. 5d), in the absence of cell apoptosis (Supplementary Fig. 7e). Similarly, silencing of *CCL28* (0.36±0.11-fold change vs NTC siRNA, $P = 0.016$) and *PMPCA* (0.19±0.04-fold change vs NTC, $P < 0.0001$) (Supplementary Fig. 7f) led to reduced cell viability (-13.8±3.9% and -29.4±5.6%, respectively vs NTC siRNA, $P = 0.002$ for both) (Fig. 5e) accompanied by decreased proportion of cells in G2/M phase (*CCL28*: -4.93±1.72, $P = 0.02$; *PMPCA*: -6.9±2.71, $P = 0.018$, vs NTC) (Fig. 5f), without increase in apoptosis (Supplementary Fig. 7g).

Fig. 4 Comparison of standard ELN2022 and ELN^{HRV} prognostic model in the *NPM1*^{mut}. **A** Kaplan-Meier curves for OS of *NPM1*^{mut} patients ($n = 165$) stratified based on HRV presence. **B** Kaplan-Meier curves for OS of *NPM1*^{mut} cohort who were risk-stratified according to ELN2022 criteria including Favorable, Intermediate and Adverse risk categories. **C** Alluvial-plot illustrating the shift of 24 (14.5%) HRV^{mut} patients from favorable ($n = 11$, 6.6%), intermediate ($n = 12$, 7.3%) and adverse ($n = 1$, 0.6%) risk category, to the novel "*very adverse*" risk category based on the presence of HRV. **D** Kaplan-Meier curves for OS of *NPM1*^{mut} patients stratified by a 4-tiers model (ELN^{HRV}) including the *very adverse* (red) risk category. **E** Forest plot showing HR and P-value calculated in the *NPM1*^{mut} cohort for OS, including ELN^{HRV} risk categories (Favorable, Intermediate, Adverse, Very Adverse). **F** Area Under the Curve (AUC) values for OS calculated using time-dependent Receiver Operating Characteristic (ROC) analysis at each time point for ELN^{HRV} (red) and ELN2022 (blue) model. The solid lines represent mean values, shaded areas indicate the 95% confidence intervals.

Table 3. Characteristics of *NPM1*^{mut}/HRV^{pos} patients.

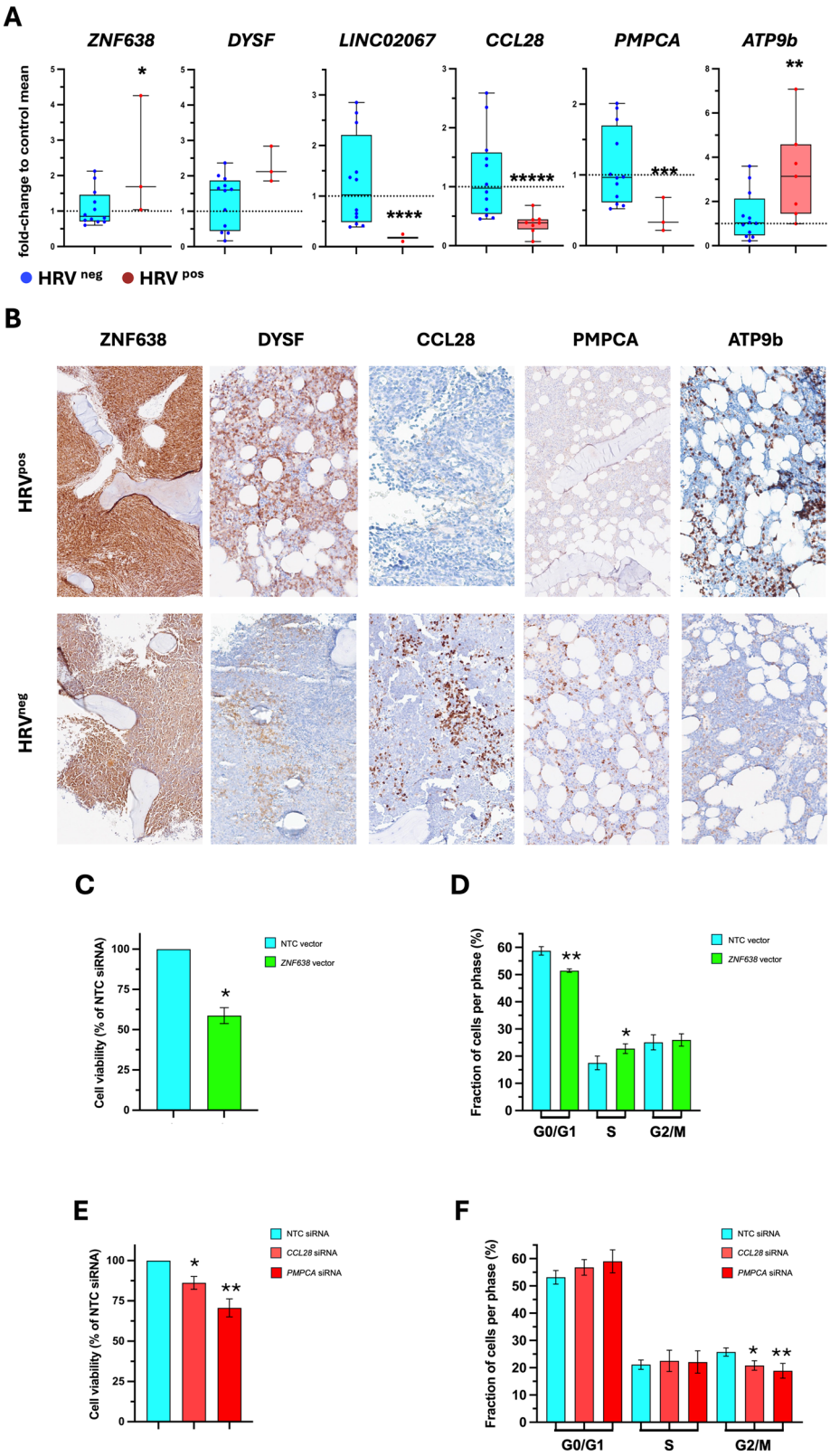
Characteristic	*NPM1*^{mut}/HRV^{pos}	*NPM1*^{mut}/HRV^{neg}	*P*-value
Patients, *n* (%)	24 (14.5)	141 (85.5)	
Age at diagnosis, months, median (95%CI)	47.92 (24.6–68.2)	53.42 (34.7–67.9)	0.193
WBCs, median (95%CI)	56350 (1182–176300)	35200 (1837–257750)	0.139
FLT3 ITD^{mut}, *n* (%)	12 (50)	44 (31.2)	0.101
NPM1^{mut}, *n* (%)	24 (100)	141 (100)	1
CEBPA^{mut}, *n* (%)	0 (0)	3 (2.1)	1
RUNX1^{mut}, *n* (%)	1 (4.2)	1 (0.7)	0.271
ASXL1^{mut}, *n* (%)	0 (0)	3 (2.1)	1
TP53^{mut}, *n* (%)	1 (4.2)	2 (1.4)	0.378
ELN2022 Favorable, *n* (%)	11 (45.8)	96 (68.1)	0.087
ELN2022 Intermediate, *n* (%)	12 (50)	40 (28.4)	
ELN2022 Adverse, *n* (%)	1 (4.2)	5 (3.5)	
ASCT, *n* (%)	3 (12.5)	34 (24.1)	0.292
CR, *n* (%)	16 (66.7)	127 (90.1)	**0.005**
Relapse, *n* (%)	18 (75)	69 (48.9)	**0.026**
Refractory, *n* (%)	8 (33.3)	14 (9.9)	**0.005**
Overall Survival, months, median (95%CI)	8.23 (95%CI: 5.55–10.91)	nr (95%CI: nr–nr)	**<0.001**[a]
OS ASCT censored, months, median (95%CI)	9.7 (95%CI: 2.01–17.39))	58.43 (95%CI: 46.06–70.8)	**<0.001**[a]
Event-free Survival, months, median (95%CI)	3.53 (95%CI: 0.45–6.61)	25.67 (95%CI: 5.88–45.45)	**<0.001**[a]
Event-free Survival ASCT censored, months, median (95%CI)	4.6 (95%CI: 0.9–8.3)	19.27 (95%CI: 9.16–29.38)	**0.002**[a]
Follow-up, months, median (95%CI)	60 (nr–nr)	50.97 (44.16–57.77)	0.264[a]

P value was calculated as two-sided by Student's t-test or by [a] Log-rank.

DISCUSSION

Present study provides evidence supporting the usefulness of analyzing DNA structural variations in the diagnostic workup of patients with AML. Recurrent chromosomal abnormalities and specific somatic mutations inform modern AML classification schemes, and represent the cornerstone of ELN2022 recommendations for risk stratification and management. Conventional cytogenetics, routinely performed in diagnostic specimens, can detect large (5–10 MB) structural abnormalities and numerical changes of chromosomes in approximately half of AML patients; karyotypic abnormalities with either favorable or adverse significance largely inform risk categories delineated by ELN2022. The remaining half of patients have "citogenetically normal" AML (CN-AML), and as such, they are considered as intermediate risk; yet, they represent the most prognostically heterogenous group of patients with AML. In fact, some patients may move to favorable category in case they harbor *NPM1* mutations and are *FLT3*^{wt}, or conversely to adverse risk category if concurrent mutations in myelodysplasia-related genes (*ASXL1, BCOR, EZH2, RUNX1, SF3B1, SRSF2, STAG2, U2AF1, ZRSR2)* and/or *TP53* are present. As a consequence, ELN2022 criteria are less powerful in separating intermediate from adverse risk CN-AML patients [20], as we documented also in our series (Fig. 3a). From a clinical standpoint, this may result in therapeutic decisions not properly aligned to the intrinsic risk of leukemia, recognizing an unmet clinical need.

With the aim to identify novel genetic characteristics potentially improving risk stratification of CN-AML patients, we focused on SVs, that are genomic rearrangements conventionally defined as >50 bp length; they are lost in standard cytogenetics, and may be identified by whole genome sequencing [40]. SVs contribute to a large extent to population genomic variability and are implicated in genetic diseases [41, 42], while less is known for SVs that are somatically acquired in cancer cells [43]; their potential as diagnostic parameters and/or disease biomarkers is actively investigated [44]. Detection of SVs faces with intrinsic technical difficulties using conventional short-read NGS sequencing [45–47]; although sensitivity and specificity depend on the characteristics and size of SV, platforms based on long-read sequencing may offer advantage and be clinically scalable [48–51]. We used the ONT platform to investigate SVs in a training cohort of CN-AML patients who received intensive chemotherapy as part of two prospective trials; although the chemotherapeutic schemes were not identical, they are considered superimposable in clinical practice. Similarly, in the validation cohort constituted of patients referred to a tertiary center, conventional intensive regimens based on anthracyclines and cytarabine were used; of note, due to the time intervals of the studies, no patient in the TC and <5% in



the VC received FLT3 inhibitor midostaurin as part of induction therapy, therefore no specific interaction with midostaurin therapy could be assesses. Using long-read sequencing approach, and after extensive and stringent filtering, we identified a set of 5 somatically-acquired SVs, which resulted significantly associated with shortened survival, independently of all genomic variables included in ELN2022 classification; their presence configured a category of patients with very adverse outcome, therefore they were defined as high-risk variations (HRV). Patients with HRV were characterized by lower rates and shorter duration of leukemia response to treatment, and mostly important with dismal OS and EFS, including after censoring for ASCT. We showed that

Fig. 5 Functional characterization of high-risk structural variations (HRV) in patients' samples and cell models. A Expression levels of *ZNF638*, *DYSF*, *LINC*, *CCL28*, *PMPCA*, and *ATP9B* in HRV^{pos} (red) and HRV^{neg} patients ($n = 12$; blue), measured by qRT-PCR. Individual patient values are shown as-fold change relative to the average of HRV^{neg} patients (dashed line). The horizontal black line in each box represents the mean; whiskers indicate the minimum and maximum values. Each dot represents the average of at least three technical replicates for a single patient. Expression of *18S* was used for normalization. A Student's t-test was used to assess differences between groups using ΔCt values. **B** Representative images of diagnostic bone marrow biopsy sections from HRV^{pos} and HRV^{neg} patients, probed with anti- *ZNF638*, *DYSF*, *LINC*, *CCL28*, *PMPCA*, and *ATP9B* antibodies (H&E counterstaining). Images were acquired at 20× magnification. **C** Cell viability assay using WST-1 reagent on HEK293 cells transfected with *ZNF638*-expressing vector and control, empty vector (NTC). Mean absorbance value of 3 biological replicates is expressed as percentage of NTC. $^{*}P < 0.001$ (Student's t-test) **D** Cell cycle phase distribution based on propidium iodide (PI) staining of HEK293 cells transfected with *ZNF638*-expressing vector and NTC. Shown is the proportion of cells in G0/G1, S, and G2/M phases. Bars represent the mean values of 3 replicates, expressed as percentage of total cells. $^{*}P = 0.042$ $^{**}P = 0.002$. (Student's t-test) **E** Cell viability assay (WST-1 reagent) on HEK293 cells transfected with *CCL28* or *PMPCA* siRNA and control NTC siRNA. Cell viability data (n = 3 biological replicates) are shown as percentage relative to NTC. $^{*}P = 0.0159$, $^{**}P = 0.0019$ (Student's t-test). **F** Shown is cell cycle phase distribution of *CCL28*- or *PMPCA*- silenced HEK293 compared to NTC siRNA transfected cells. The graph displays the proportions of cells in G0/G1, S, and G2/M phases from 3 independent biological replicates. $^{*}P = 0.02$, $^{**}P = 0.018$ (Student's t-test).

introduction of a novel, "*very adverse*", category including HRV^{pos} patients significantly improved the performance of ELN2022 risk score for patients with CN-AML. This was particularly relevant for patients with *NPM1*^{mut}, whose favorable impact was overwhelmed by HRV. Therefore, we surmise that, for decision making, identification of HRV^{pos} patients might facilitate decisions regarding early ASCT, if feasible, adoption of intensified consolidation regimens and stringent MRD monitoring or, preferably, enrollment in clinical trials. We anticipate that prospective trials are needed to ascertain whether any of the above clinical interventions might improve prognosis of HRV^{pos} patients.

Interpretation of pathogenicity of SVs is in cancer is problematic, since the majority of oncogenic SVs lie outside the coding genome [52], as it was the case of the HRVs we identified. A functional role of SVs that determine copy number changes may be reasonably attributed to differences in mRNA dosage, while for intergenic or intragenic variations that may cause gene fusions or gene disruption, assessment of pathogenicity is more complex. However, it was reported that SVs alter the expression of nearby genes at greater extent than single nucleotide variations or small indels [53]. In other instances, SVs may alter the three-dimensional chromatin architecture and disrupt or reshuffle TADs, thereby tuning gene expression, as recently described in AML [54, 55]. Although not the primary focus of this manuscript, we attempted initial studies to get insights into functional consequences of HRVs. We provided evidence that, for each of the 5 HRVs, the levels of mRNA of either directly involved or flanking genes were changed compared to a control group of HRV^{neg} patients; those changes were consistent with results of immunohistochemistry assessment of diagnostic biopsies. Furthermore, forced expression of 3 involved genes, *ZNF638*, *CCL28* and *PMPCA*, resulted in changes of cell cycle regulation and cell survival. Therefore, while how the genes targeted by HRV, that have a variety of predicted functions (*ZNF638* [56–58], *DYSF* [59, 60], *LINC02067* [61], *CCL28* [62–64], *PMPCA* [65–67], *ATP9B* [68, 69]), ultimately affect the behavior of leukemic cells of CN-AML patients remains to be fully ascertained, these data offer proof-of-evidence of functional implications of HRVs.

Our work has some limitations to acknowledge, particularly the fact that the analyzed variable (SVs) was not collected prospectively in either clinical trials (TC) or real-world dataset (VC). Although changes in risk assessment, ASCT allocation criteria and some differences in chemotherapy protocols, might influence the results, the use of intensive treatments with curative intent across all patients should have minimized their impact, as suggested by balanced distribution of pretreatment risk factors and similar outcomes when comparing VC and TC (Tables 1 and 2).

We deliberately restricted our study to cytogenetically normal AML patients, to avoid inferences of concurrent karyotypic abnormalities on the prognostic role of HRVs. However, HRV may not be restricted to CN-AML and represent a more general phenomenon in AML. Interpretation of a karyotype-agnostic role of HRV warrants analysis of larger series that account for the cytogenetic and mutational heterogeneity of AML patients. Also, it would be important to assess whether HRVs remain informative in patients treated with novel target agents, including FLT3 and menin inhibitors.

DATA AVAILABILITY

The raw data (fastq) of each TC sample sequenced by long-read sequencing platform generated in this study have been deposited in the NCBI Sequence Read Archive (SRA) under accession number PRJNA1338842. Additional data will be made available upon reasonable request. Requests should be addressed to amvannucchi@unifi.it.

REFERENCES

1. Khwaja A. The role of Janus kinases in haemopoiesis and haematological malignancy. Br J Haematol. 2006;134:366–84.
2. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. Genomic classification and prognosis in acute myeloid leukemia. N Engl J Med. 2016;374:2209–21.
3. Khoury JD, Solary E, Abla O, Akkari Y, Alaggio R, Apperley JF, et al. The 5th edition of the World Health Organization classification of haematolymphoid tumours: myeloid and histiocytic/dendritic neoplasms. Leukemia. 2022;36:1703–19.
4. Arber DA, Orazi A, Hasserjian RP, Borowitz MJ, Calvo KR, Kvasnicka H-M, et al. International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. Blood. 2022;140:1200–28.
5. Duncavage EJ, Bagg A, Hasserjian RP, DiNardo CD, Godley LA, Iacobucci I, et al. Genomic profiling for clinical decision making in myeloid neoplasms and acute leukemia. Blood. 2022;140:2228–47.
6. Döhner H, Wei AH, Appelbaum FR, Craddock C, DiNardo CD, Dombret H, et al. Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. Blood. 2022;140:1345–77.
7. Rausch C, Rothenberg-Thurley M, Dufour A, Schneider S, Gittinger H, Sauerland C, et al. Validation and refinement of the 2022 European LeukemiaNet genetic risk stratification of acute myeloid leukemia. Leukemia. 2023;37:1234–44.
8. Heuser M, Freeman SD, Ossenkoppele GJ, Buccisano F, Hourigan CS, Ngai LL, et al. 2021 Update on MRD in acute myeloid leukemia: a consensus document from the European LeukemiaNet MRD Working Party. Blood. 2021;138:2753–67.
9. Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DA, Gralnick HR, et al. Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. Br J Haematol. 1976;33:451–8.
10. Grimwade D, Walker H, Harrison G, Oliver F, Chatters S, Harrison CJ, et al. The predictive value of hierarchical cytogenetic classification in older adults with acute myeloid leukemia (AML): analysis of 1065 patients entered into the United Kingdom Medical Research Council AML11 trial. Blood. 2001;98:1312–20.
11. Byrd JC, Mrózek K, Dodge RK, Carroll AJ, Edwards CG, Arthur DC, et al. Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461). Blood. 2002;100:4325–36.
12. Mrózek K, Döhner H, Bloomfield CD. Influence of new molecular prognostic markers in patients with karyotypically normal acute myeloid leukemia: recent advances. Curr Opin Hematol. 2007;14:106–14.
13. Gu Z, Churchman ML, Roberts KG, Moore I, Zhou X, Nakitandwe J, et al. PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. Nat Genet. 2019;51:296–307.

14. Shanker S, Hasserjian RP, Madanat Y, Weinberg OK, Cantu MD. Impact of myelodysplasia-related mutations on 2022 European LeukemiaNet genetic risk classification in de novo acute myeloid leukemia with normal karyotype. Am J Hematol. 2025;100:180–4.
15. Mawad R, Estey EH. Acute myeloid leukemia with normal cytogenetics. Curr Oncol Rep. 2012;14:359–68.
16. Baldus CD, Mrózek K, Marcucci G, Bloomfield CD. Clinical outcome of de novo acute myeloid leukaemia patients with normal cytogenetics is affected by molecular genetic alterations: a concise review. Br J Haematol. 2007;137:387–400.
17. Marcucci G, Mrózek K, Bloomfield CD. Molecular heterogeneity and prognostic biomarkers in adults with acute myeloid leukemia and normal cytogenetics. Curr Opin Hematol. 2005;12:68–75.
18. Swansbury GJ, Lawler SD, Alimena G, Arthur D, Berger R, Van den Berghe H, et al. Long-term survival in acute myelogenous leukemia: a second follow-up of the Fourth International Workshop on Chromosomes in Leukemia. Cancer Genet Cytogenet. 1994;73:1–7.
19. Arthur DC, Berger R, Golomb HM, Swansbury GJ, Reeves BR, Alimena G, et al. The clinical significance of karyotype in acute myelogenous leukemia. Cancer Genet Cytogenet. 1989;40:203–16.
20. Farag SS, Ruppert AS, Mrózek K, Mayer RJ, Stone RM, Carroll AJ, et al. Outcome of induction and postremission therapy in younger adults with acute myeloid leukemia with normal karyotype: a cancer and leukemia group B study. J Clin Oncol Off J Am Soc Clin Oncol. 2005;23:482–93.
21. Romagnoli S, Bartalucci N, Vannucchi AM. Resolving complex structural variants via nanopore sequencing. Front Genet. 2023;14:1213917.
22. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinforma Oxf Engl. 2018;34:3094–100.
23. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, et al. Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods. 2018;15:461–8.
24. Smolka M, Paulin LF, Grochowski CM, Horner DW, Mahmoud M, Behera S, et al. Detection of mosaic and population-level structural variants with Sniffles2. Nat Biotechnol. 2024;42:1571–80.
25. Heller D, Vingron M. SVIM: structural variant identification using mapped long reads. Bioinformatics. 2019;35:2907–15.
26. Jiang T, Liu Y, Jiang Y, Li J, Gao Y, Cui Z, et al. Long-read-based human genomic structural variation detection with cuteSV. Genome Biol. 2020;21:189.
27. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. Biometrics. 2005;61:92–105.
28. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. Nucleic Acids Res. 2019;47:W556–W560.
29. Venditti A, Piciocchi A, Candoni A, Arena V, Palmieri R, Filì C, et al. Risk-adapted MRD-directed therapy for young adults with acute myeloid leukemia: 6-year update of the GIMEMA AML1310 trial. Blood Adv. 2024;8:4410–3.
30. Venditti A, Piciocchi A, Candoni A, Melillo L, Calafiore V, Cairoli R, et al. GIMEMA AML1310 trial of risk-adapted, MRD-directed therapy for young adults with newly diagnosed acute myeloid leukemia. Blood. 2019;134:935–45.
31. Bassan R, Intermesoli T, Masciulli A, Pavoni C, Boschini C, Gianfaldoni G, et al. Randomized trial comparing standard vs sequential high-dose chemotherapy for inducing early CR in adult AML. Blood Adv. 2019;3:1103–17.
32. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinforma Oxf Engl. 2009;25:2078–9.
33. Pedersen BS, Quinlan AR. Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics. 2018;34:867–8.
34. Desvillechabrol D, Bouchier C, Kennedy S, Cokelaer T. Sequana coverage: detection and characterization of genomic variations using running median and mixture models. GigaScience. 2018;7:giy110.
35. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29:24–6.
36. Nattestad M, Aboukhalil R, Chin C-S, Schatz MC. Ribbon: intuitive visualization for complex genomic variation. Bioinforma Oxf Engl. 2021;37:413–5.
37. Pölsterl S. scikit-survival: A Library for Time-to-Event Analysis Built on Top of scikit-learn. J Mach Learn Res. 2020;21:1–6.
38. Fino J, Marques B, Dong Z, David D. SVInterpreter: a comprehensive topologically associated domain-based clinical outcome prediction tool for balanced and unbalanced structural variants. Front Genet. 2021;12:757170.
39. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014;159:1665–80.
40. Duncavage EJ, Schroeder MC, O'Laughlin M, Wilson R, MacMillan S, Bohannon A, et al. Genome Sequencing as an Alternative to Cytogenetic Analysis in Myeloid Cancers. N Engl J Med. 2021;384:924–35.
41. Schöpflin R, Melo US, Moeinzadeh H, Heller D, Laupert V, Hertzberg J, et al. Integration of Hi-C with short and long-read genome sequencing reveals the structure of germline rearranged genomes. Nat Commun. 2022;13:6470.
42. Spielmann M, Lupiáñez DG, Mundlos S. Structural variation in the 3D genome. Nat Rev Genet. 2018;19:453–67.
43. Dubois F, Sidiropoulos N, Weischenfeldt J, Beroukhim R. Structural variations in cancer and the 3D genome. Nat Rev Cancer. 2022;22:533–46.
44. Narang S, Ghebrechristos Y, Evensen NA, Murrell N, Jasinski S, Ostrow TH, et al. Clonal evolution of the 3D chromatin landscape in patients with relapsed pediatric B-cell acute lymphoblastic leukemia. Nat Commun. 2024;15:7425.
45. Bartalucci N, Tarantino D, Enderti A, Colazzo D, Guglielmelli P, Vannucchi AM. A novel approach for highly sensitive and rapid identification of HMGA2 submicroscopic deletions in myeloproliferative neoplasms. Leukemia. 2025;39:2042–5.
46. Romagnoli S, Bartalucci N, Gesullo F, Balliu M, Bonifacio S, Fernandez AGL, et al. Nanopore sequencing for the screening of myeloid and lymphoid neoplasms with eosinophilia and rearrangement of PDGFRα, PDGFRβ, FGFR1 or PCM1-JAK2. Biomark Res. 2021;9:83.
47. Bartalucci N, Romagnoli S, Vannucchi AM. A blood drop through the pore: nanopore sequencing in hematology. Trends Genet TIG. 2022;38:572–86.
48. Ebert P, Audano PA, Zhu Q, Rodriguez-Martin B, Porubsky D, Bonder MJ, et al. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. Science. 2021;372:eabf7117.
49. Chaisson MJP, Sanders AD, Zhao X, Malhotra A, Porubsky D, Rausch T, et al. Multi-platform discovery of haplotype-resolved structural variation in human genomes. Nat Commun. 2019;10:1784.
50. Schaal W, Ameur A, Olsson-Strömberg U, Hermanson M, Cavelier L, Spjuth O. Migrating to Long-Read Sequencing for Clinical Routine BCR-ABL1 TKI Resistance Mutation Screening. Cancer Inform. 2022;21:11769351221110872.
51. Sala-Torra O, Reddy S, Hung L-H, Beppu L, Wu D, Radich J, et al. Rapid detection of myeloid neoplasm fusions using single-molecule long-read sequencing. PLOS Glob Public Health. 2023;3:e0002267.
52. Hnisz D, Weintraub AS, Day DS, Valton A-L, Bak RO, Li CH, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science. 2016;351:1454–8.
53. Wang X, Yue F. Hijacked enhancer-promoter and silencer-promoter loops in cancer. Curr Opin Genet Dev. 2024;86:102199.
54. Xu J, Song F, Lyu H, Kobayashi M, Zhang B, Zhao Z, et al. Subtype-specific 3D genome alteration in acute myeloid leukaemia. Nature. 2022;611:387–98.
55. Klever M-K, Sträng E, Hetzel S, Jungnitsch J, Dolnik A, Schöpflin R, et al. AML with complex karyotype: extreme genomic complexity revealed by combined long-read sequencing and Hi-C technology. Blood Adv. 2023;7:6520–31.
56. Zhu Y, Wang GZ, Cingöz O, Goff SP. NP220 mediates silencing of unintegrated retroviral DNA. Nature. 2018;564:278–82.
57. Du C, Ma X, Meruvu S, Hugendubler L, Mueller E. The adipogenic transcriptional cofactor ZNF638 interacts with splicing regulators and influences alternative splicing. J Lipid Res. 2014;55:1886–96.
58. Meruvu S, Hugendubler L, Mueller E. Regulation of adipocyte differentiation by the zinc finger protein ZNF638. J Biol Chem. 2011;286:26516–23.
59. Krahn M, Béroud C, Labelle V, Nguyen K, Bernard R, Bassez G, et al. Analysis of the DYSF mutational spectrum in a large cohort of patients. Hum Mutat. 2009;30:E345–375.
60. Therrien C, Dodig D, Karpati G, Sinnreich M. Mutation impact on dysferlin inferred from database analysis and computer-based structural predictions. J Neurol Sci. 2006;250:71–8.
61. Liu Y, Hu Z, Zhang Y, Wang C. Long non-coding RNAs in Epstein–Barr virus-related cancer. Cancer Cell Int. 2021;21:278.
62. Wang W, Soto H, Oldham ER, Buchanan ME, Homey B, Catron D, et al. Identification of a novel chemokine (CCL28), which binds CCR10 (GPR2). J Biol Chem. 2000;275:22313–23.
63. Karlsson C, Baudet A, Miharada N, Soneji S, Gupta R, Magnusson M, et al. Identification of the chemokine CCL28 as a growth and survival factor for human hematopoietic stem and progenitor cells. Blood. 2013;121:3838–42. S1-15.
64. Pan J, Kunkel EJ, Gosslar U, Lazarus N, Langdon P, Broadwell K, et al. A novel chemokine ligand for CCR10 and CCR3 expressed by epithelial cells in mucosal tissues. J Immunol Baltim Md 1950. 2000;165:2943–9.
65. Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. Cell. 2016;167:1415–29.e19.
66. Jobling RK, Assoum M, Gakh O, Blaser S, Raiman JA, Mignot C, et al. PMPCA mutations cause abnormal mitochondrial protein processing in patients with non-progressive cerebellar ataxia. Brain. 2015;138:1505–17.
67. Charif M, Chevrollier A, Gueguen N, Kane S, Bris C, Goudenège D, et al. Next-Generation Sequencing Identifies Novel PMPCA Variants in Patients with Late-Onset Dominant Optic Atrophy. Genes. 2022;13:1202.
68. Takatsu H, Baba K, Shima T, Umino H, Kato U, Umeda M, et al. ATP9B, a P4-ATPase (a putative aminophospholipid translocase), localizes to the trans-golgi network in a CDC50 protein-independent manner. J Biol Chem. 2011;286:38159–67.

69. Yagi T, Nakabuchi R, Muranaka Y, Tanaka G, Katoh Y, Nakayama K, et al. Lipid flippases ATP9A and ATP9B form a complex and contribute to the exocytic pathway from the Golgi. Life Sci Alliance. 2025;8:e202403163.

ACKNOWLEDGEMENTS

Research was supported by AIRC 5×1000 call, MYNERVA Project #21267; by GIMEMA under FI 2018 project ID.13; by University of Florence through institutional funds from Zottola legacy. AIRC IG IG #2023 #28966 to SB, and EU MUR PNRR funding within the "National Center for Gene Therapy and Drugs based on RNA Technology" (Project no. CN00000041 CN3 Spoke #6 "RNA chemistry") and within "National Center for HPC, Big Data and Quantum Computing" (Project no. CN00000013 CN1 Spoke #8 "In Silico Medicine and Omics Data") to SB.

AUTHOR CONTRIBUTIONS

N.B., F.M., P.G. and A.M.V. conceptualized the study. N.B. led the study design and analyses. T.O., A.V., M.T.V. established the Training Cohort (GIMEMA AML1310 trial). S.S., A.C. and A.R. established the Training Cohort (NILG02/06 trial). F.M., B.S., G.G., M.P. and A.B. established the Validation Cohort (AOUC). A.P., P.F., M.V. participated in creating the study dataset. N.B., S.R. performed WGS. A.E. and S.R. performed the pre-processing and joint calling of WGS data. D.C., L.S., F.I. performed NGS analysis on Validation Cohort. D.T. and D.C. developed HRV screening assay and performed molecular characterization of HRV. N.B., S.R. and A.E. were responsible for statistical analysis. R.S., M.V., M.C.S. and R.B. performed histology and immuno-histochemistry analysis. S.O. and S.B. aided in analyses and provided key insights. N.B. wrote the manuscript. N.B., A.E. and D.T. produced figures. F.M. and A.M.V. edited the manuscript and supervised the study. All the authors critically revised the manuscript for intellectual content and approved the final version of the manuscript.

COMPETING INTERESTS

A.V., research funding from Jazz Pharmaceuticals, AstraZeneca, Abbvie, Otsuka, Servier, Beigene; consultancy for Amgen, Servier, AstraZeneca, Pfizer, Kite-Gilead, AbbVie, Janssen, Astellas, Astex, Otzuka, Stemline Menarini, Bristol Myers Squibb, Glycostem, Novartis, and Delbert (all are not related to this manuscript). A.M.V., honoraria from Menarini Stemline, GSK, Novartis, Blueprint, AbbVie, Incyte, AOP, Italfarmaco (all are not related to this manuscript); consultancy, Italfarmaco. P.G. honoraria from GSK, Novartis, AOP (all are not related to this manuscript). All other authors declare no conflict of interest.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41408-026-01465-3.

Correspondence and requests for materials should be addressed to Niccolò. Bartalucci or Alessandro M. Vannucchi.