nature communications

Article

https://doi.org/10.1038/s41467-024-54254-6

Longitudinal ultra-sensitive mutation burden sequencing for precise minimal residual disease assessment in AML

Received: 24 February 2024

Accepted: 1 November 2024

Published online: 14 November 2024

Check for updates

Yitian Wu[1,5], Shuai Zhang[2,3,5], Ru Feng[2], Kangming Xiao[4], Ting Wang[2], Jiefei Bai[2], Xiaoyu Zhou[1], Yuji Wang[1], Peng Dai[4] ✉, Hui Liu[2] ✉ & Lucia Ruojia Wu[1] ✉

Relapse is one of the major challenges in clinical treatment of acute myeloid leukemia (AML). Though minimal residual disease (MRD) monitoring plays a crucial role in quantitative assessment of the disease, molecular MRD analysis has been mainly limited to patients diagnosed with gene fusions and *NPM1* mutations. Here, we report a longitudinal ultra-sensitive mutation burden (UMB) monitoring strategy for accurate MRD analysis in AML patients regardless of genetic abnormality types. Using a Quantitative Blocker Displacement Amplification (QBDA) sequencing panel with limit of detection below 0.01% variant allele frequency (VAF), a hazard ratio of 14.8 ($p < 0.001$) is observed in cumulative incidence of relapse analysis of 20 patients with ≥ 2 samples during complete remission (CR). The ROC area under curve (AUC) is 0.98 when predicting relapse within 30 weeks of CR timepoint 2 ($N = 20$). Furthermore, we demonstrate quantitating VAF below 0.01% is essential for accurate relapse prediction.

Acute myeloid leukemia (AML) is a malignant disorder of hematopoietic stem cells characterized by the accumulation of immature myeloid precursors (myeloblasts) in the bone marrow and peripheral blood. While the development of novel chemotherapeutic drugs and targeted therapy have enabled a majority of AML patients to achieve cytomorphologic complete remission (CR), approximately 50% of patients ultimately experience relapse[1]. Potentially curative therapy allogeneic hematopoietic stem cell transplantation (HSCT) during remission clearly reduces relapse, but the high cost, and nonrelapse mortality and morbidity counterbalance the beneficial effect. Therefore, in addition to assessing transplant-related factors when choosing between allogeneic HSCT and nontransplant strategies, precise identification of patients at high risk for relapse is the cornerstone of risk-adapted post-remission therapy selection approaches[2]. Such stratification in AML patients depends on initial disease risk assessment at diagnosis and the detection of minimal residual disease (MRD) during remission. As accumulating evidence suggests cytogenetic and molecular prognostic markers that are identified at diagnosis may be insufficient to estimate the clinical outcome[3–5], sensitive and precise molecular detection of MRD during CR is highly desired[6,7].

MRD detection based on immunophenotyping by multicolor flow cytometry (MFC) is informative with a detection limit of approximately between 0.1% and 0.01%[8,9]. However, 20%–70% of patients experienced relapse with low or negative MFC results, indicating limited clinical sensitivity[10]. Quantitative detection of leukemia-specific genetic aberrations based on RT-PCR, next-generation sequencing (NGS), and droplet digital PCR (ddPCR) to predict disease relapse has been explored, which enabled molecular MRD detection in patients with leukemia fusion genes (such as *RUNX1-RUNX1T1*, *CBFβ-MYH11*, *PML-RARα*) and *NPM1* founder mutations[7,11]. The applicability is limited to about 40% of AML patients diagnosed with targetable abnormalities. The significance of other genetic variants including *KIT*[12], *IDH1/2*[13],

[1]School of Pharmaceutical Sciences, Capital Medical University, Beijing, China. [2]Department of Hematology, Beijing Hospital, National Center of Gerontology; Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing, China. [3]Department of Geriatrics, Beijing Friendship Hospital, Capital Medical University, Beijing, China. [4]College of Chemistry and Molecular Engineering, Peking University, Beijing, China. [5]These authors contributed equally: Yitian Wu, Shuai Zhang. ✉e-mail: pengdai@pku.edu.cn; liuhui8140@126.com; ruojiawu@ccmu.edu.cn

FLT3-ITD[14], *GATA2*[15], and *JAK2*[16], remains to be further demonstrated. Even for patients with mutations or gene fusions, de novo genetic aberrations can occur as an alternative cause of relapse which may cause false negative MRD calling[5,17]. It is challenging and highly desired to develop highly sensitive and specific MRD detection methods to predict clinical outcomes that are applicable to general AML patients regardless of their genetic cause for disease.

Here we report an MRD assessment strategy by longitudinal monitoring of ultra-sensitive mutation burden (UMB) change in genomic regions that are commonly mutated in AML patients. The mutation burden analysis is based on the assumption that cancer cells, regardless of their genetic variation, are more frequently mutated in the genome than normal cells in the same patient. An ultra-sensitive mutation quantitation method Quantitative Blocker Displacement Amplification (QBDA) with a limit of detection (LoD) below 0.01% variant allele frequency (VAF) recently developed by us is employed to detect residual disease at an early stage of relapse[17]. UMB monitoring strategy allowed early prediction of relapse in cumulative incidence of relapse (CIR) analysis. A hazard ratio (HR) of 14.8 ($p < 0.001$) was observed in patients with at least 2 CR sample timepoints ($N = 20$). For all patients with at least 1 CR sample timepoint ($N = 23$), an HR of 11.6 ($p < 0.001$) was achieved. The area under the curve (AUC) of the receiver operating characteristic (ROC) curve was 0.98 when predicting relapse within 30 weeks of CR timepoint 2 ($N = 20$). Furthermore, we demonstrated the UMB strategy is applicable to AML patients with various genetic abnormality types at diagnosis, various treatments, or within different European Leukemia Net (ELN) risk groups.

Results

Design and calibration of UMB quantitation panel

UMB is defined as the sum of VAF of all mutations detected within the selected genomic regions. According to our hypothesis, cancer cells have higher genome instability than normal cells and therefore higher UMB. In bone marrow (BM) aspirate samples, leukemia cell counts drop to lower values after reaching CR than at diagnosis, contributing to a decrease in UMB. For patients with relapse, an increase in leukemia cells occurs before relapse; while for patients with continued CR, the number of leukemia cells remains stable or declines during CR. Therefore, elevated UMB during CR is expected to be a precursor to relapse (Fig. 1a). Because the MRD during CR starts with ultra-low levels of <0.01% VAF[18], we decided to use QBDA technology to quantify various mutations including single base substitutions and indels even below <0.01% VAF. QBDA integrates unique molecular identifier (UMI) with blocker displacement amplification (BDA)[19] for variant enrichment to achieve calibration-free accurate VAF quantitation with low-depth sequencing. In previous work, QBDA was demonstrated in 5 AML cases diagnosed with *NPM1* mutations for MRD detection as a proof-of-concept study[17].

As a measure of residual disease and a potential marker for AML relapse, the central idea of UMB is to evaluate the mutation burden in genomic regions that are commonly mutated in AML with ultra-sensitive quantitation technology. Though a panel covering random genomic regions may be used for measuring the genome instability, considering the potential of hotspot mutations to guide the choice of targeted therapy and considering previous traditional driver mutation-based MRD methods[15–17], the panel was designed to cover AML-related

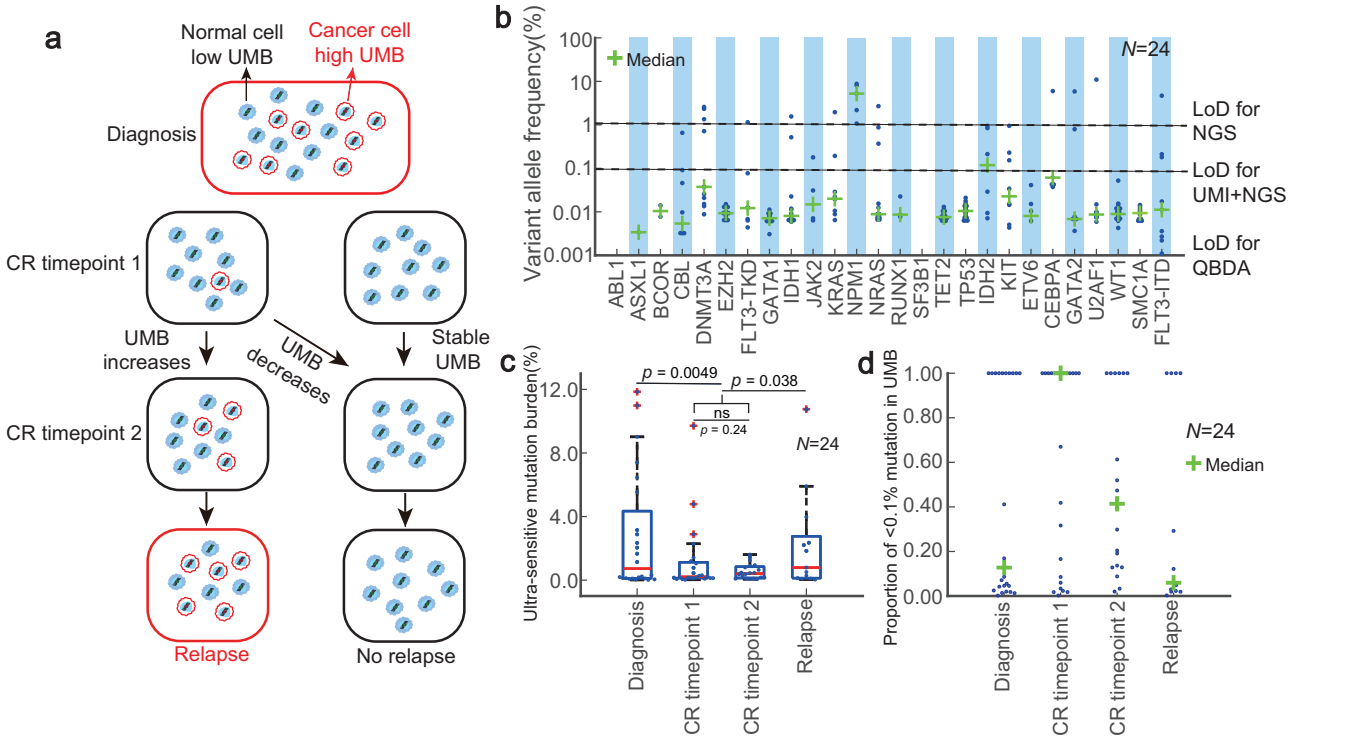

Fig. 1 | Quantitation of low VAF mutations and ultra-sensitive mutation burden (UMB) in AML bone marrow (BM) aspirate samples using a hotspot QBDA NGS panel. a Schematic showing AML disease progression and fluctuation of UMB. Based on the assumption that DNA in cancer cells is more frequently mutated than in normal cells, an increased UMB indicates a higher likelihood of relapse than a decreased or stable UMB does. **b** Summary of mutation VAF detected in AML BM samples at diagnosis. **c** Summary of UMB values at diagnosis, two CR timepoints, and relapse. UMB was lower during CR than at diagnosis or relapse, with $p = 0.0049$ between at diagnosis and during CR, and $p = 0.038$ between during CR and at relapse. During CR, UMB at CR timepoints 1 and 2 were not statistically different (two-tailed t-test). Central lines represent median values, and the bottom and top edges of the box represent the 25th and 75th percentile. The whiskers extend to 1.5 times the interquartile range, and data points beyond the whiskers are marked by + symbols. **d** Proportion of ≤0.1% mutations in UMB. In each sample, the ratio between the sum of VAF for mutations with VAF ≤ 0.1% and the sum of VAF for all mutations (total UMB) was calculated and plotted as dots. Median proportion in diagnosis, CR timepoint 1, CR timepoint 2, and relapse groups were 0.13, 1.00, 0.41, and 0.06. Source data are provided as a Source Data file.

https://doi.org/10.1038/s41467-024-54254-6

Table. 1 | Summary of clinical characteristics of the cases used in this study

Characteristic	Total (*N* = 24)	UMB + (*N* = 9)	UMB−(*N* = 11)	*p*-value
Age at diagnosis (years)[b]	67 (45–79)	67 (49–76)	69 (45–79)	0.528
Gender[a]				0.293
Male	12 (50%)	5 (55.6)	5 (45.5)	
Female	12 (50%)	4 (44.4)	6 (54.5)	
Laboratory data[b]				
WBC (×10^9/L)	4.735 (0.69–260.9)	2.24 (0.69–260.9)	5.95 (1.04–90)	0.524
Hemoglobin (g/L)	92 (51–141)	90 (51–138)	88 (53–141)	0.868
Platelets (×10^9/L)	44 (6–192)	45 (12–119)	50 (6–192)	0.703
BM blasts (%)	52.75 (12.5–91)	33 (12.5–90.5)	71.5 (17–91)	0.115
Peripheral blood blasts (%)	15 (0–94)	10 (0–94)	32 (1–68)	0.281
AML type[a]				
Primary AML	23 (95.8)	8 (88.9)	11 (100)	0.281
MDS history	5 (20.8)	4 (44.4)	0	0.011
FAB subtype[a]				
M1	2 (8.3)	1 (11.1)	1 (9.1)	0.888
M2	15 (62.5)	7 (77.8)	7 (63.6)	0.518
M3	1 (4.2)	0	0	-
M4	3 (12.5)	1 (11.1)	1 (9.1)	0.888
M5	3 (12.5)	0	2 (18.2)	0.196
Molecular markers[a]				
AML1-ETO	1 (4.2)	0	1 (9.1)	0.38
CBFβ-MYH11	1 (4.2)	1 (11.1)	0	0.281
***FLT3*-ITD/TKD**	5 (20.8)	2 (22.2)	3 (27.3)	0.808
***IDH* mutations**	3 (12.5)	0	3 (27.3)	0.098
NPM1	1 (4.2)	1 (11.1)	0	0.281
KIT	3 (12.5)	1 (11.1)	2 (18.2)	0.679
WT1	4 (16.7)	4 (44.4)	0	0.011
2022 ELN risk stratification[a]				
Favorable	4 (16.7)	3 (33.3)	1 (9.1)	0.196
Intermediate	16 (66.6)	4 (44.4)	8 (72.7)	0.220
Adverse	4 (16.7)	2 (22.2)	2 (18.2)	0.833
HSCT[a]	1 (4.3)	0	1 (9.1)	0.38
Relapse (days)[bc]	312 (80–1431)	163 (80–406)	711 (192–883)	0.00029

The UMB+ and UMB− are evaluated after excluding DTA mutations. The percentage may not sum to 100 because of rounding. *p*-values are from a two-tailed t-test.
BM bone marrow, *WBC* white blood cells, *AML* acute myeloid leukemia, *MDS* myelodysplastic syndromes, *ELN* European Leukemia Net, *HSCT* hematopoietic stem cell transplantation.
[a]Number of patients (%).
[b]Median (range).
[c]*p*-value from a two-tailed log-rank test.

mutation hotspots so that potential driver mutations were also captured in UMB calculation. The blocker-based, hypothesis-free mutation enrichment nature of QBDA technology allowed us to detect both residual driver mutation and non-driver mutations caused by genomic instability in the enrichment regions[17].

To efficiently detect mutations associated with AML, we rationally selected commonly mutated hotspot genomic regions in AML based on My Cancer Genome (MCG) and COSMIC databases[20,21], and designed a QBDA panel covering a total of 738 nucleotide sites in 28 hotspots of 22 genes. Detailed information of the panel can be found in Supplementary Data 1. The *FLT3*-ITD mutation cannot be easily detected by QBDA because of its large range, thus we used PCR amplification and standard NGS for the *FLT3*-ITD site (Supplementary Data 2). In order to calibrate the quantitation accuracy of the AML QBDA panel, we mixed gDNA extracted from healthy human peripheral blood mononuclear cells (PBMC) sample with Myeloid DNA Reference Standard from Horizon Discovery, as well as plasmids containing mutant sequences, to formulate a synthetic positive sample (Supplementary Data 3), with a mutation VAF range of 0.0023% to 0.28%. All 28 spike-in mutations in the synthetic positive sample were observed using the AML QBDA panel (Supplementary Fig. 1). Healthy PBMC showed significantly lower UMB than AML patients' CR timepoint 2 samples based on a two-tailed t-test. CR timepoint 1 samples were not significantly different from healthy PBMC, probably due to the short duration after achieving CR (Supplementary Fig. 2). Mutation observed in the healthy PBMC samples were likely low-level clonal hematopoiesis of indeterminate potential (CHIP) mutations. *FLT3*-ITD panel was similarly calibrated using healthy human PBMC mixed with cell line DNA (MV-4-11) containing *FLT3*-ITD mutation (Supplementary Fig. 1).

Mutation VAF and UMB in AML BM aspirate samples

Next, we tested gDNA extracted from BM aspirate samples at diagnosis, relapse (only in relapsed cases), and at least one CR timepoint using the AML QBDA and *FLT3*-ITD panel. A total of 83 samples from 24 AML cases were sequenced. Detailed information can be found in Supplementary Data 4 and 5. A summary of clinical characteristics of the AML cases is shown in Table 1. The patients were not selected based on genomic variation at diagnosis. A total of 221 mutations were detected at diagnosis with VAF ranging between 0.0031% and 11%. The LoD is about 1% VAF for conventional NGS methods such as whole-exome sequencing and other hybrid-capture-based panels and is about 0.1% for standard UMI-based NGS methods[22–24]. Of all the mutations detected by our AML panel, only 9.01% could be detected by conventional NGS, and 14.9% could be detected by NGS with UMIs. In the 24 samples at diagnosis, only 14 contained >0.1% VAF mutations, and 10 of these contained >1% VAF mutations. However, mutations above 0.001% were detected in all samples using QBDA. The results indicate that there are a large number of ultra-low VAF mutations in AML samples, which have not been examined in detail in past studies due to limitations in detection methods; precluding the study of their roles in diagnosis, monitoring, and prognosis (Fig. 1b).

Next, we performed UMB analysis on all diagnosis, CR, and relapse samples. UMB was calculated by summing all mutation VAF above LoD detected by the panel, except for *DNMT3A*, *TET2*, and *ASXL1* (DTA) and *FLT3*-ITD mutations (details can be found in Supplementary Data 4). DTA mutations are related to clonal hematopoiesis (CH) and are likely not markers of true MRD based on literature[6,12]. Moreover, DTA mutation VAF is high during both CR and active disease periods in some cases, accounting for a large proportion of the total mutation burden. Therefore, the quantitation error of DTA VAF may result in an elevated error of the total UMB. In order to evaluate the effect of DTA and other CH mutations potentially covered by the QBDA panel[12], we performed UMB analysis excluding different types of CH mutations (Supplementary Figs. 3 and 4). Excluding only DTA showed the best overall results. *FLT3*-ITD was not covered in the UMB calculation and was analyzed separately.

The mean UMB for all cases (*N* = 24) at diagnosis and relapse were significantly higher than during CR, which is consistent with our hypothesis (Fig. 1c). Next, we calculated the sum of mutation VAF ≤ 0.1% in each sample and analyzed its proportion in UMB. Analyzing by median, CR samples had a higher proportion of low-frequency mutations with VAF ≤ 0.1%. 41% of CR samples had no >0.1% mutations. This further illustrates the existence and potential significance of low VAF mutations for MRD monitoring (Fig. 1d). A color-coded version was added in Supplementary Fig. 5, to distinguish between relapsed and non-relapsed samples.

Relapse prediction by UMB analysis

According to our hypothesis, an increase in UMB during CR is likely a precursor of relapse. We categorized AML cases into two groups based on the UMB at CR sample timepoint 1 (b_{CR1}) and UMB at CR sample timepoint 2 (b_{CR2}): UMB+ represents $b_{CR2} > b_{CR1}$, i.e., UMB increases during CR; and UMB− represents $b_{CR2} \le b_{CR1}$, i.e., UMB decreases or remains stable during CR. Cases with at least 2 CR timepoints ($N = 20$) were used in the analysis. CR timepoint 1 samples were collected within 0–159 days after reaching CR, with a median of 0 days. CR timepoint 2 samples were collected 36–493 days after CR, with a median of 121 days. In the cumulative incidence of relapse (CIR) analysis, there was a significant difference in CR duration between UMB+ and UMB− groups according to the log-rank test (Fig. 2a), with HR = 14.8 (95% CI, 3.9–56.3), and $p = 0.00029$. The median relapse time of the UMB+ group was 548 days shorter than the UMB− group (163 days and 711 days after CR for UMB+ and UMB- groups, respectively). In the 17 relapsed patients, the ratio b_{CR2}/b_{CR1} generally decreased with later relapse, and an R^2 of 0.45 was observed (Supplementary Fig. 6).

In the overall survival (OS) analysis, the HR between UMB+ and UMB− was 2.9 (95% CI, 0.81–10.7), and $p = 0.19$ ($N = 20$). The distinction between the two groups was less pronounced than in CIR analysis, which might be a result of varying severity of illness, treatment regimen, and response after relapse. Therefore, the duration of survival was not directly correlated with the duration of remission (Supplementary Fig. 7).

We proceed to make UMB analysis compatible with patients with only one CR timepoint, by estimating the UMB background using diagnosis samples and compare it with the UMB of a single CR sample for analysis of relapse. For any BM sample, total UMB, normal cell UMB, and tumor cell UMB are b_{total}, b_{normal}, and b_{tumor}, respectively. The percentage of normal and tumor cells are p_{normal} and p_{tumor}, respectively; then we have $p_{normal} + p_{tumor} = 1$, and

$$b_{total} = p_{normal} \times b_{normal} + p_{tumor} \times b_{tumor} \quad (1)$$

We assumed that the UMB background $b_{background} = b_{normal}$ for the same patient, which remains constant across different timepoints. In diagnosis samples with high BM blast percentage (p_{blast}), it can be assumed that $p_{tumor} \approx p_{blast}$. In addition, because healthy individuals have few >1% somatic mutations, mutations greater than 1% can be considered to be mainly brought by tumor cells. Therefore, the sum of the VAF of >1% mutations $b_{>1\%} \approx p_{tumor} \times b_{tumor}$. Therefore, $b_{background}$ can be estimated as:

$$b_{background} = b_{normal} = \frac{b_{total} - p_{tumor} \times b_{tumor}}{1 - p_{tumor}} \approx \frac{b_{total} - b_{>1\%}}{1 - p_{blast}} \quad (2)$$

Based on the ratio between CR sample UMB (b_{CR}) and $b_{background}$, we categorized the samples with 1 CR timepoint into two groups: adjUMB+ and adjUMB−. AdjUMB+ represented $b_{CR}/b_{background} > 1$, and the rest were adjUMB−. For cases with ≥2 CR timepoints, $b_{CR2}/b_{CR1} > 1$ were categorized as adjUMB+, and the rest as adjUMB−. In CIR analysis, there was a significant difference in CR duration between adjUMB+ and adjUMB- groups according to the log-rank test (Fig. 2b): HR = 11.6 (95% CI, 3.4–39.5), and $p = 0.00031$. The median difference in CR duration between the two groups was 532 days (179 days and 711 days for adjUMB+ and adjUMB−, respectively). Because of inaccuracies in the $b_{background}$ estimation, the HR was lower than the analysis using two CR timepoints. Case #10 was lost to follow-up immediately after CR and thus was not included in the analysis. Data for adjusted UMB ratio using only one CR timepoint for relapse prediction can be found in Supplementary Fig. 8. The estimation of $b_{background}$ was mainly focused on filtering out high VAF non-background mutations. As it is challenging to set a judicious cutoff especially considering it might differ from patient to patient, the $b_{background}$ may be inaccurate. When there is no mutation above 1% VAF at diagnosis, the $b_{background}$ is likely overestimated. Our recommended best approach for relapse prediction is by comparing b_{CR1} and b_{CR2}.

It is worth noting that relapse cannot be predicted using only the absolute value of UMB without comparing it to the UMB of other CR samples or the estimated UMB background of the same patient. Due to differences in age, lifestyle, genetic factors, etc., the UMB background varies greatly from person to person and cannot be used directly for relapse monitoring (Supplementary Fig. 9).

Because UMB+ and UMB− groups had significantly different CR durations, next we studied whether the ratio between UMB in 2 CR timepoints is predictive of relapse time. Using b_{CR2}/b_{CR1} to predict whether relapse would occur within 30 weeks of CR timepoint 2, we calculated specificity and sensitivity at different b_{CR2}/b_{CR1} values. The AUC of the ROC curve was 0.98 ($N = 20$). Under the condition of $b_{CR2}/b_{CR1} = 1$, sensitivity was 90% and specificity was 100%. The result indicates that the b_{CR2}/b_{CR1} parameter can accurately predict relapse within a time window of 30 weeks after the second CR timepoint (Fig. 2c). ROC analysis of different intervals after CR timepoint 2 are shown in Supplementary Fig. 10. ROC AUC were all greater than 0.82 between 15 and 45 weeks, and the highest ROC AUC was 0.98 when the interval was 30 weeks. Analysis of null AUC distribution calculated by 1000 simulations of randomly permuted outcome labels is shown in Supplementary Fig. 11.

We also tested the adjusted UMB ratio for the prediction of relapse in $N = 23$ cases with ≥1 CR samples. Adjusted UMB ratio was defined as $b_{CR}/b_{background}$ for cases with 1 CR sample, and b_{CR2}/b_{CR1} for cases with ≥2 CR samples. The ROC AUC of predicting relapse within 30 weeks of CR timepoint 2 (for cases with ≥2 CR samples) or CR timepoint 1 (for cases with 1 CR sample) was 0.90; under the condition of adjusted UMB ratio = 1, sensitivity = 83.3% and specificity = 100%. Because of inaccuracies in $b_{background}$ estimation, the predictive accuracy was lower than using 2 CR timepoints (Fig. 2d).

An M3 case with only 1 CR timepoint was included in Fig. 2b, d analysis. Though M3 is different from other AML subtypes in terms of leukemic cellular characteristics, prognosis, and therapeutic approaches, we were able to successfully predict no relapse within 30 weeks after CR timepoint 1 using adjusted UMB ratio $b_{CR}/b_{background} = 0.50 < 1$.

Significance of quantitating mutations with VAF < 0.01% for UMB-based relapse prediction

To investigate the effect of ultra-low VAF mutations on UMB-based relapse analysis, we next calculated UMB using only >0.01%, >0.1%, or >1% mutations, simulating the experimental results of various detection techniques with different VAF LoD. We performed the CIR and ROC analyses as described above, and compared them with the results using all mutations detected by QBDA (i.e., >0.001% VAF). The results illustrate that $p < 0.05$ between UMB+ and UMB− groups in the CIR analysis was not achieved when only >0.01%, >0.1%, or >1% mutations were used (Fig. 2e). HR was also the highest when using all mutations (Supplementary Fig. 12 and 13). The ROC AUC for predicting relapse increased with improving LoD (Fig. 2f). These results collectively suggest that UMB-based relapse prediction requires <0.01% VAF LoD. The ultra-sensitive QBDA technology enables UMB as a biomarker for monitoring AML disease progression. All the HR, p values, and ROC AUC values were summarized in Supplementary Data 6.

In addition to predicting relapse in the relatively short term, UMB can potentially be used for long-term MRD monitoring with more longitudinal samples. In case #4, CR timepoint 2 with decreased UMB compared to CR timepoint 1 successfully predicted no relapse within 30 weeks of CR timepoint 2. Moreover, an additional sample at CR timepoint 3 was collected. UMB at CR timepoint 3 (day 309 post-CR) increased compared to CR timepoint 2, which was consistent with the following relapse on day 491, demonstrating the potential utility of

https://doi.org/10.1038/s41467-024-54254-6

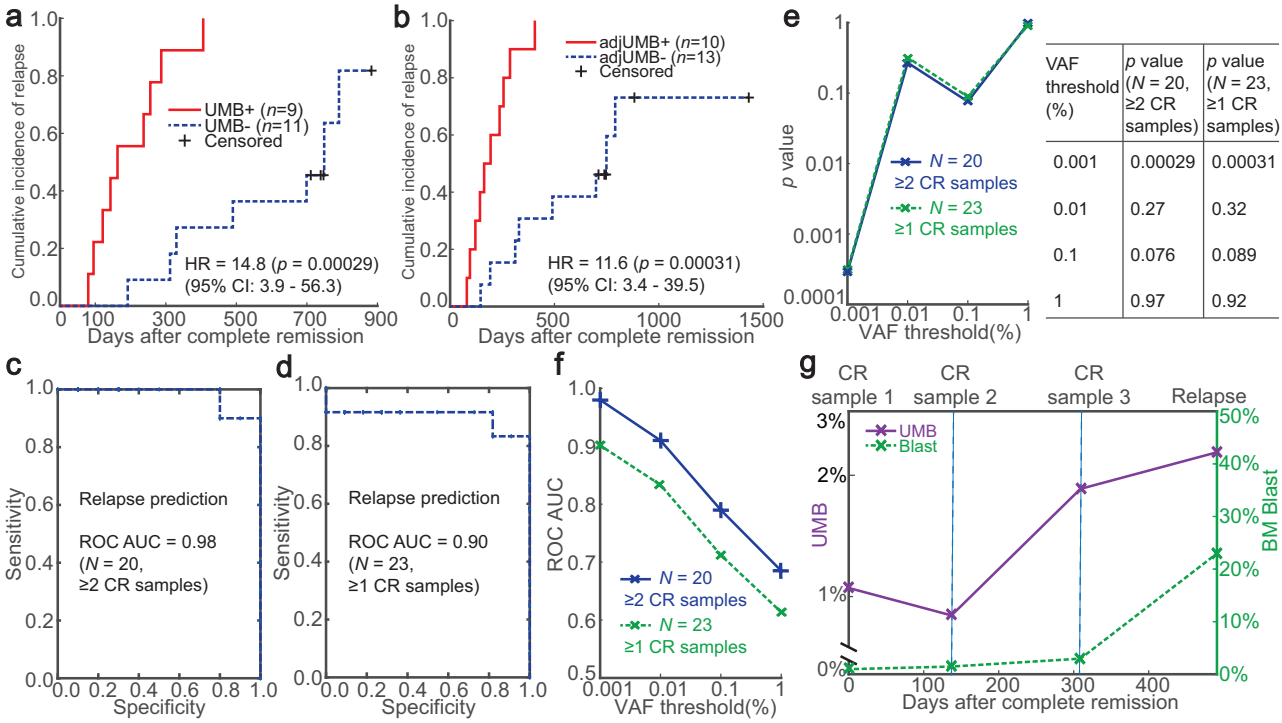

Fig. 2 | Early prediction of relapse in AML patients by UMB. a CIR analysis using 2 samples collected at different times during CR. $N = 20$ cases with at least 2 CR sample timepoints are shown; those with higher UMB at the second sample than at the first sample were classified as UMB+, and the rest as UMB−. Median relapse time was shorter in UMB+ group (163 days) than in UMB− group (711 days) ($p = 0.00029$, two-tailed log-rank test). Censored: Patients not known to have relapsed are censored on the date they were last assessed for response. **b** CIR analysis including all samples with ≥1 CR timepoints ($N = 23$). Cases with adjusted UMB ratio >1 were classified as adjUMB+, and the rest as adjUMB−. Median relapse time was shorter in adjUMB+ group (179 days) than in adjUMB− group (711 days) ($p = 0.00031$, two-tailed log-rank test). **c** ROC of using UMB ratio between CR timepoint 2 and 1 for predicting relapse within 30 weeks of CR timepoint 2 ($N = 20$). **d** ROC analysis using adjusted UMB ratio for predicting relapse within 30 weeks of CR timepoint 2 (for cases with ≥2 CR timepoints) or CR timepoint 1 (for cases with 1 CR timepoint) ($N = 23$). **e** UMB prediction of relapse requires ultra-sensitive VAF LoD <0.01%. UMB was calculated using mutations above different VAF threshold values, and the corresponding p values (two-tailed log-rank test) between UMB+ and UMB− CIR groups in (**a**) and (**b**) are presented. $p < 0.05$ was achieved with LoD <0.01%. **f** ROC AUC of relapse prediction in (**c**) and (**d**) corresponding to different VAF threshold values are plotted. ROC AUC increases with improved VAF LoD. **g** UMB quantitation for continuous relapse monitoring. Example case showed decreased UMB in CR timepoint 2 (day 137 after CR) and increased UMB in CR timepoint 3 (day 309 after CR), signaling relapse at day 491 after CR. Source data are provided as a Source Data file.

continuous UMB monitoring to capture signs of relapse and to guide prompt treatment (Fig. 2g).

Based on the WES results of 8 diagnosis samples, we observed a small number of pathogenic mutations (median, 2; range, 0–3) per patient (Supplementary Data 7), which was consistent with the literature[25]. >99.99% of mutation calls were germline, nonpathogenic SNVs. In the further longitudinal WES analysis of the diagnosis, CR timepoint 1, 2, and the relapse sample of patient #8, both WES and QBDA detected the driver mutations at diagnosis (*CEBPA* and *KIT* mutations). WES did not detect *CEBPA* or *KIT* mutations during CR or at relapse, while QBDA was able to detect and quantify them (Supplementary Data 8), because of the difference in VAF LoD.

MRD detection in cases with different genetic abnormality types at diagnosis

According to the 2022 ELN recommendations, the confirmed biomarkers for MRD detection in AML are *NPM1* mutations and gene fusions including *RUNX1-RUNX1T1*, *CBFB-MYH11*, and *KMT2A-MLLT3*; the significance of other gene mutations such as IDH1 and IDH2 are not clear[13,26]. In this study, *NPM1*-positive cases accounted for 20.8% of the tested cases, which is close to the percentage reported in literature[27]. *RUNX1-RUNX1T1*, *CBFB-MYH11*, and *KMT2A-MLLT3* gene fusion-positive cases together accounted for 8.3% of the tested cases; literature reported 17.5% fusion-positive in all cases (Fig. 3a, b)[28]. The difference between experimental and reported proportions might be a result of the high average age of the cases in this study.

We defined double negative as neither positive for *NPM1* mutations nor for gene fusions. The MRD of the remaining 70% double negative cases cannot be accurately analyzed by conventional methods. Therefore, we specifically investigated whether UMB can assess MRD and predict relapse in these cases. Similar to the section "Relapse prediction by UMB analysis", we performed CIR analysis and prediction of relapse within 30 weeks of CR timepoint 2 (for cases with ≥2 CR samples) or CR timepoint 1 (for cases with 1 CR sample) for only the double negative cases using the adjusted UMB ratio. In CIR analysis, there was a significant difference in CR duration between adjUMB+ and adjUMB− groups according to the log-rank test (Fig. 3c), with HR = 16.7 (95% CI, 4.1–67.6), and $p = 0.00034$. The median difference in CR duration between the two groups was 553 days (194 days and 747 days for adjUMB+ and adjUMB−, respectively). Using adjusted UMB ratio (b_{CR2}/b_{CR1} and $b_{CR}/b_{background}$) to predict whether relapse would occur within 30 weeks showed ROC AUC = 1.00, indicating that changes in UMB can accurately predict relapse for double negative cases (Fig. 3d).

Among the cases with *NPM1* mutations at diagnosis, QBDA reported no *NPM1* mutation during CR in case #16, which was consistent with the UMB− result and a favorable outcome of no relapse at the last follow-up visit (343 days after CR timepoint 2). Case #3 and #14 had positive *NPM1* results but negative UMB during CR. Their outcome of no relapse within 30 weeks of CR timepoint 2 was consistent with the UMB− results. These 2 cases eventually relapsed, therefore the very low VAF *NPM1* mutations (0.0051% and 0.068%, respectively) may predict relapse in the long term (Fig. 3a).

https://doi.org/10.1038/s41467-024-54254-6

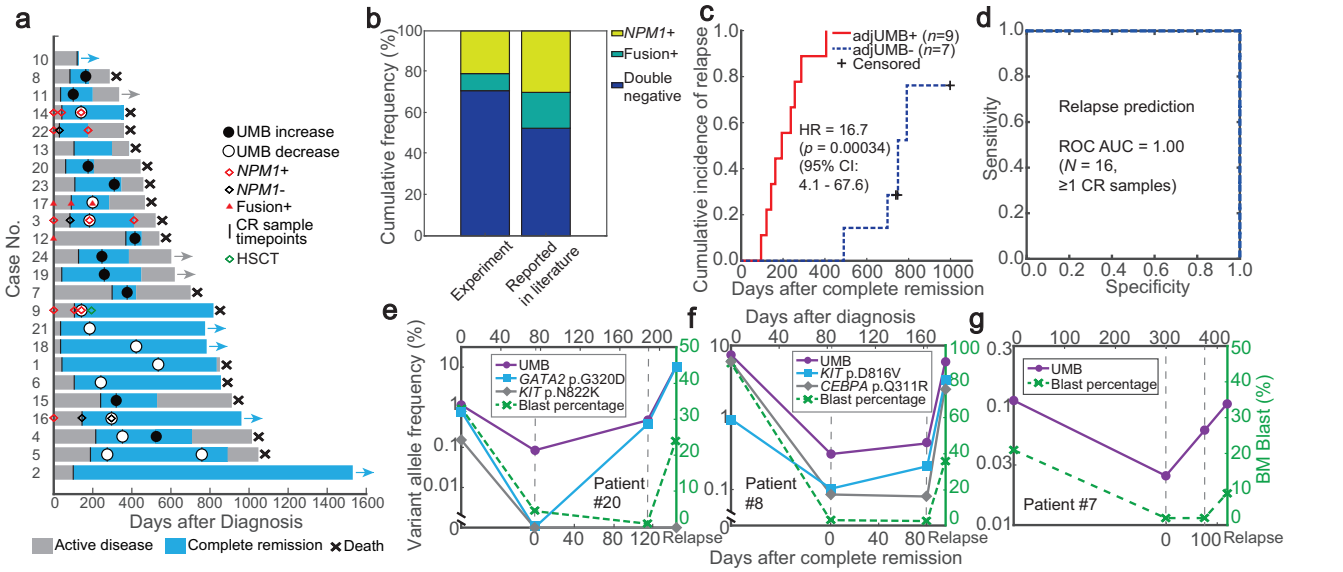

Fig. 3 | Assessment of MRD in AML cases with different mutation characteristics. a Swimmer plot of clinical course and molecular findings in 24 AML patients. **b** Proportion of *NPM1*-positive, gene fusion-positive, and double negative cases, and comparison with reported values[27,28]. *NPM1*-positive represents cases with >0.1% VAF pathogenic *NPM1* mutation observed at diagnosis. Gene-fusion-positive represents *RUNX1-RUNX1T1*, *CBFB-MYH11*, or *KMT2A-MLLT3* detected at diagnosis. **c** CIR analysis of double negative cases in adjUMB+ and adjUMB− groups ($N = 16$) ($p = 0.00034$, two-tailed log-rank test). **d** ROC of predicting relapse within 30 weeks of the latter CR sample timepoint in double negative cases by adjusted UMB ratio ($N = 16$). **e**–**g** Double negative case examples of UMB, mutation VAF, and blast percentage fluctuation during CR and at relapse in double negative cases. Only mutations with >0.1% VAF at diagnosis are shown. Source data are provided as a Source Data file.

In the double negative cases, 50% of them bore >0.1% VAF pathogenic mutations detected at diagnosis. Based on previous research[29], these mutations may be used as a biomarker for MRD (Supplementary Data 4). For example, in case #20, *GATA2* mutation was detected at diagnosis and was below the detection limit at CR timepoint 1. A VAF of 0.38% was observed at CR timepoint 2, and this patient relapsed 143 days after reaching CR. The upward trend of UMB during CR was consistent with the *GATA2* VAF, which also predicts relapse in advance (Fig. 3e). UMB was also successfully applied to cases in which residual diseases cannot be well inferred by a single pathogenic mutation. In case #8, there were *KIT* and *CEBPA* mutations with >0.1% VAF at diagnosis. *KIT* mutation VAF increased and *CEBPA* VAF decreased during CR, making it difficult to predict relapse by a single mutation. On the other hand, UMB clearly increased during CR, which successfully predicted relapse within 30 weeks of CR timepoint 2 (Fig. 3f). In case #7, no >0.1% VAF mutation was detected at diagnosis, but relapse within 30 weeks of CR timepoint 2 could also be predicted by the rise of UMB during CR (Fig. 3g). Collectively, these results demonstrate that UMB provided accurate MRD detection and early warning of relapse for double negative cases, which account for more than 60% of all AML cases lacking methods to accurately analyze molecular MRD. UMB-based MRD analysis may improve the prognosis of more AML patients by performing timely medical interventions. VAF, UMB, and blast percentage fluctuation of other cases are shown in Supplementary Fig. 14.

We also tried to predict relapse using the VAF fluctuation of one >0.1% VAF pathogenic mutation in each case instead of using the UMB approach. In the 20 cases with at least 2 CR samples, only 12 (60%) have mutations with >0.1% VAF at diagnosis, including the *NPM1*-positive and some double negative cases. In the 12 cases, the mutation with the highest VAF at diagnosis was used for relapse prediction. We observed an HR of 3.0 with $p = 0.45$ ($N = 12$) in the CIR analysis (Supplementary Fig. 15a); ROC for predicting relapse within 30 weeks of the second CR timepoint was 0.71 (Supplementary Fig. 15b). The prediction accuracy was lower than that of the UMB approach.

UMB analysis using different panel sizes and the impact of existing and emerging mutations

Currently, 738 nucleotide loci in 22 genes were used to calculate UMB, and we next investigated the effect of the number of genes included in the UMB panel for relapse prediction. The 22 genes were ranked according to the mutation frequency in AML in the MCG database (Fig. 4a). Using all samples from CR timepoint 1, we calculated the average mutation rate per base for each gene. The mean mutation rate of the first 11 genes in Fig. 4a was greater than that of the last 11 genes ($p = 0.063$ when comparing the two groups by t-test), suggesting that genes with high MCG ranking might be more prone to mutation in AML (Fig. 4b). This indicates that it could be beneficial to use AML-associated gene hotspots when calculating UMB, and these locations capture mutations more efficiently than other locations in the genome. Next, we calculated UMB using different panel sizes and investigated the discrimination between UMB+ and UMB− groups in CIR analysis. We calculated UMB using mutations in top m genes ($m \le 22$) based on the rank in Fig. 4a. The HR of UMB+ to UMB− tended to be positively correlated with the number of genes ($R^2 = 0.87$) (Fig. 4c). When the number of genes is not lower than 15, $p < 0.05$ was achieved in CIR analysis (Supplementary Fig. 16). We also performed relapse prediction within 30 weeks of CR timepoint 2 using UMB with different panel sizes (Fig. 4d). ROC AUC was also positively correlated with the number of genes ($R^2 = 0.92$), with ROC AUC > 0.80 when the number of genes m ≥ 10. Overall, more accurate predictions were obtained using larger panels, with the smallest feasible panel size being about 15 genes.

During the tumor disease progression, the profile of mutation is constantly changing due to subclone evolution. Some mutations that were prominent at diagnosis may decline significantly or even disappear after treatment because they are sensitive to therapeutic approaches, while new mutations and previously low VAF subclones that are resistant to therapy may become dominant. We categorized mutations during CR into existing and emerging types and studied their roles in UMB. Existing mutations are those detected in diagnosis samples and also present during CR; emerging mutations are those not

https://doi.org/10.1038/s41467-024-54254-6

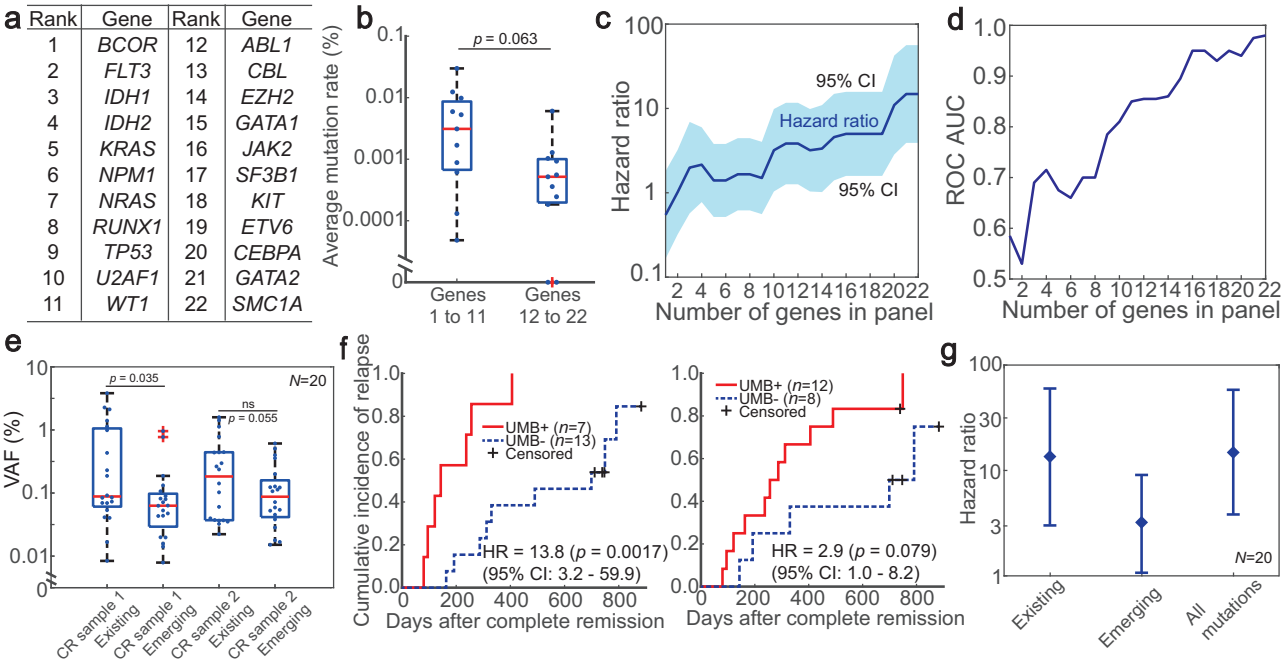

Fig. 4 | UMB-based relapse analysis using subsets of mutations. a Genes in panel ranked by mutated frequency in My Cancer Genome (MCG) database[20]. **b** Comparison of average gene mutation rate in 24 samples collected at CR timepoint 1 between rank 1–11 and 12–2 genes in (**a**). Consistent with the MCG database, the average mutation rate of rank 1–11 genes was higher than that of rank 12–22 genes ($p = 0.063$, two-tailed t-test). Central lines represent median values, and the bottom and top edges of the box represent the 25th and 75th percentile. The whiskers extend to 1.5 times interquartile range and data points beyond the whiskers are marked by + symbols. **c** Using a different number of genes for relapse prediction. Number of genes in panel = m indicates that the top m genes in (**a**) are used in UMB calculation. HR of UMB-based CIR analysis improved as the number of genes increased ($N = 20$). The central line shows the HR values and the shaded region represents 95% CI. **d** ROC AUC of predicting relapse within 30 weeks of CR timepoint 2 improved as the number of genes used for UMB calculation increased ($N = 20$). **e** VAF summary of existing and emerging mutations. Mutations observed both at diagnosis and during CR were classified as existing; those only observed during CR but not at diagnosis were classified as emerging ($N = 20$, $p = 0.035$ for existing, $p = 0.055$ for emerging, two-tailed t-test). Boxplot marker definitions are the same as in (**b**). **f** CIR analysis of UMB+ and UMB− groups using only existing or emerging mutations for UMB calculation. ($p = 0.0017$ for existing, $p = 0.079$ for emerging, two-tailed log-rank test). **g** HR of CIR analysis using existing, emerging, or all mutations for UMB calculation ($N = 20$). The diamond symbols represent HR, and the whiskers represent 95% CI. Source data are provided as a Source Data file.

detected in diagnosis samples and newly emerged during CR. In CR timepoint 1 samples, the existing mutations had higher VAF than the emerging mutations ($p = 0.035$ by t-test), but in CR timepoint 2 samples, the two groups were closer ($p = 0.055$), indicating there is a tendency for emerging mutation to rise during CR (Fig. 4e). When using only existing or emerging mutations to calculate UMB and perform CIR analysis, the HR for UMB+ to UMB− were both greater than 1. The HR using existing mutations was greater than that of emerging mutations, indicating that existing mutations are more effective than emerging mutations in predicting relapse. The result implies that relapse in the time period of this study (within 883 days) is more often associated with uncleared subclone present at diagnosis than with newly generated mutations (Fig. 4f). It was worth noting that the standard UMB analysis considering all mutations exhibited the best performance (Fig. 4g) and streamlined bioinformatics analysis.

UMB-based AML risk assessment in ELN intermediate risk group and hypomethylating agents + venetoclax cases

The current consensus for AML risk assessment is the 2022 ELN risk classification, which divides patients into three risk categories based on their genetics at initial diagnosis: favorable, intermediate, and adverse. In this study, the proportion of favorable, intermediate, and adverse were 16.7%, 66.7%, and 16.7%, respectively (Fig. 5a). An effective MRD monitoring method is especially desired for cases assigned to the intermediate group. To demonstrate that UMB is applicable to these patients, we performed UMB analysis specifically on the ELN intermediate group. In CIR analysis, the HR of UMB+ to UMB− according to the log-rank test was 93.5 (95% CI, 8.7–1008.3), with $p = 0.0017$ (Fig. 5b). The median difference in CR duration between the two groups was 525 days (200 days and 725 days). Using b_{CR2}/b_{CR1} to predict whether recurrence would occur within 30 weeks of CR timepoint 2 showed an ROC AUC of 1.0 (Fig. 5c). The UMB-based risk assessment can potentially be used to guide medical interventions for the higher-risk cases promptly.

Among the cases in this study, 62.5% were treated with hypomethylating agents in combination with venetoclax (Fig. 5d) as these patients were evaluated as unsuitable for intensive chemotherapy based on a comprehensive geriatric assessment[30]. Here the hypomethylating agents and venetoclax group had a mean age of 69 years, which was higher than that of the intensive chemotherapy group (59 years old). To demonstrate compatibility with cases receiving the combination therapy, we performed UMB analysis on the hypomethylating agents and venetoclax group. In CIR analysis, UMB+ had a significant effect with HR = 8.9 (95% CI, 2.2–36.0), $p = 0.0069$ (Fig. 5e). The median difference in CR duration between the two groups was 539 days (200 days and 739 days). Using b_{CR2}/b_{CR1} to predict whether recurrence would occur within 30 weeks of CR timepoint 2 had ROC AUC of 0.97 (Fig. 5f). Collectively, we demonstrate that UMB may be used as companion diagnostics for monitoring treatment outcomes and assessing disease progression in cases with various genetic abnormality types, ELN risk groups, or treatments. UMB-based MRD assessment may further assist medical decision-making, such as adjustment of treatment and dosages.

https://doi.org/10.1038/s41467-024-54254-6

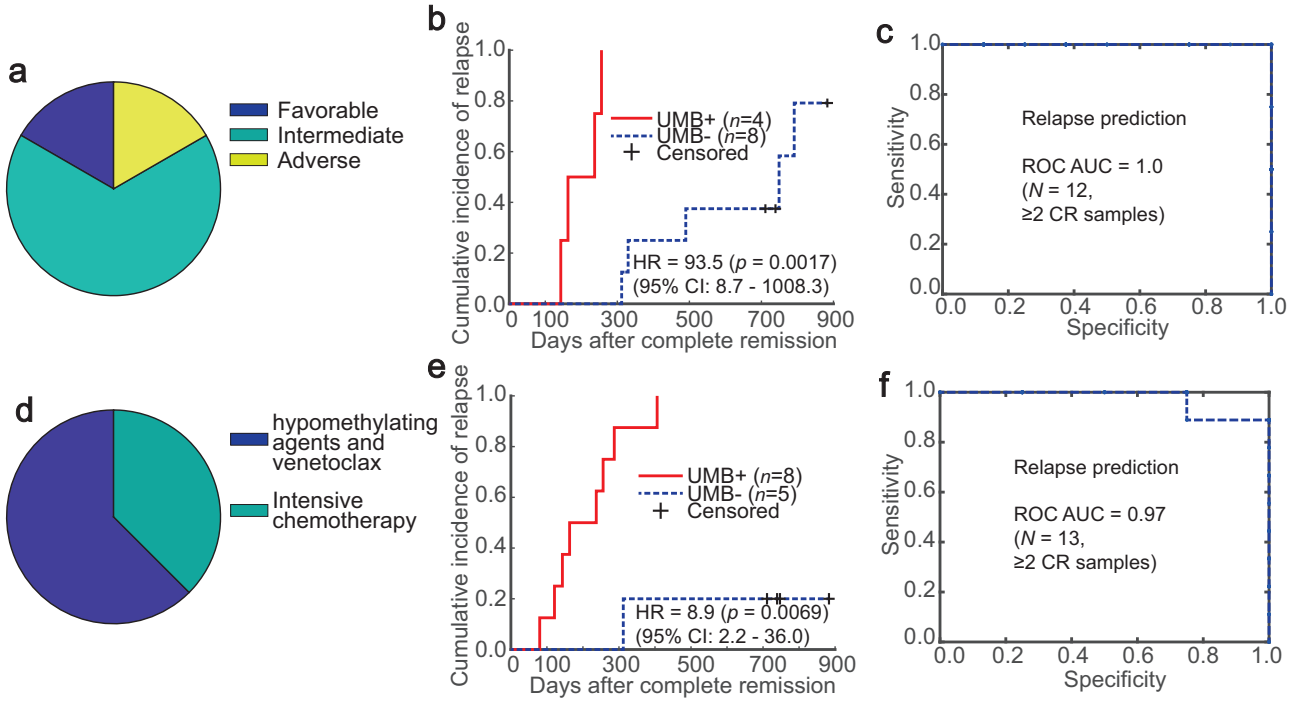

Fig. 5 | UMB-based relapse prediction is applicable to patients in the ELN intermediate risk group or treated with targeted therapy. a Proportion of European Leukemia Net (ELN) risk groups in tested cases ($N = 24$). **b** UMB analysis classifies the ELN intermediate group into high and low risk of relapse. UMB+ group showed shorter median relapse time (200 days) than UMB− did (725 days) ($p = 0.0017$, two-tailed log-rank test). **c** ROC of predicting relapse within 30 weeks of CR timepoint 2 in the ELN intermediate group. **d** Proportion of intensive chemotherapy, and hypomethylating agents + venetoclax in tested cases ($N = 24$). **e** UMB analysis in hypomethylating agents and venetoclax cases. UMB+ cases were at higher risk of relapse (median relapse time 200 days) than UMB− cases (739 days) ($p = 0.0069$, two-tailed log-rank test). **f** ROC of predicting relapse within 30 weeks of CR timepoint 2 in hypomethylating agents and venetoclax groups. Source data are provided as a Source Data file.

Discussion

Sensitive detection of MRD in AML plays a critical role in both disease monitoring and clinical decision-making. Current methods based on analysis of gene aberrations have enabled MRD detection only in patients with *NPM1* mutations or gene fusions and are not widely applicable. We established an ultra-low VAF mutation enrichment NGS panel based on QBDA technology to target hotspot mutation regions for quantitation of <0.01% VAF mutations. Based on this technique, we were able to accurately detect small changes in a patient's mutation profile during CR and quantify this metric using ultra-sensitive mutation burden (UMB). Consistent with our hypothesis that UMB may increase before relapse, we performed CIR analysis and observed that the HR for UMB+ (i.e., the group of patients with elevated UMB during CR) to UMB- was 14.8 with $p < 0.001$. ROC for predicting relapse within 30 weeks of the second CR timepoint was 0.98. Different initial mutation types and treatment regimens did not affect the prediction of relapse. We demonstrated that UMB prediction of relapse clearly requires ultra-sensitive mutation detection with VAF LoD <0.01%. Mutation burden only considering variants with allele frequency >0.01% failed to distinguish the positive group from the negative group in CIR analysis.

We performed OS analysis and observed that the HR for UMB+ to UMB- was 2.9 with $p > 0.05$. ROC for predicting survival within 50 weeks of the second CR timepoint was 0.66. The effect of UMB in predicting survival was not as good as relapse, as OS is related to several factors, including the patient's age, functional status, nutritional status, hyperleukocytosis, comorbidities, genetic and molecular characteristics, treatment after relapse and so on[31]. We have previously conducted research on the AML prognostic assessment system[30,32], and hope to combine it with UMB to better assess the prognosis of AML patients in the future.

We also performed a relapse analysis using MFC and compared the results to UMB. Based on the ELN guidelines[26], we used the first CR timepoint for prediction and categorized AML cases based on whether there were ≥0.1% leukemic blasts in the bone marrow. In CIR analysis, there was no significant difference in CR duration between the 2 groups according to the log-rank test (Supplementary Fig. 17a, b). Therefore, using the MFC absolute blast percentage alone did not seem effective in separating early and late relapse cases. Based on our analysis of UMB, we speculated that the fluctuation could be more effective in relapse prediction than absolute values, thus we next analyzed MFC data based on the ratio between the 2 CR timepoints. We categorized AML cases into two groups based on whether the MFC blast percentage had increased during CR. Cases with at least 2 CR timepoints ($N = 20$) were used in the analysis. In CIR analysis, we observed HR = 4.0 (95% CI, 1.3–13.0), and $p = 0.040$, which was significant but not as good as UMB (Supplementary Fig. 17c). We next tried to predict whether relapse would occur within 30 weeks of CR timepoint 2 using the change in blast percentage, and observed that the AUC of the ROC curve was 0.53 ($N = 20$, Supplementary Fig. 17d). The predictive accuracy was lower than using UMB (AUC = 0.98).

Due to the relatively small cohort size, we did not perform relapse prediction with an independent test dataset. A larger cohort size with an independent test dataset will help to further evaluate the clinical applicability of the UMB-based strategy.

The current clinical standard practice for AML sample collection and storage is to use BM samples. We also analyzed 2 CR PBMC samples from 2 different cases by QBDA, and the results preliminarily indicate that the mutation profiles are similar in PBMC and BM samples, and the UMB of PBMC is lower than that of the corresponding BM (Supplementary Fig. 18 and Supplementary Data 9). We are also planning to collect more PBMC samples for validation in our follow-up work.

We hypothesize that a significant portion of stochastic mutations caused by genome instability in tumor cells have VAFs lower than 0.01%, so traditional mutation detection methods, such as conventional NGS with LoD around 1% and standard commercially available UMI-based NGS with LoD around 0.1%, were unable to effectively detect small changes in UMB. Compared to other ultra-sensitive UMI-based methods (e.g., DuplexSeq[33], SaferSeqS[34], etc.), QBDA enriches any variants that differ from the WT sequence by multiplexed PCR, which has the advantages of low sequencing depth requirement, accurate quantification, high conversion yield of the original biological DNA molecules, fast library preparation and bioinformatics for use in clinical research. ddPCR-based methods also have sensitive LoD and are relatively fast compared to NGS-based methods, but the mutation detection range is generally lower than 10 mutation sites, which is insufficient to calculate UMB.

We validated the effectiveness of a panel containing 22 genes for UMB calculation and found that the prediction accuracy (HR, ROC AUC) improved with increasing gene number. The current panel has not clearly arrived plateau in predicting accuracy, thus using a panel that covers a larger region may further improve the performance. On the other hand, using a panel covering a smaller genomic region, it may be possible to achieve UMB quantitation using a multiplexed real-time PCR-based approach with a simpler experimental process than NGS.

The current panel evaluates both AML-related driver mutations and genome instability-related non-driver mutations. In this manuscript, we used the sum of both as a measure of residual disease and did not distinguish the contribution components in UMB. Based on the results, this UMB strategy was accurate in predicting relapse regardless of the mutation type the patient carries. Comparing the performance of a non-hotspot, instability-based panel with the standard hotspot panel might provide a further understanding of the biological mechanisms of UMB.

All the patients selected in this study were AML patients, but the hypothesis of higher UMB in tumor cells may be applicable in other cancer types as well. In liquid tumors other than AML, it is likely that relapse can be predicted by monitoring UMB changes in BM or blood samples. The UMB method may also be applied in monitoring the relapse of solid tumors using different sample types, e.g. urine samples (tumors of the urinary system), fecal samples (colorectal cancer), or cell-free DNA in the blood. In addition, the mutation profile varies in different cancers, thus designing panels covering highly mutated genes and hotspot regions of the target cancer type may be important for UMB-based analysis.

Methods

Ethical statement

Our research complies with all relevant ethical regulations. All human studies were performed under the approved protocol of the Beijing Hospital Research Ethics Committee (approval letter No. 2018BJYYEC-154-02).

Oligonucleotides and reagents

Oligonucleotides and plasmids were ordered from Sangon Biotech and RuiBiotech (see Supplementary Data 1 and 2 for nucleotide sequences); Hot Start Flex DNA Polymerase and Dual Index Primers Set1 were purchased from NEW ENGLAND BioLabs Int; dNTP Mix was purchased from Vazyme; AMPure XP beads were purchased from Beckman Coulter Life Sciences; Myeloid DNA Reference Standard was purchased from Horizon Discovery. Healthy human PBMC samples were voluntarily provided by the authors of this work. MV-4-11 cell line was purchased from Fenghuishengwu, China.

Patient selection and sample collection

Twenty-four AML patients diagnosed at the Department of Hematology, Beijing Hospital, Beijing, China, between January 2018 and June 2023 were collected for this study (clinical information summarized in Supplementary Data 10). AML was diagnosed according to the World Health Organization (WHO) diagnostic criteria[35]. All patients had achieved morphologic CR, defined by the 2022 ELN recommendation after standard induction chemotherapy.

A total of 83 BM aspirate samples (including 24 samples at diagnosis, 24 samples of CR timepoint 1, 20 samples of CR timepoint 2, 2 samples of CR timepoint 3, and 13 samples at relapse) were collected for this study (sampling time information summarized in Supplementary Data 11). All samples were stored in a 1.5 mL RNase-free polypropylene (PP) tube at −80 °C until DNA was extracted. Genomic DNA of all BM samples was extracted from samples using Lab-Aid 824 Nucleic Acid Extraction instrument (Zeesan, China) and QIAsymphony DNA Mini *Kit* (Qiagen, Hilden, Germany) according to the manufacturer's guidelines and quantified with NanoDrop One (Thermo Fisher). The BM aspirate smear was stained using the Wright–Giemsa stain, the blast percentage was determined by counting 200 nucleated BM cells under a microscope.

The retrospective study was approved by the Beijing Hospital Research Ethics Committee (approval letter No. 2018BJYYEC-154-02), and written informed consents were obtained from all participants in accordance with Institutional Review Board guidelines and the Declaration of Helsinki. Clinical information was obtained from electronic medical records. All data were deidentified. This study did not provide participant compensation. Gender was not taken into account in the experimental design. The sex and/or gender of participants were ascertained according to self-report. Sex or gender analysis was not conducted in the study, as this study aims to develop a longitudinal UMB monitoring strategy for MRD analysis of various AML patients, regardless of their gender.

QBDA and *FLT3*-ITD library preparation and sequencing

QBDA library preparation was conducted according to our previous study[17]. 1000 ng of gDNA was used to prepare sequencing libraries and QBDA Library preparation consisted of three PCR reactions: UMI addition and pre-amplification, BDA amplification, and index PCR, all performed on T100 Thermal Cyclers (Bio-Rad). DNA sample was mixed with the specific forward primer (SfP), and specific reverse primer (SrP) and amplified using Phusion polymerase. The concentration for each SfP and SrP was 15 nM. Thermal cycling condition was: 98 °C:30 s − (98 °C:10 s − 63 °C:30 min − 72 °C:60 s) × 2 − (98 °C:10 s − 63 °C:20 s − 72 °C:60 s) × 2 − (98 °C:10 s − 71 °C:20 s − 72 °C:60 s) × 5 − (72 °C:5 min) − 4 °C:hold. During the last 5 min of the second 30 min at 63 °C, 1.5 μM of universal forward primer (UfP) and 1.5 μM of universal reverse primer (UrP) were added while keeping the reactions inside the thermal cycler. Double-side size selection (0.3×, 1.6× ratio) was performed to remove long gDNA, followed by another 1.6× AMPure XP beads purification. Next, BDA forward primer (fP), BDA blocker, Phusion polymerase, dNTPs, and PCR buffer were mixed with the purified PCR product for BDA amplification. Thermal cycling condition was: 98 °C:30 s − (98 °C:10 s − 63 °C:5 min − 72 °C:60 s) × 23 − 4 °C:hold. The reaction mixture was purified using 1.8× AMPure XP beads. BDA adapter primer (Adapter_fP) and UrP are mixed with the purified PCR reaction mixture and amplified. Thermal cycling condition was: 98 °C:30 s − (98 °C:10 s − 63 °C:5 min − 72 °C:1 min) × 2 − 4 C:hold. The reaction mixture was purified using 1.6× AMPure XP beads. Standard NGS index PCR and library normalization are performed afterward. Sequencing was performed using NovaSeq (PE150, Illumina, San Diego, CA, USA).

FLT3-ITD library preparation consisted of PCR amplification and standard NGS, all performed on a T100 Thermal Cycler (Bio-Rad). First, a 100 ng DNA sample was mixed with the barcoded forward primer (BfP) and barcoded reverse primer (BrP) and amplified using high-fidelity Phusion polymerase. Each sample corresponds to a unique sample barcode. The final concentration for each BfP and BrP was 500 nM. Thermal cycling condition was: 98 °C:30 s − (98 °C:10 s − 63 °

C:60 s − 72 °C:60 s) × 15 − 4 °C:hold. The reaction mixture was purified using AMPure XP beads (1.8× ratio) once, followed by standard NGS index PCR. Libraries were normalized, and sequencing was performed on NovaSeq (PE250).

QBDA and *FLT3*-ITD data analysis

QBDA data analysis, including NGS data preprocessing, UMI-based mutation calling, and mutation filtering steps, was conducted using the analysis pipeline for QBDA variant calling available from GitHub (https://github.com/wyt228/UMB) according to our previous study[17]. For *FLT3*-ITD data analysis, adapter sequences were removed, and sample barcode sequences were extracted (4 nt in BfP, and 4 nt in BrP) using custom Python software. The processed sequences were then aligned to *FLT3* amplicons using Bowtie2[36]. Sample barcode analysis and ITD calling were performed using custom Python software. In this manuscript, the VAF of a specific variant was defined as VAF = 100% × M_v/M_t, where M_v is the number of DNA molecules bearing this variant, and M_t is the total number of DNA molecules at this locus, including wild type and all sequence variants. All VAF values were expressed as percentages. Detailed calculations of QBDA mutation calls can be found in our previous publication[17].

Gene fusion detection

Take 2 mL of BM samples anticoagulated with EDTA and separate mononuclear cells (PBMCs) using Ficoll solution (Solarbio Life Sciences, Beijing, China), according to the manufacturer's protocol. Use the TRIzol method (provided by Invitrogen in the United States) to extract RNA from nucleated cells, reverse transcription, and synthesize cDNA. AML-related fusion genes were detected using Promega GoTaq® GreenMaster Mix according to the manufacturer's protocol. The tested fusion genes include *MLL-FOXO4(MLL-AFX1), MLL-MLLT4(MLL-AF6), MLL-ELL, MLL-EPS15(MLL-AF1P), MLL-MLLT6(MLL-AF17), MLL-MLLT10(MLL-AF10), MLL-PTD, TCF3-PBX1(E2A-PBX1), TCF3-HLF(E2A-HLF), ETV6-RUNX1(TEL-AML1), STIL-TAL1, RUNX1-RUNX1T1(AML1-ETO), RUNX1-MDS1-EVI1(AML1-MDS1-EVI1), FUS-ERG(TLS-ERG), PML-RARA, MLL-AFF1(MLL-AF4), MLL-MLLT1(MLL-ENL), MLL-MLLT3(MLL-AF9), MLL-MLLT11(MLL-AF1Q), CBFB-MYH11, BCR-ABL1, ETV6-ABL1(TEL-ABL1), ETV6-PDGFRB(TEL-PDGFRB), ETV6-MN1(TEL-MN1), DEK-NUP214(DEK-CAN), SET-NUP214(SET-CAN), RUNX1-RPL22P1(AML1-EAP), ZBTB16-RARA(PLZF-RARA), NPM1-RARA, and NPM1-MLF1*.

Multiparameter flow cytometry (MFC)

MFC assays were performed to assess MRD in BM aspirates. The assay aimed to detect blasts or immature cells based on the immunephenotypic expression of molecular markers associated with AML. Mononuclear Cells (MNCs) were isolated from the BM by density centrifugation. Approximately 1×10^6 MNCs per sample were then incubated with different fluorescein-labeled antibodies (purchased from BD Biosciences). Antibodies were diluted with phosphate-buffered saline according to the manufacturer's instructions. Dilution ratio was 1:20 for CD45-V500 (560777), 1:5 for CD34-PerCP (347203), 1:20 for CD117-PE-Cy7 (339195), 1:5 for CD33-APC (561817), 1:5 for CD13-PE (347406), 1:5 for HLA-DR-FITC (347363), 1:5 for CD36-APC (550956), 1:5 for CD64-PE (558592), 1:20 for CD56-V450 (560360), 1:20 for CD4-APC-H7 (560158), 1:5 for CD123-PE (340545), 1:20 for CD16-FITC (561308), 1:20 for CD38-V450 (561378), 1:20 for CD14-APC-H7 (560270), and 1:20 for CD19-PE-CY7 (341113). Cells were then detected by a flow cytometer (BD FACSCantoTM II) and analyzed by FlowJo software v10.6.2 (BD Biosciences).

Whole-exosome sequencing (WES)

Briefly, DNA was extracted from fresh frozen tissue blocks using a QIAGEN QIAamp DNA Mini Kit. For WES, library construction and whole-exome capture of genomic DNA were performed using the SureSelectXT Reagent Kit (Agilent) and SureSelectXT Human All Exon V6 Kit (Agilent). The captured DNA was then sequenced on an Illumina HiSeq 2500 sequencing platform with 150-bp paired-end sequencing. Sequencing and bioinformatics services were provided by Clinical Biobank, Beijing Hospital, and Novogene.

Statistical analyses

All statistical analyses were performed with MATLAB R2018b (MathWorks Inc, MathWorks, Natick, USA), Python2.7 (Python Software Foundation, Netherlands), and Microsoft Excel (Microsoft Corp, Redmond, WA, USA). Cumulative incidences of relapse (CIR) were defined from the date of complete remission to the date of relapse or death. Overall survival (OS) was defined from the date of diagnosis to the date of death or last follow-up. CIR and OS curves were plotted using the Kaplan–Meier method and analyzed with the log-rank test. A Cox proportional hazards analysis was used to evaluate the relationship between the relapse (or survival) outcome and exposure variables. Results were expressed as the HR with a 95% confidence interval (95% CI). ROC AUC values were calculated as the area under the ROC curve. A $p < 0.05$ was considered statistically significant. ns, not significant.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The sequences of the DNA oligonucleotides used for AML QBDA panels and *FLT3*-ITD, QBDA, and *FLT3*-ITD test results, and clinical sample information are included in Supplementary Data 1–2, 4–5, and 10–11. Source data are provided with this paper. The sequencing data has been deposited in the National Genomics Data Center (NGDC) GSA for human database and can be found at https://ngdc.cncb.ac.cn/gsa-human/browse/HRA006832. The sequencing data is publicly available. Source data are provided with this paper.

Code availability

NGS data analysis pipeline for QBDA variant calling is available from GitHub (https://github.com/wyt228/UMB) and can be found at https://doi.org/10.5281/zenodo.13908530.

References

1. Klco, J. M. et al. Association between mutation clearance after induction therapy and outcomes in acute myeloid leukemia. *JAMA* **314**, 811–822 (2015).
2. Cornelissen, J. J. et al. The European LeukemiaNet AML Working Party consensus statement on allogeneic HSCT for patients with AML in remission: an integrated-risk adapted approach. *Nat. Rev. Clin. Oncol.* **9**, 579–590 (2012).
3. Walter, R. B. et al. Resistance prediction in AML: analysis of 4601 patients from MRC/NCRI, HOVON/SAKK, SWOG and MD Anderson Cancer Center. *Leukemia* **29**, 312–320 (2015).
4. Khwaja, A. et al. Acute myeloid leukaemia. *Nat. Rev. Dis. Prim.* **2**, 16010 (2016).
5. Cocciardi, S. et al. Clonal evolution patterns in acute myeloid leukemia with NPM1 mutation. *Nat. Commun.* **10**, 2031 (2019).
6. Jongen-Lavrencic, M. et al. Molecular minimal residual disease in acute myeloid leukemia. *N. Engl. J. Med.* **378**, 1189–1199 (2018).
7. Ivey, A. et al. Assessment of minimal residual disease in standard-risk AML. *N. Engl. J. Med.* **374**, 422–433 (2016).
8. Terwijn, M. et al. High prognostic impact of flow cytometric minimal residual disease detection in acute myeloid leukemia: data from the HOVON/SAKK AML 42A study. *J. Clin. Oncol.* **31**, 3889–3897 (2013).
9. Jaso, J. M., Wang, S. A., Jorgensen, J. L. & Lin, P. Multi-color flow cytometric immunophenotyping for detection of minimal residual disease in AML: past, present and future. *Bone Marrow Transpl.* **49**, 1129–1138 (2014).

10. Li, S.-Q. et al. An LSC-based MRD assay to complement the traditional MFC method for prediction of AML relapse: a prospective study. *Blood* **140**, 516–520 (2022).
11. Heuser, M. et al. 2021 Update on MRD in acute myeloid leukemia: a consensus document from the European LeukemiaNet MRD Working Party. *Blood* **138**, 2753–2767 (2021).
12. Hasserjian, R. P., Steensma, D. P., Graubert, T. A. & Ebert, B. L. Clonal hematopoiesis and measurable residual disease assessment in acute myeloid leukemia. *Blood* **135**, 1729–1738 (2020).
13. Hammond, D. et al. Response patterns and impact of MRD in patients with IDH1/2-mutated AML treated with venetoclax and hypomethylating agents. *Blood Cancer J.* **13**, 148 (2023).
14. Grob, T. et al. Prognostic value of FLT3-internal tandem duplication residual disease in acute myeloid leukemia. *J. Clin. Oncol.* **41**, 756–765 (2023).
15. Wang, J. et al. Detection of measurable residual disease may better predict outcomes than mutations based on next-generation sequencing in acute myeloid leukaemia with biallelic mutations of CEBPA. *Br. J. Haematol.* **190**, 533–544 (2020).
16. Onecha, E. et al. Monitoring of clonal evolution of acute myeloid leukemia identifies the leukemia subtype, clinical outcome and potential new drug targets for post-remission strategies or relapse. *Haematologica* **106**, 2325–2333 (2021).
17. Dai, P. et al. Calibration-free NGS quantitation of mutations below 0.01% VAF. *Nat. Commun.* **12**, 6123 (2021).
18. Murdock, H. M. et al. Impact of diagnostic genetics on remission MRD and transplantation outcomes in older patients with AML. *Blood* **139**, 3546–3557 (2022).
19. Wu, L. R., Chen, S. X., Wu, Y., Patel, A. A. & Zhang, D. Y. Multiplexed enrichment of rare DNA variants via sequence-selective and temperature-robust amplification. *Nat. Biomed. Eng.* **1**, 714–723 (2017).
20. Holt, M.E. et al. My Cancer Genome: Coevolution of Precision Oncology and a Molecular Oncology Knowledgebase. *JCO Clin Cancer Inform* **5**, 995–1004 (2021).
21. Sondka, Z. et al. COSMIC: a curated database of somatic variants and clinical data for cancer. *Nucleic Acids Res* **52**, D1210–d1217 (2024).
22. Song, P. et al. Limitations and opportunities of technologies for the analysis of cell-free DNA in cancer diagnostics. *Nat. Biomed. Eng.* **6**, 232–245 (2022).
23. Narayan, A. et al. Ultrasensitive measurement of hotspot mutations in tumor DNA in blood using error-suppressed multiplexed deep sequencing. *Cancer Res.* **72**, 3492–3498 (2012).
24. Newman, A. M. et al. Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat. Biotechnol.* **34**, 547–555 (2016).
25. Tsai, C. H. et al. Clinical implications of sequential MRD monitoring by NGS at 2 time points after chemotherapy in patients with AML. *Blood Adv.* **5**, 2456–2466 (2021).
26. Döhner, H. et al. Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. *Blood* **140**, 1345–1377 (2022).
27. Ranieri, R. et al. Current status and future perspectives in targeted therapy of NPM1-mutated AML. *Leukemia* **36**, 2351–2367 (2022).
28. Grimwade, D. et al. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood* **116**, 354–365 (2010).
29. El Chaer, F., Hourigan, C. S. & Zeidan, A. M. How I treat AML incorporating the updated classifications and guidelines. *Blood* **141**, 2813–2823 (2023).
30. Feng, R. et al. Low-dose decitabine for previously untreated acute myeloid leukemia ineligible for intensive chemotherapy aged 65 years or older: a prospective study based on comprehensive geriatric assessment. *Ther. Adv. Hematol.* **14**, 20406207231208979 (2023).
31. Pollyea, D. A. et al. Acute myeloid leukemia, version 3.2023, NCCN clinical practice guidelines in oncology. *J. Natl Compr. Canc Netw.* **21**, 503–513 (2023).
32. Zhang, C. et al. Treatment patterns and a prognostic scoring system for elderly acute myeloid leukemia patients: a retrospective multicenter cohort study in China. *Cancer Biol. Med.* **19**, 871–883 (2021).
33. Schmitt, M. W. et al. Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl Acad. Sci. USA* **109**, 14508–14513 (2012).
34. Cohen, J. D. et al. Detection of low-frequency DNA variants by targeted sequencing of the Watson and Crick strands. *Nat. Biotechnol.* **39**, 1220–1227 (2021).
35. Arber, D. A. et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **127**, 2391–2405 (2016).
36. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

Acknowledgements

This study is supported by the National Natural Science Foundation of China (Grant no. 22274101 to L.R.W., and 22304004 to P.D.), Beijing Natural Science Foundation to H.L. (No. 7232137), CAMS Innovation Fund for Medical Sciences (CIFMS) to H.L. (No.2021-I2M-C&T-A-020), National High-Level Hospital Clinical Research Funding to H.L. (No.BJ-2022-127), Capital Medical University startup fund to L.R.W., and Beijing Friendship Hospital Seed Project Funding to S.Z. (No.YYZZ202311).

Author contributions

Y. Wu, Y. Wang, P.D., H.L., and L.R.W. conceived the project. Y. Wu and K.X. conducted the sequencing experiments. S.Z., R.F., T.W., J.B., and H.L. collected clinical AML samples and patient information. Y. Wu, K.X., X.Z., and L.R.W. designed the oligonucleotide sequences. Y. Wu, S.Z., R.F., K.X., P.D., H.L., and L.R.W. analyzed the data. Y. Wu, S.Z., P.D., H.L., and L.R.W. wrote the paper with input from all authors.

Competing interests

There is a patent pending (No. 202410760423.7) with the title of A device for predicting disease recurrence based on ultra-sensitivity mutation burden. L.R.W. and Y. Wu are the inventors. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-54254-6.

Correspondence and requests for materials should be addressed to Peng Dai, Hui Liu or Lucia Ruojia Wu.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.