nature genetics

Article

https://doi.org/10.1038/s41588-024-01929-x

Multiomic profiling identifies predictors of survival in African American patients with acute myeloid leukemia

Received: 20 December 2023

Accepted: 23 August 2024

Published online: 4 October 2024

Check for updates

A list of authors and their affiliations appears at the end of the paper

Genomic profiles and prognostic biomarkers in patients with acute myeloid leukemia (AML) from ancestry-diverse populations are underexplored. We analyzed the exomes and transcriptomes of 100 patients with AML with genomically confirmed African ancestry (Black; Alliance) and compared their somatic mutation frequencies with those of 323 self-reported white patients with AML, 55% of whom had genomically confirmed European ancestry (white; BeatAML). Here we find that 73% of 162 gene mutations recurrent in Black patients, including a hitherto unreported *PHIP* alteration detected in 7% of patients, were found in one white patient or not detected. Black patients with myelodysplasia-related AML were younger than white patients suggesting intrinsic and/or extrinsic dysplasia-causing stressors. On multivariable analyses of Black patients, *NPM1* and *NRAS* mutations were associated with inferior disease-free and *IDH1* and *IDH2* mutations with reduced overall survival. Inflammatory profiles, cell type distributions and transcriptional profiles differed between Black and white patients with *NPM1* mutations. Incorporation of ancestry-specific risk markers into the 2022 European LeukemiaNet genetic risk stratification changed risk group assignment for one-third of Black patients and improved their outcome prediction.

Acute myeloid leukemia (AML) is characterized by clonal expansion of myeloid progenitor cells due to genomic alterations that inhibit myeloid development[1]. The importance of genomic alterations for the pathogenesis of AML and their prognostic significance was first appreciated through recurrent chromosomal alterations decades ago[1].

Many landmark genomics studies have transformed our understanding of the molecular underpinnings of AML[2–6]. These efforts led to the incorporation of several somatic mutations into clinically applied molecular classification and risk stratification systems[7,8]. In most cancer genomic studies to date (including AML), self-reported white patients predominate. In fact, patients with AML who self-reported as Black include less than 2% of patients in these studies despite Black patients representing 9% of patients with AML[2–6,9–13].

The disparity of AML genomic data between ancestry-diverse populations results in the inequitable application of molecular medicine, which increases the potential for inadequate treatment. Previous studies showed that self-reported Black patients with AML have inferior outcomes compared with white patients[14–18]. Multiple interconnected factors, including structural racism and socioeconomic contributors, are associated with this survival disparity[14–18]. Additionally, the frequency of established recurrent AML-associated genomic alterations and their prognostic significance is different for Black patients[19,20], suggesting the presence of as yet unappreciated differences in the biology of AML in Black patients.

An unbiased depiction of the somatic genetic landscape of AML in Black patients is thus far unknown because previous reports analyzed only known AML-associated genes that were initially identified almost

e-mail: Elaine.Mardis@nationwidechildrens.org; Ann-Kathrin.Eisfeld@osumc.edu

exclusively in white patients. Likewise, resulting differences in the transcriptomes that could provide insights into disease biology have also not been assessed.

Thus, we conducted a comprehensive genomic analysis including whole-exome sequencing (WES) and RNA sequencing (RNA-seq) of paired germline and tumor samples from 100 Black patients with AML with genotype-confirmed African ancestry (hereafter referred to as Black), who were treated with intensive induction chemotherapy on Cancer and Leukemia Group B (CALGB)/Alliance for Clinical Trials in Oncology (Alliance) trials. This resulted in the depiction of the genomic landscape of AML in Black patients, including the identification of ancestry-associated mutations and biological features that differ from traditional European ancestry studies. The somatic mutation frequency in Black patients was compared with results of matched workflow WES of 323 white patients from the Beat AML cohort, 177 (55%) of whom had available genotyping data confirming European ancestry.

Total-transcriptome-based gene expression profiling of tumor RNA from Black and white patients allowed for principal component-based clustering with respect to established molecular drivers and determination of a gene-expression-based inflammation score (iScore)[21]. Moreover, multiomic RNA single-cell sequencing studies of Black and white patients with *NPM1* mutations showed differences in cell type distribution and differentially activated pathways.

Lastly, we compared the clinical outcomes of Black and white patients treated on the same CALGB/Alliance protocols. We found substantial differences between Black and white patients in the prognostic significance of some genomic features used in the 2022 European LeukemiaNet (ELN) genetic risk stratification system[7]. This suggests the need to incorporate ancestry-associated features into this system, thus having direct clinical implications for Black patients.

Results

Variants in coding sequences of Black patients with AML

A WES was conducted on FACS-sorted leukemic blasts and paired B and T cells as a germline equivalent from 100 Black patients (Table 1). Two patients, including one harboring an *MSH6* mutation and another with a *POLE* variant, were hypermutated having 269 and 760 coding variants, respectively, and were excluded from the frequency comparison analyses of nonrecurrently mutated genes. In the remaining 98 patients, we identified 1,971 nonsynonymous coding variants, including single-nucleotide variants (SNVs), small insertions and deletions (indels) and internal tandem duplications (ITDs; Extended Data Fig. 1). The median somatic coding variants per patient was 16 (range: 1–65). There were four predominant mutation signatures, three of which resembled established Catalogue Of Somatic Mutations In Cancer single-base substitution signatures for defective homologous recombination DNA damage repair and defective mismatch repair (Extended Data Fig. 2a–c).

One hundred and sixty-two genes were recurrently mutated (that is, mutated in two or more patients) and 33 genes were mutated in 4% or more patients (Supplementary Table 1). Established AML-associated mutations in *FLT3* (32%), *DNMT3A* (25%), *NRAS* (23%) and *NPM1* (20%) were the most common, followed by mutations in *IDH2* (11%), *RUNX1* (10%), *KRAS* (9%), *ASXL1* (8%) and *WT1* (8%). Notably, *IDH1*, *PHIP*, *TET2* and *TP53* were mutated in 7% of patients (Fig. 1a and Supplementary Table 1). Co-mutational patterns are shown in Extended Data Fig. 3. Additionally, single-cell DNA sequencing (scDNA-seq) was performed for three patients and the observed clonality patterns were consistent with published datasets[22] (Fig. 1b).

In 39 patients with cytogenetically normal AML (CN-AML), *FLT3* (49%), *NPM1* (44%) and *DNMT3A* (44%) alterations were the most common (Extended Data Fig. 4), thereby resembling established molecular distributions in this cytogenetic subset[3,4,23]. However, in 8% of patients with CN-AML, we detected several gene mutations not routinely associated with AML, including *IFNL3*, *MYC* and *PRAMEF15* mutations.

Table 1 | Clinical and cytogenetic characteristics of 100 Black patients with AML included in the integrated genomic profiling study

Characteristic	
Age, years	
Median (range)	43 (18–92)
Sex, *n* (%)	
Female	48 (48)
Hemoglobin, g dl^{-1}	
Median (range)	8.7 (2.3–13.2)
Platelet count, $\times 10^9$ per liter	
Median (range)	48 (5–499)
White blood cell count, $\times 10^9$ per liter	
Median (range)	28.3 (0.4–308.8)
Blood blasts, %	
Median (range)	52 (0–98)
Bone marrow blasts, %	
Median (range)	68 (17–96)
Extramedullary involvement, *n* (%)	22 (25)
2022 ELN genetic risk group, *n* (%)	
Favorable	41 (41)
Intermediate	22 (22)
Adverse	37 (37)
CN-AML, *n* (%)	38 (38)
CBF-AML, *n* (%)	21 (21)
t(8;21), *n* (%)	10 (10)
inv(16), *n* (%)	11 (11)
Complex karyotype, *n* (%)	11 (11)
KMT2A-rearranged, *n* (%)	6 (6)
Myelodysplasia-related AML, *n* (%)[a]	12 (15)

[a]Mutations defining myelodysplasia-related AML include mutations in the *SRSF2*, *SF3B1*, *RUNX1*, *U2AF1*, *ZRSR2*, *ASXL1*, *EZH2*, *BCOR* and *STAG2* genes as per the 2022 ELN[7].

Similarly, in 21 patients with core binding factor AML (CBF-AML), the most common were established mutations in *NRAS* (33%), *KIT* (24%) and *FLT3* (19%). However, variants in *PHIP*, *LRP6*, *PRH2*, *RHPN2* and *ZBTB7A* were all found in 14% of patients, thereby establishing additional, relatively frequent genetic events in our patient cohort (Supplementary Fig. 1).

Among 11 Black patients with a complex karyotype, 45% had mutations in *TP53*. Notably, *TP53* wild-type patients carried variants in targetable genes, including *IDH2* in two and *ALK* in one patient (Supplementary Fig. 2), suggesting the need to test for such alterations in Black patients with complex karyotype as they might be eligible for alternative treatment options.

Gene mutation frequencies in Black and white patients

To compare the gene mutation frequencies of Black and white patients, including assessment of potentially new variants, paired tumor-normal WES data for 323 white patients from the BeatAML 1.0 studies were reanalyzed using the same variant calling workflow, including identical variant allele fraction cutoff (see Extended Data Fig. 5 and Methods for details).

Of 162 recurrently mutated genes in Black patients, only 43 (27%) were recurrently mutated in 323 white patients (BeatAML), while 119 genes (73%) were mutated in one white patient or not mutated. The frequencies of several known and hitherto unreported gene

https://doi.org/10.1038/s41588-024-01929-x

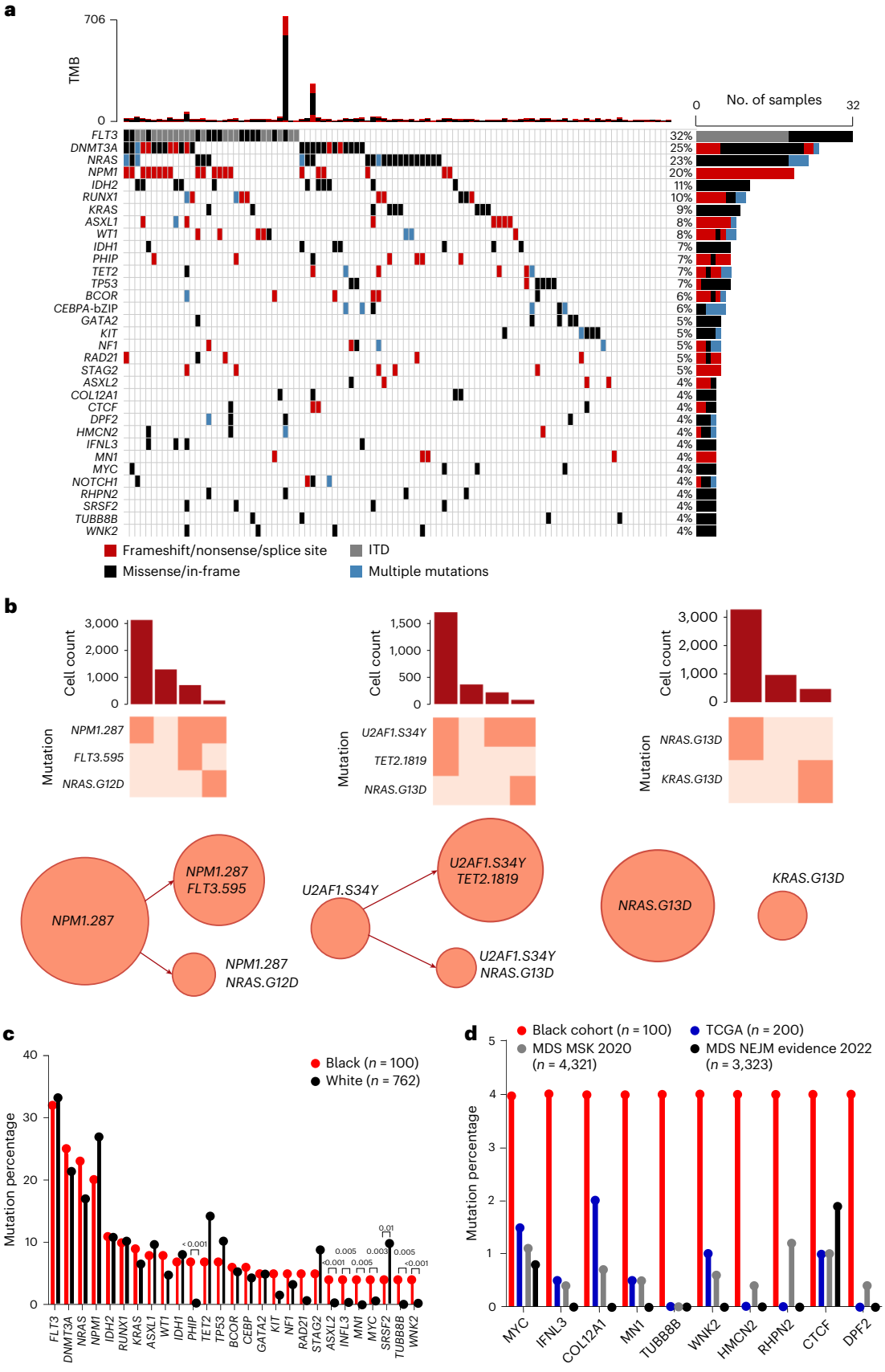


Fig. 1 | Mutational landscape of Black patients with AML who were treated on CALGB/Alliance study protocols. a, Oncoprint showing mutations detected in 4% or more of Black patients by WES. **b**, Clonality of driver mutations detected in three Black patients with AML, identified using scDNA-seq. **c**, Difference in mutation percentage between white (Beat AML) and Black patients in genes mutated in 4% or more of Black patients. The *P* value was calculated using a two-sided Fisher's exact test. **d**, Mutation percentage for genes mutated in 4% of Black patients compared with other major AML sequencing studies. TMB, tumor mutational burden.

https://doi.org/10.1038/s41588-024-01929-x

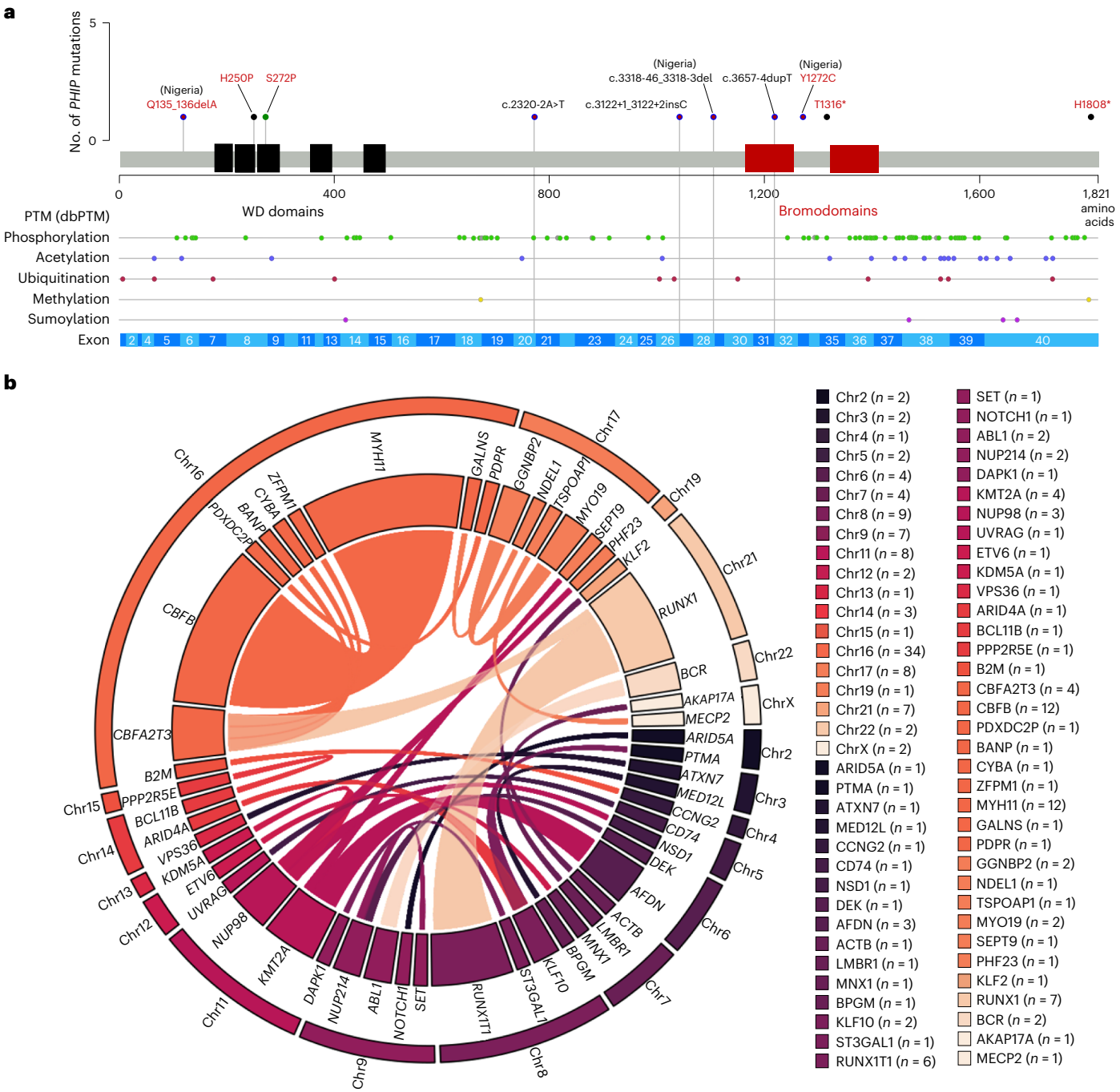


Fig. 2 | Hitherto unreported mutations in the *PHIP* gene and depiction of the fusion gene landscape. a, Schematic of the *PHIP* gene with location of posttranslational modifications and the location of mutations detected in Black patients by WES. Depicted are *PHIP* mutations found in the CALGB/Alliance cohort and mutations detected using targeted sequencing in three patients with AML who were diagnosed and treated in Nigeria. **b**, Circos plot displaying fusion genes detected using RNA-seq-based discovery in Black patients. The width of each arc represents the frequency with numbers indicated in the legend. dbPTM, database of Protein Post-Translational Modifications.

mutations differed between ancestries (Fig. 1c and Supplementary Table 1). Ten genes were mutated in 4% of our cohort, but in only 2% or less of white patients and other previous hallmark sequencing efforts (Supplementary Table 1 and Fig. 1d). Moreover, *PHIP* mutations present in 7% of Black patients were identified in only one white patient (7% versus 0.3%, $P < 0.001$; Fig. 2a and Supplementary Table 2). As germline variants, *PHIP* mutations cause Chung–Jansen syndrome, a rare inherited form of intellectual disability[24]. Somatic *PHIP* mutations were reported in less than 2% of myelodysplastic syndromes, blast crisis chronic myeloid leukemia and clonal hematopoiesis (CH)[6,25,26]. Notably, targeted sequencing of *PHIP* in Black patients consisting of two African patient cohorts we additionally analyzed (Nigerian, 46 patients; South African, 23 patients; Methods) detected three *PHIP* mutations (7%) in Nigerian patients (two males, one female; aged 5, 31 and 42 years, respectively), but none in the South African patients.

Based on large sequencing studies, next-generation sequencing (NGS) panels used in the clinic to test newly diagnosed patients commonly included 56 genes recurrently mutated in AML (Supplementary Table 3). We found that only one of 323 white patients lacked a mutation

in at least one of these genes. In contrast, 10% of Black patients lacked mutations in these genes, thus being 'mutation negative' in current standard clinical molecular testing.

Structural variants

RNA-seq-based fusion gene discovery[27] found 12 *CBFB*::*MYH11*, seven *RUNX1*::*RUNX1T1*, four *KMT2A* rearrangements (three *KMT2A*::*AFDN* and one *KMT2A*::*SEPT9*) and two *BCR*::*ABL* fusions (Fig. 2b and Supplementary Table 4). Four cryptic in-frame fusions affecting *CBFA2T3* were detected, including three chromosome 16 rearrangements, in which *CBFA2T3* was fused to either *BANP*, *CYBA* or *ZFPM1*, and a t(16;21) with *CBFA2T3*::*RUNX1* fusion (Supplementary Table 4). These fusions did not cluster with CBF-AML according to gene expression, indicating a different biological impact (Supplementary Fig. 3). Four cryptic *NUP98* and two *NUP214* fusions were identified, which are recurrent abnormalities in AML. Interestingly, a recurrent *GGNBP2*::*MYO19* fusion involving chromosome 17, which has been identified in giant congenital nevi, but not AML[28], was found in two patients. The copy number alterations detected are listed in Supplementary Tables 5 and 6.

Clonal Hematopoiesis

CH was evaluated through comparison of variant calls identified in WES from both B and T cells, to those found in leukemic blasts, after exclusion of common SNVs. Variants in established CH-associated genes that were present in both B and T cells as well as the leukemic blasts (and absence of other non-CH-associated variants in B and T cells that would be indicative of residual disease, such as *NPM1*) were considered to represent CH. Such CH variants were identified in 11% of Black patients, including known CH-associated variants in *DNMT3A*, *TET2*, *TP53*, *ASXL1*, *MPL*, *CBL*, *SRSF2* and *SMC1A* (Supplementary Table 7). CH occurred across the age spectrum, with a median of 56 years (range: 32–75 years), which is comparable to previously reported frequencies.

Molecular features and clonal dynamics of relapsed AML

To assess clonal dynamics at relapse, we performed integrated genomic profiling (WES + RNA-seq) on 18 Black patients with paired primary and relapse samples available. All patients relapsed with their dominant clone from diagnosis. However, only one patient relapsed with the identical molecular profile, while others gained at least one subclone (Extended Data Fig. 6a). Seven patients had new copy number alterations affecting genes mutated at diagnosis. Five patients had mutational switches at relapse, suggestive of oncogene addiction. One patient gained a new *PHIP* variant and another gained a new *HMCN2* variant, both of which were identified as recurrently mutated in Black patients at diagnosis. Gene expression analysis revealed transcriptional changes of some relapse samples (Extended Data Fig. 6b).

Transcriptional profiles based on oncogenic drivers

In transcriptome-based gene expression profiling and *t*-distributed stochastic neighbor embedding (*t*-SNE) data visualization, Black and white patients clustered together according to known driver mutations (Fig. 3a).

A relatively large, central *t*-SNE cluster contained Black and white patients with myelodysplasia-related mutations as defined by the 2022 ELN[7] (Fig. 3b). Gene expression analysis showed that these patients had expression profiles resembling myelodysplastic neoplasms[29] (Fig. 3c). Notably, Black patients in the myelodysplasia-related cluster were younger than white patients (median, 50 versus 58 years, $P = 0.04$; Fig. 3d,e), suggestive of intrinsic and/or extrinsic dysplasia-causing stressors. Additionally, while white patients within the myelodysplasia-related AML cluster were older than white patients outside this cluster ($P < 0.001$), there was no significant age difference between Black patients ($P = 0.17$), further highlighting the unusually early onset of myelodysplasia-related AML in Black patients.

Clinical features and treatment outcomes

Black patients with AML have poor survival even when considering socioeconomic data, both in population-based analyses and in clinical trials[14–20]. For a more accurate assessment of treatment response and survival, we compared our molecularly characterized cohort of 103 Black patients younger than 60 years, for whom intensive induction is still standard of care, with an age-matched, sex-matched and study date-matched cohort of 206 white patients (size-matched 1:2) with similar performance status treated on the CALGB/Alliance protocols. There were no significant differences in pretreatment parameters between Black and white patients, including white blood cell count, percentage of bone marrow (BM) and blood blasts, performance status or 2022 ELN genetic risk group distribution (Supplementary Table 8). Both cohorts were similarly treated on the same frontline CALGB/Alliance clinical trials, with all patients receiving standard induction chemotherapy followed by consolidation with cytarabine or autologous hematopoietic stem cell transplantation (HSCT), except for three patients who received consolidation with mitoxantrone/diaziquone and cyclophosphamide/etoposide (Supplementary Information and Supplementary Table 8).

Although the complete remission (CR) rates of Black and white patients did not differ significantly (73% versus 79%, $P = 0.25$; Fig. 4a), Black patients more frequently died within the 30 days after treatment initiation (11% versus 3%, $P = 0.02$), had higher relapse rates (63% versus 48%, $P = 0.05$) and worse disease-free survival (DFS) (3-year rates, 30% versus 47%, $P = 0.01$; Supplementary Fig. 4), event-free survival (EFS) (3-year rates, 22% versus 37%, $P = 0.003$; Fig. 4a) and overall survival (OS) (3-year rates, 31% versus 47%, $P = 0.007$; Fig. 4b) compared with white patients.

Molecular features associated with response and survival

To identify the features associated with the treatment response and survival of Black patients, we performed multivariable analyses for DFS, which identified *NPM1* (hazard ratio (HR) = 2.29, 95% confidence interval (CI) = 1.23–4.25, $P = 0.009$) and *NRAS* (HR = 1.95, 95% CI = 1.03–3.69, $P = 0.04$) mutations as adverse prognostic factors (Supplementary Tables 9 and 10 and Fig. 4c). *NPM1* mutations without *FLT3*-ITD are a favorable-risk marker in clinical guidelines[7], contrasting their strong adverse prognostic impact in Black patients. In contrast to established risk association, in which the presence of *FLT3*-ITD negatively affects the positive survival association with *NPM1* mutations, Black patients with *NPM1* mutations who were *FLT3*-ITD$^-$ had worse DFS (3-year rates: 13% versus 55%, $P = 0.002$), EFS (3-year rates: 10% versus 46%, $P = 0.01$) and OS (10% versus 61%, $P < 0.001$) than white patients, and their OS was not significantly different from the OS of patients in the 2022 ELN adverse-risk group ($P = 0.88$; Fig. 4d). Notably, although definitive conclusions are limited by a small sample size, *NPM1* mutations were also associated with inferior OS in Black patients younger than 60 years treated with intensive chemotherapy in the analysis of real-world data from the Flatiron Health database (Extended Data Fig. 7). *NRAS* mutations, which are not included in the 2022 ELN genetic risk stratification system[7], also conferred adverse prognosis in Black patients (Fig. 4e and Supplementary Table 11). Furthermore, myelodysplasia-related mutations were predictive of inferior DFS, which is consistent with their established negative survival association in current clinical guidelines (Supplementary Table 9 and Extended Data Fig. 8)[7].

Mutations in *IDH1* and *IDH2* were associated with decreased OS ($P = 0.05$; HR = 1.73, 95% CI = 1.01–2.97; Supplementary Table 12), which was again specific for Black patients with *IDH1* and *IDH2* mutations (Supplementary Table 13 and Fig. 4f). Given the approval of new targeted agents for older and unfit patients harboring *IDH1* and *IDH2* variants, the survival association in Black patients suggests the possibility of using these agents in the frontline setting at an earlier age. No recurrent gene mutation stayed in the final multivariable model for EFS (Supplementary Table 14).

https://doi.org/10.1038/s41588-024-01929-x

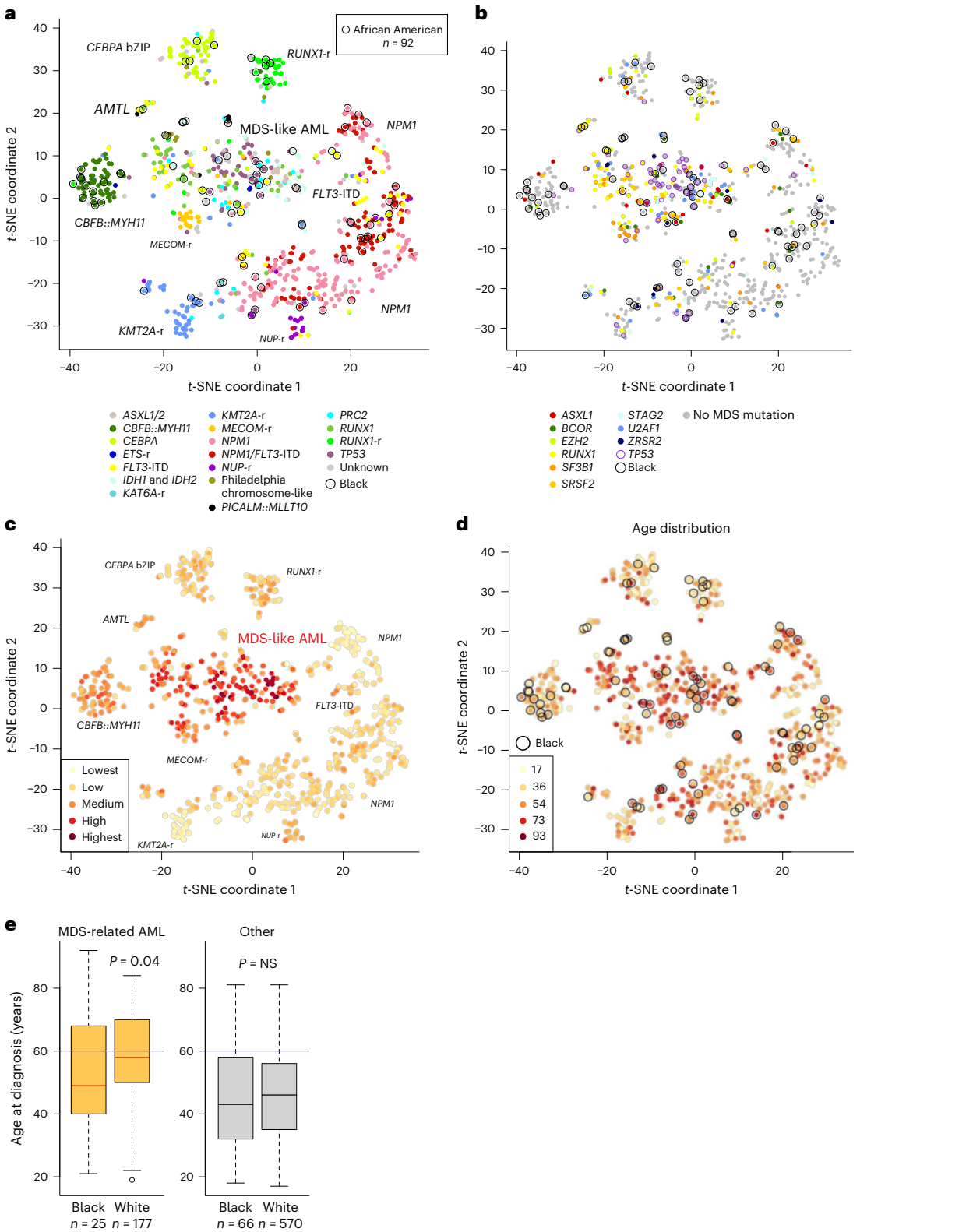
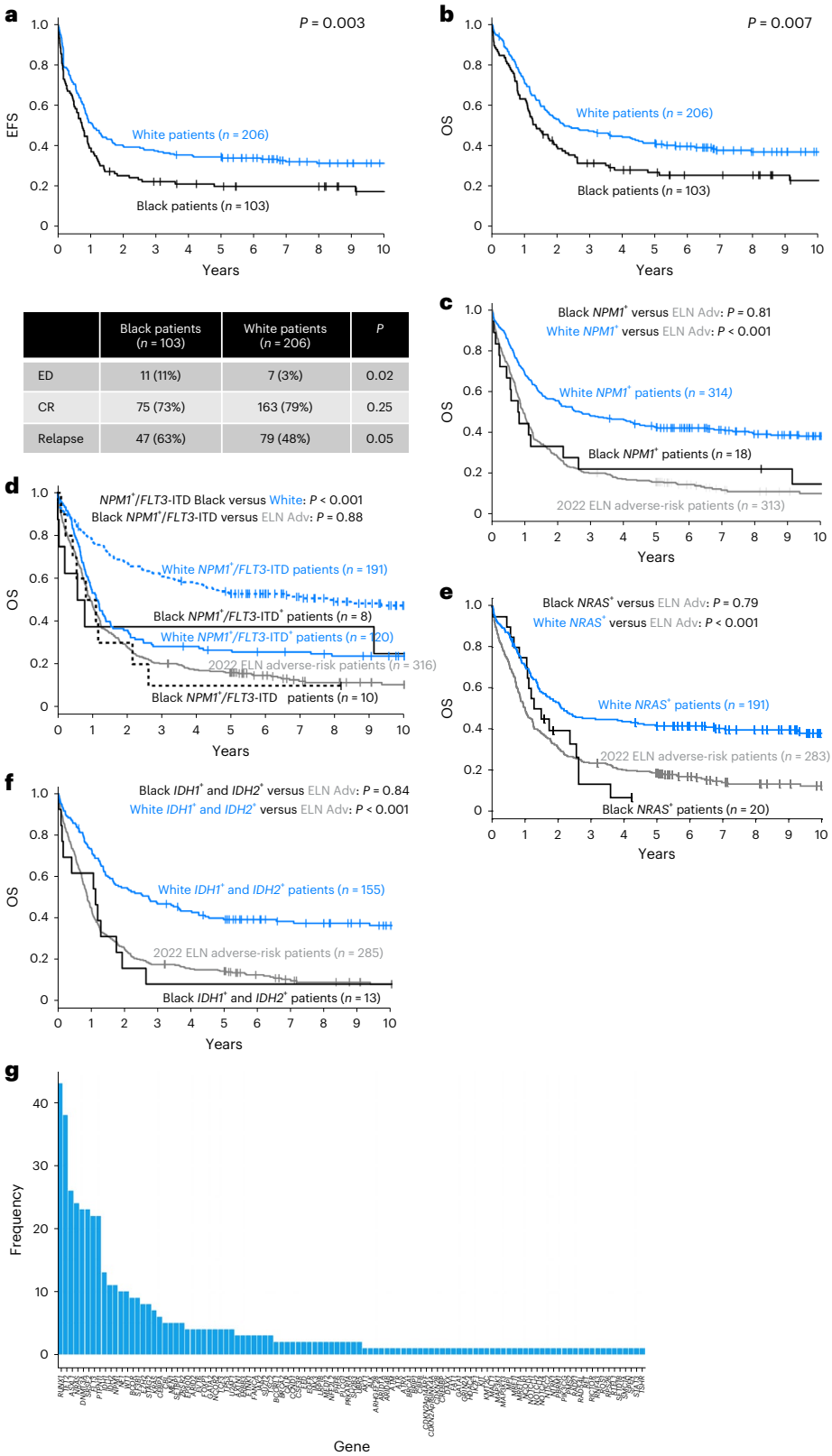


Fig. 3 | Clustering of patients with AML using transcriptome-based gene expression profiling and *t*-SNE visualization. a, Clustering of Black and white patients according to gene expression with the presence of major oncogenic driver mutations. The dot color corresponds to the presence of the indicated driver mutation. Black patients are circled. **b**, Presence of myelodysplasia-related mutations in Black and white patients clustered according to gene expression. The dot color corresponds to the presence of the indicated myelodysplasia-related mutation. **c**, Similarity of Black and white patient gene expression profiles to a previously published myelodysplasia gene expression signature[29]. **d**, Age distribution of patients clustered according to gene expression. The dot colors correspond to patient age as indicated in the legend. Black patients are circled. **e**, Age comparison between Black and white patients with myelodysplasia-related mutations. *P* values were calculated using a two-sided Wilcoxon rank-sum test (box plots: centerline, median; box limits, first and third quartiles; whiskers, minimum and maximum). '-r' behind a gene symbol indicates fusion genes involving the gene indicated and other partner gene(s). *CEBPA* bZIP, in-frame mutations affecting the basic leucine zipper (bZIP) region of the *CEBPA* gene; MDS, myelodysplastic syndrome; NS, not significant.

https://doi.org/10.1038/s41588-024-01929-x



	Black patients (*n* = 103)	White patients (*n* = 206)	*P*
ED	11 (11%)	7 (3%)	0.02
CR	75 (73%)	163 (79%)	0.25
Relapse	47 (63%)	79 (48%)	0.05

Fig. 4 | Comparison of clinical outcomes of Black and white patients with AML. a,**b**, Comparison of treatment response and survival of age-matched, sex-matched and study date-matched cohorts of Black and white patients, **a**, EFS, rates of early death (ED), CR and relapse. **b**, OS. **c**, OS of Black and white patients with *NPM1* mutations compared with the OS of 2022 ELN adverse-risk patients. **d**, OS of Black and white patients with *NPM1* mutations and the presence or absence of a co-occurring *FLT3*-ITD in comparison with OS of 2022 ELN adverse-risk patients. **e**, OS of Black and white patients with *NRAS* mutations compared with the OS of 2022 ELN adverse-risk patients. **f**, OS of Black and white patients with *IDH1* and *IDH2* mutations compared with the OS of 2022 ELN adverse-risk patients. **g**, Frequencies of gene mutations detected in 43 relapsed or refractory adult Black patients cared for at the Memorial Sloan Kettering Comprehensive Cancer Center, profiled using an MSK-IMPACT assay. In **a**, for ED, CR and relapse rates, *P* values were calculated using a two-sided Fisher's exact test. **a**–**f**, for time-to-event analyses, survival estimates were calculated using the Kaplan–Meier method and compared using a two-sided log-rank test. Adv, adverse.

https://doi.org/10.1038/s41588-024-01929-x

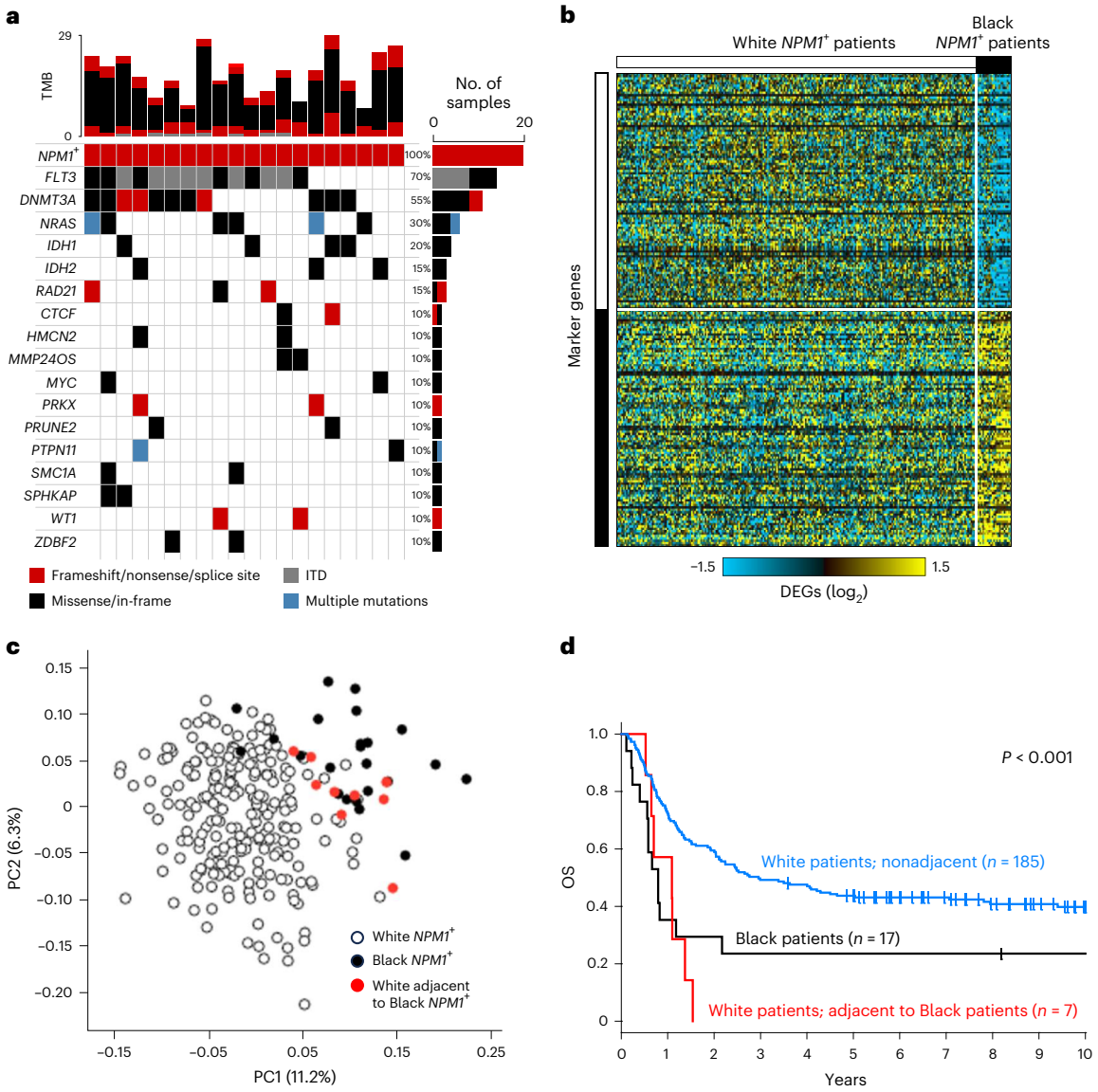


Fig. 5 | Ancestry-associated differences in transcriptional profiles. **a**, Mutational oncoprint of Black patients with *NPM1* mutations. The red color denotes frameshift, nonsense and splice site mutations; black denotes missense and in-frame mutations; gray denotes ITDs of the *FLT3* gene; and blue represents multiple mutations. **b**, Marker heatmap (MarkerFinder algorithm) of ancestry-associated gene expression differences in bulk RNA-seq data derived from patients with *NPM1*-mutated AML with genotype-confirmed ancestry ($n = 260$ patients). **c**, Visualization of the top two principal components of the *NPM1* gene expression variation. White patients of European ancestry who clustered together with Black patients of African ancestry ('white adjacent to Black', $n = 9$), based on supervised classification, are denoted separately in red. **d**, OS of patients with *NPM1* mutations aged younger than 60 years with respect to their ancestry and separate depiction of the identified white patients adjacent to Black patients ($n = 7$). Survival estimates were calculated using the Kaplan–Meier method and compared using a two-sided log-rank test.

We next analyzed the mutation frequencies in 43 self-reported Black patients with relapsed or refractory disease whose gene mutation profiles were obtained after unsuccessful first-line treatment (median age: 60 years; range: 18–86 years) via MSK-IMPACT with subsequent enrichment of more adverse-risk features, contrasting the genomic profiling of the newly diagnosed Alliance patient cohort. The most frequent mutations included several adverse-risk markers identified in our study, supporting their risk association in Black patients (Fig. 4g).

Ancestry-associated transcriptomic impact in *NPM1*-mutated AML

To understand the possible causes for the contrasting survival impact of *NPM1* mutations in Black compared to white patients, we first assessed co-mutational patterns. In addition to established coexisting mutational features, 11 of 20 Black patients with *NPM1* mutations harbored at least one predicted pathogenic variant in genes not yet implicated in AML (Fig. 5a); however, no new recurrent variants were identified.

Next, to evaluate possible molecular underpinnings, we analyzed the transcriptomes of Black and white patients with *NPM1* mutations with genotype-confirmed ancestry. Interestingly, using the principal component and reclassification analyses, we found distinguishing marker genes that were differentially expressed between Black and white patients (Fig. 5b). Only one Black patient's AML resembled in its expression the phenotype of white patients, while nine white patients resembled the expression profile found in Black patients (Fig. 5c). Remarkably, the OS of these white patients who presented as transcriptomically 'adjacent' to Black patients was also poor (Fig. 5d).

https://doi.org/10.1038/s41588-024-01929-x

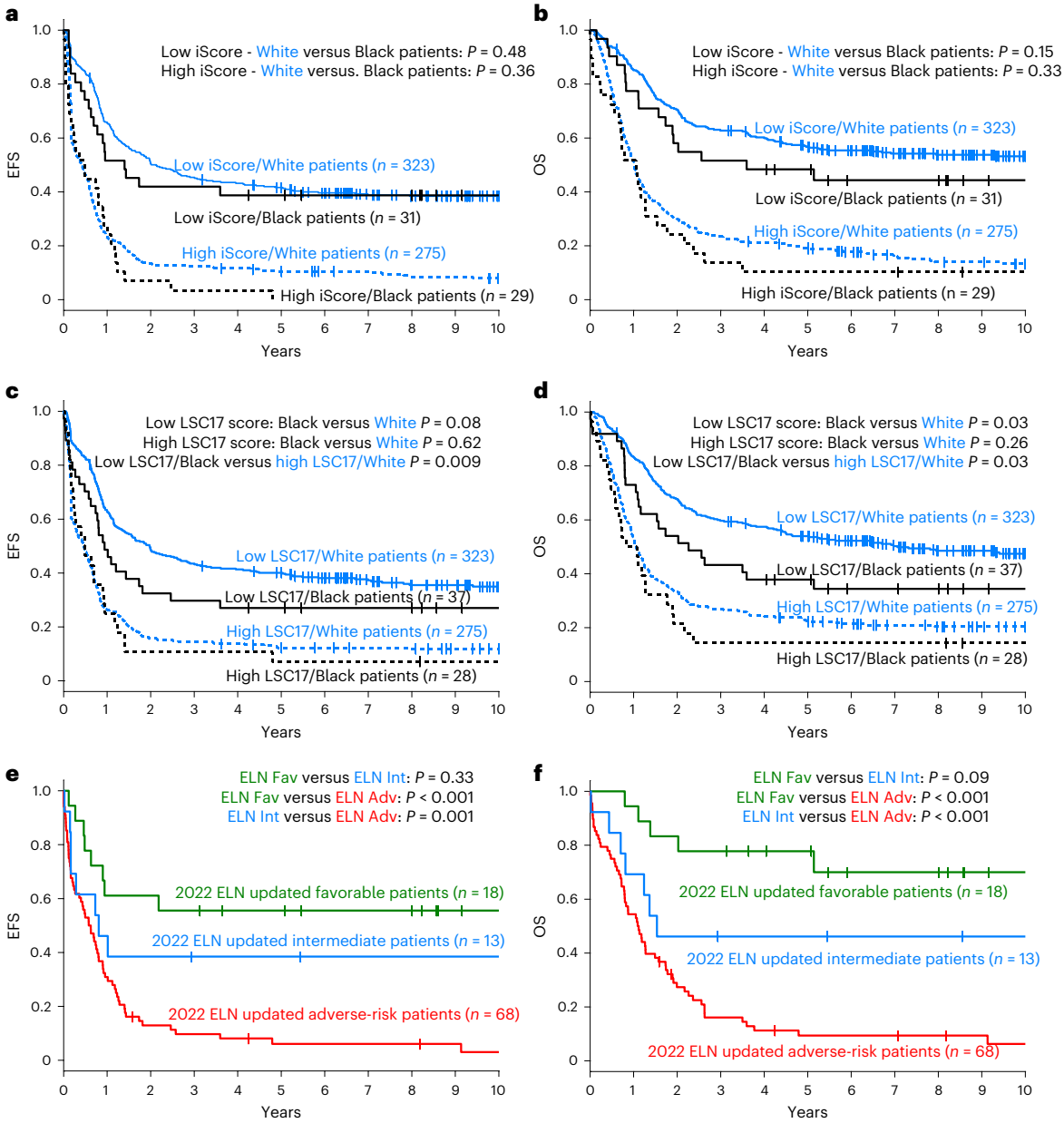


Fig. 6 | Treatment outcomes of Black and white patients classified according to three genetic risk stratification systems. a, EFS of Black and white patients according to high and low iScore status. **b**, OS of Black and white patients according to high and low iScore status. **c**, EFS of Black and white patients with low or high LSC17 scores. **d**, OS of Black and white patients with low or high LSC17 scores. **e**,**f**, EFS (**e**) and OS (**f**) of Black patients categorized into genetic risk groups according to our modification of the 2022 ELN genetic risk classification by the inclusion of *NPM1*, *NRAS* and *IDH1* and *IDH2* mutations as adverse-risk markers. Survival estimates were calculated using the Kaplan–Meier method and compared using a two-sided log-rank test. Adv, adverse; Fav, favorable; Int, intermediate.

Treatment outcomes according to iScore, LSC17 and 2022 ELN

Inflammation has a role in the development and progression of AML; high inflammation has emerged as a negative prognostic factor independent of established genetic features[21,30]. Using the recently published gene-expression-based iScore[30] as a proxy for inflammation, we noted an enrichment of high inflammation in Black patients harboring *NPM1* mutations compared to white patients (69% versus 45%; $P = 0.08$; Supplementary Table 15). This is notable given the previously described association of high inflammation with poor-risk genetic features and treatment response, and the identified inferior survival of Black patients with *NPM1* mutations[19,21]. High inflammation was predictive of poor EFS and OS for both Black and white patients ($P < 0.001$; Fig. 6a,b).

The LSC17 score represents a validated survival-associated gene expression signature thought to be reflective of 'stemness' in AML with high score[31,32]. LSC17 scores were predictive of EFS and OS in both white and Black patients (Fig. 6c,d). Only 53% of Black 2022 ELN favorable-risk patients had a low LSC17 score, compared with 69% of white 2022 ELN favorable-risk patients. Notably, the incidence of *NPM1* mutations was discrepant: only 13% of low LSC17 Black patients had *NPM1* mutations compared with 38% of white patients, suggesting possible biological differences between Black and white patients that require further study. Similarly, only 16% of Black patients in the low LSC17 group carried *NRAS* mutations versus 26% of white patients, which is consistent with the observed contrasting survival impact of *NRAS* mutations (Supplementary Table 16).

Refinement of the 2022 ELN genetic risk classification

The 2022 ELN genetic risk classification performed suboptimally for Black patients, especially with respect to identifying patients who benefit from cytotoxic chemotherapy alone. Compared with white patients, Black patients tended to have inferior EFS and OS in each risk group, with the difference reaching statistical significance for EFS in the 2022 ELN favorable-risk group (Extended Data Fig. 9). The unfavorable impact of *NRAS*, *IDH1*, *IDH2* and *NPM1* mutations on OS and DFS in our Black patients suggest that these mutations should be included as adverse-risk markers for Black patients. Indeed, implementing this change improved genetic risk assignment, with 3-year EFS and OS predicted rates for the 2022 ELN favorable-risk group improving from 37% to 56% and from 50% to 78%, and for the intermediate-risk patients from 28% to 38% and from 35% to 43%, respectively (Fig. 6e,f). This revised, ancestry-inclusive genetic risk assignment resulted in reclassification of 34% of Black patients to the adverse-risk group. This has important clinical ramifications because these patients would be considered for alternative frontline regimens or allogeneic HSCT in first CR.

Cell type distribution and pathway activation of *NPM1*-mutated AML

To elucidate the potential underlying biological mechanisms responsible for the observed phenotype differences, we performed single-cell RNA-seq (scRNA-seq) profiling with cellular indexing of transcriptomes and epitopes (CITE-seq) on BM samples from six Black and seven white patients with *NPM1* mutations, obtaining a total of 112,017 cells (Fig. 7a). To understand possible lineage relationships, we aligned cells from these patients to our recent comprehensive multimodal BM progenitor cell atlas[33]. Among 74 aligned cell populations, the most frequently aligned cell state was to the annotated MultiLin-GMP-1 cluster (34% of cells on average), which usually is the most infrequently found transitional intermediate state in nonleukemic BM. Interestingly, when comparing cell type proportions with respect to ancestry, lymphoid myeloid primed progenitor (LMPP-2) cells showed consistent differences between Black and white patients ($P = 0.04$; Fig. 7b). While the leukemic cell state is probably an important mediator of response to therapy, transcriptomic impacts differentiated according to ancestry could reveal unique AML survival gene programs. To exclude bias in cell capture frequency, we compared gene expression among the reference-annotated pseudobulk cell populations using the software cellHarmony. cellHarmony revealed cell state-specific and lineage impacts when comparing the scRNA-seq profiles of Black and white patients that nominated broadly shared AML blast ancestry programs and cell state-specific impacts (Fig. 7c and Extended Data Fig. 10).

A rare intermediate mixed-lineage progenitor, megakaryocyte–erythroid progenitor (MEP-2) cell population showed the greatest gene expression changes according to ancestry. Notably, we also observed consistent downregulation of *NPM1* in the MEP-2 cells of Black compared to white patients ($P = 0.008$), but not in any other cell populations. As identified using gene set enrichment analysis, Black patients had lower expression of splicing factors, and transcription, translation and mitochondrial energy pathway, which was consistent among patients (Fig. 7d). These pathways and other were substantially enriched when considering shared progenitor lineage impacts from cellHarmony (MPP and MEP combined), including apoptosis, Wnt signaling, EGFR signaling and stress granule production (*AGO4*, *GIGYF2*, *MBNL1*, *CIRBP*, *DDX3X*, *HNRNPK*, *PABPC1*, *YBX1*, *ZFP36*) (Fig. 7e). These data point to contributing posttranscriptional regulatory mechanisms driving ancestry-associated survival of the leukemic clone.

Discussion

Current clinical NGS panels for AML-associated mutations and genetic risk stratification assume that these clinical standards perform equally for all patients. However, our results suggest that these panels are not delivering adequate results for Black patients. This conclusion is highlighted by the 10% of 'mutation-negative' Black patients in the routinely used testing panels, and the identification of recurrent gene mutations in Black patients, which were not recurrent in previous AML hallmark studies[3,6]. These findings are consistent with studies on other cancers, which found that genetic ancestry contributes to patterns of somatic mutations[34–37] and suggest the need to expand current clinical NGS panels to include ancestry-associated mutations. However, their addition to routine clinical testing of AML is at present premature; further studies are necessary to gather information on the biological and clinical relevance of these variants.

The presence of an early-onset myelodysplastic phenotype in Black patients associated with poor survival suggests the presence of intrinsic and/or extrinsic stressors that cause these changes at an unusually young age. Racial discrimination and safety-related insecurities have been linked to inflammation, which is known to influence mutagenesis and promote the development of myeloid neoplasms, highlighting the importance of considering ethnicity and ancestry in the assessment of disparities as they provide additive information[38–40]. Notably, this observation is consistent with epidemiological data on prostate and breast cancer showing earlier disease onset in Black patients than white patients[41,42]. The mechanisms driving the early myelodysplastic phenotype are unknown, but lung cancer tumors from patients with African ancestry have higher levels of genomic instability and ancestry-associated defects in homologous recombination repair[37].

Genetic risk stratification has a central role in therapeutic decision-making for patients with AML. Our findings indicate that the current risk stratification system does not capture important ancestry-associated biological differences with treatment implications. As all patients received cytarabine/daunorubicin-based therapy, we were able to perform rigorous outcome analyses. We identified contrasting survival predictors between Black and white patients, indicating an inferior genetic risk assignment for almost one-third of Black patients based on current classification systems.

Specifically, the presence of *NPM1* mutations confers an adverse risk in Black patients, while current clinical guidelines consider *NPM1* mutations to bestow favorable or intermediate risk depending on *FLT3*-ITD status[7], suggesting a potential undertreatment of Black patients with *NPM1* mutations. Notably, our data provide evidence of underlying biological differences in patients with *NPM1* mutations with respect to genetic ancestry, including differences in cell type distribution, with an increase of LMPP-2 cells in Black patients with confirmed African ancestry. Remarkably, this distinct phenotype identified in all but one Black patient was identified in 3.8% of white patients with genotype-confirmed European ancestry, who also had poor survival.

We believe our results indicate the need to modify the current AML risk stratification by inclusion of ancestry-specific molecular prognostic factors, and to prospectively test their validity in randomized clinical trials. Furthermore, they indicate the urgent need to understand the functional significance of the hitherto unrecognized putative pathogenic variants and their relevance in leukemogenesis using murine models. Future genomic studies of patients with cancer from ancestry-diverse populations should include ancestry-weighted models and a more refined assessment of ancestry that may uncover additional differences within different African patient populations. Such efforts are necessary for the equitable distribution of molecular medicine and to address long-standing disparities in survival of patients with AML.

Importantly, our results call for large-scale follow-up efforts. While adults of all ages were included in the analyses of mutation frequencies in Black and white patients, the outcome analyses were limited to fit patients aged younger than 60 years, for whom intensive induction is still standard of care. None of these patients received an allogeneic HSCT in first CR, which represents a limitation of our study and thus

https://doi.org/10.1038/s41588-024-01929-x

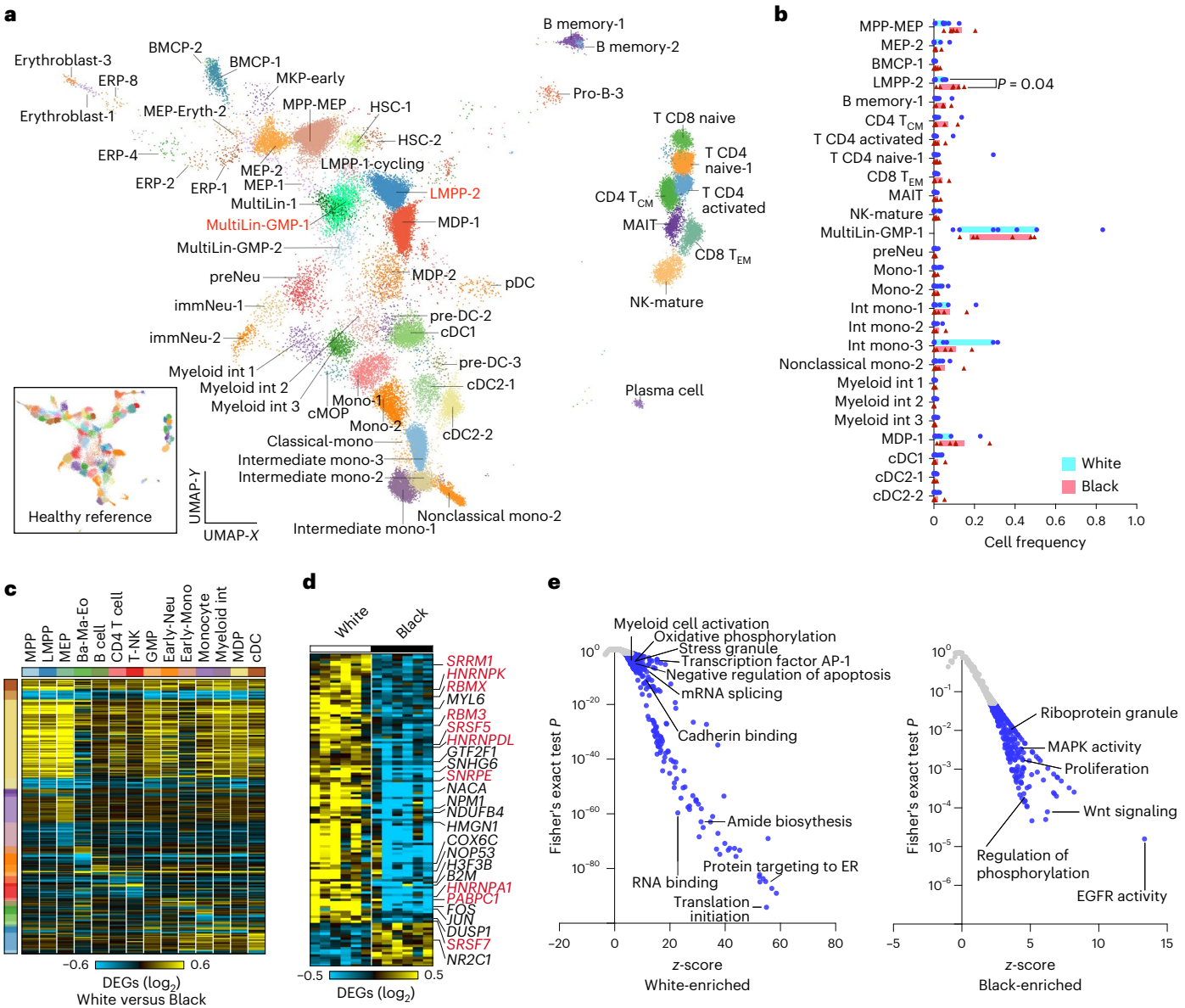


Fig. 7 | Identification of a durable clonal gene expression program in patients with *NPM1*-mutated AML based on ancestry. a, Uniform manifold approximation and projection (UMAP) of scRNA-seq profiles from 13 patients with *NPM1*-mutated AML (six with African ancestry, seven with European ancestry), and to a nonleukemic reference BM cell populations (inset). **b**, Cell population frequency for each patient with AML with respect to genetic ancestry (two-sided uncorrected Welch *t*-test). **c**, Differential expression heatmap of pseudobulk clusters (cellHarmony), comparing patients with European and African ancestry (fold change greater than 1.2, eBayes *t*-test $P < 0.01$). **d**, Heatmap showing differentially expressed genes (DEGs) in MEP-2 according to ancestry derived from cellHarmony. Red, splicing regulator. **e**, Gene set enrichment of European-enriched (left) and African-enriched (right) differentially expressed Gene Ontology terms in MPP/MEP combined pseudobulk RNA-seq data (raw *P* values from a two-sided Fisher's exact test). BMCP, basophil/mast cell progenitor; cDC, conventional dendritic cell; DC, dendritic cell; EGFR, epidermal growth factor receptor; ER, endoplasmic reticulum; eryth, erythroid; GMP, granulocyte-monocyte progenitor; LMPP, lympho-myeloid primed progenitor; MAIT, mucosal associated invariant T; MAPK, mitogen-activated protein kinase; MDP, monocyte-dendritic progenitor; mono, monocyte; MKP, megakaryocyte progenitor; MPP, multipotent progenitor; Neu, neutrophil; NK, natural killer; pDC, plasmacytoid dendritic cell; T_{CM}, central memory T; T_{EM}, effector memory T.

warrants performing similar analyses in the transplant setting. Finally, given the retrospective nature of this work, it is unclear if improvements in risk stratification for Black patients would translate into improved outcomes. For this, future prospective clinical trials are needed.

We hope that this large-scale study focused on patients with African ancestry with AML sets a precedent for future genomic profiling efforts, as the current underrepresentation of minority patients not only constrains our ability to provide the best possible care, but also limits our understanding of AML biology. Future studies should occur in conjunction with increased efforts to address major contributors to the long-standing survival disparities reflective of structural racism, including clinical trial enrollment, access to care and social determinants of health.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-024-01929-x.

References

1. Döhner, H., Weisdorf, D. J. & Bloomfield, C. D. Acute myeloid leukemia. *N. Engl. J. Med.* **373**, 1136–1152 (2015).
2. Duncavage, E. J. et al. Genome sequencing as an alternative to cytogenetic analysis in myeloid cancers. *N. Engl. J. Med.* **384**, 924–935 (2021).
3. Ley, T. J. et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).
4. Papaemmanuil, E. et al. Genomic classification and prognosis in acute myeloid leukemia. *N. Engl. J. Med.* **374**, 2209–2221 (2016).
5. Mardis, E. R. et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361**, 1058–1066 (2009).
6. Tyner, J. W. et al. Functional genomic landscape of acute myeloid leukemia. *Nature* **562**, 526–531 (2018).
7. 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).
8. Khoury, J. D. et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. *Leukemia* **36**, 1703–1719 (2022).
9. Davis, M. B. Genomics and cancer disparities: the justice and power of inclusion. *Cancer Discov.* **11**, 805–809 (2021).
10. Spratt, D. E. et al. Racial/ethnic disparities in genomic sequencing. *JAMA Oncol.* **2**, 1070–1074 (2016).
11. Tan, D. S. W., Mok, T. S. K. & Rebbeck, T. R. Cancer genomics: diversity and disparity across ethnicity and geography. *J. Clin. Oncol.* **34**, 91–101 (2016).
12. Hantel, A. et al. Racial and ethnic enrollment disparities and demographic reporting requirements in acute leukemia clinical trials. *Blood Adv.* **5**, 4352–4360 (2021).
13. Hantel, A. et al. Inequities in Alliance acute leukemia clinical trial and biobank participation: defining targets for intervention. *J. Clin. Oncol.* **40**, 3709–3718 (2022).
14. Abraham, I. E. et al. Structural racism is a mediator of disparities in acute myeloid leukemia outcomes. *Blood* **139**, 2212–2226 (2022).
15. Ho, G. et al. Early mortality and complications in hospitalized adult Californians with acute myeloid leukaemia. *Br. J. Haematol.* **177**, 791–799 (2017).
16. Jabo, B., Morgan, J. W., Martinez, M. E., Ghamsary, M. & Wieduwilt, M. J. Sociodemographic disparities in chemotherapy and hematopoietic cell transplantation utilization among adult acute lymphoblastic and acute myeloid leukemia patients. *PLoS ONE* **12**, e0174760 (2017).
17. Master, S., Munker, R., Shi, Z., Mills, G. & Shi, R. Insurance status and other non-biological factors predict outcomes in acute myelogenous leukemia: analysis of data from the National Cancer Database. *Anticancer Res.* **36**, 4915–4921 (2016).
18. Borate, U. M., Mineishi, S. & Costa, L. J. Nonbiological factors affecting survival in younger patients with acute myeloid leukemia. *Cancer* **121**, 3877–3884 (2015).
19. Bhatnagar, B. et al. Poor survival and differential impact of genetic features of Black patients with acute myeloid leukemia. *Cancer Discov.* **11**, 626–637 (2021).
20. Larkin, K. T. et al. High early death rates, treatment resistance, and short survival of Black adolescents and young adults with AML. *Blood Adv.* **6**, 5570–5581 (2022).
21. Lasry, A. et al. An inflammatory state remodels the immune microenvironment and improves risk stratification in acute myeloid leukemia. *Nat. Cancer* **4**, 27–42 (2023).
22. Miles, L. A. et al. Single-cell mutation analysis of clonal evolution in myeloid malignancies. *Nature* **587**, 477–482 (2020).
23. Baldus, C. D., Mrózek, K., Marcucci, G. & Bloomfield, C. D. Clinical outcome of *de novo* acute myeloid leukaemia patients with normal cytogenetics is affected by molecular genetic alterations: a concise review. *Br. J. Haematol.* **137**, 387–400 (2007).
24. Dietrich, J., Lovell, S., Veatch, O. J. & Butler, M. G. *PHIP* gene variants with protein modeling, interactions, and clinical phenotypes. *Am. J. Med. Genet. A* **188**, 579–589 (2022).
25. Taylor, J. et al. Single-cell genomics reveals the genetic and molecular bases for escape from mutational epistasis in myeloid neoplasms. *Blood* **136**, 1477–1486 (2020).
26. Weinstock, J. S. et al. Aberrant activation of TCL1A promotes stem cell expansion in clonal haematopoiesis. *Nature* **616**, 755–763 (2023).
27. LaHaye, S. et al. Discovery of clinically relevant fusions in pediatric cancer. *BMC Genomics* **22**, 872 (2021).
28. Martins da Silva, V. et al. Genetic abnormalities in large to giant congenital nevi: beyond *NRAS* mutations. *J. Invest. Dermatol.* **139**, 900–908 (2019).
29. Im, H. et al. Distinct transcriptomic and exomic abnormalities within myelodysplastic syndrome marrow cells. *Leuk. Lymphoma* **59**, 2952–2962 (2018).
30. Trowbridge, J. J. & Starczynowski, D. T. Innate immune pathways and inflammation in hematopoietic aging, clonal hematopoiesis, and MDS. *J. Exp. Med.* **218**, e20201544 (2021).
31. Ng, S. W. K. et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* **540**, 433–437 (2016).
32. Bill, M. et al. Mutations associated with a 17-gene leukemia stem cell score and the score's prognostic relevance in the context of the European LeukemiaNet classification of acute myeloid leukemia. *Haematologica* **105**, 721–729 (2020).
33. Zhang, X. et al. An immunophenotype-coupled transcriptomic atlas of human hematopoietic progenitors. *Nat. Immunol.* **25**, 703–715 (2024).
34. Carrot-Zhang, J. et al. Genetic ancestry contributes to somatic mutations in lung cancers from admixed Latin American populations. *Cancer Discov.* **11**, 591–598 (2021).
35. Guda, K. et al. Novel recurrently mutated genes in African American colon cancers. *Proc. Natl Acad. Sci. USA* **112**, 1149–1154 (2015).
36. Gomez, F., Griffith, M. & Griffith, O. L. Genetic ancestry correlations with driver mutations suggest complex interactions between somatic and germline variation in cancer. *Cancer Discov.* **11**, 534–536 (2021).
37. Sinha, S. et al. Higher prevalence of homologous recombination deficiency in tumors from African Americans versus European Americans. *Nat. Cancer* **1**, 112–121 (2020).
38. Cuevas, A. G. et al. Discrimination and systemic inflammation: a critical review and synthesis. *Brain Behav. Immun.* **89**, 465–479 (2020).
39. Ferguson, L. R. Chronic inflammation and mutagenesis. *Mutat. Res.* **690**, 3–11 (2010).
40. Stubbins, R. J., Platzbecker, U. & Karsan, A. Inflammation and myeloid malignancy: quenching the flame. *Blood* **140**, 1067–1074 (2022).
41. Lillard, J. W. Jr., Moses, K. A., Mahal, B. A. & George, D. J. Racial disparities in Black men with prostate cancer: a literature review. *Cancer* **128**, 3787–3795 (2022).
42. Stringer-Reasor, E. M., Elkhanany, A., Khoury, K., Simon, M. A. & Newman, L. A. Disparities in breast cancer associated with African American identity. *Am. Soc. Clin. Oncol. Educ. Book* **41**, e29–e46 (2021).

Andrew Stiff[1,31], Maarten Fornerod[2,31], Bailee N. Kain[3,31], Deedra Nicolet[1,4,5], Benjamin J. Kelly[6], Katherine E. Miller[6,7], Krzysztof Mrózek[1,5], Isaiah Boateng[1], Audrey Bollas[6], Elizabeth A. R. Garfinkle[6], Omolegho Momoh[8], Foluke A. Fasola[9], Hannah O. Olawumi[10], Nuria Mencia-Trinchant[11], Jean F. Kloppers[12,13], Anne-Cecilia van Marle[12,13], Eileen Hu[1,14], Saranga Wijeratne[6], Gregory Wheeler[6], Christopher J. Walker[1], Jill Buss[1,5], Adrienne Heyrosa[1,5], Helee Desai[1,5], Andrea Laganson[1,5], Ethan Hamp[1,5], Yazan Abu-Shihab[1,5], Hasan Abaza[1,5], Parker Kronen[1], Sidharth Sen[15], Megan E. Johnstone[16], Kate Quinn[16], Ben Wronowski[16], Erin Hertlein[16], Linde A. Miles[17], Alice S. Mims[1,18], Christopher C. Oakes[1,18], James S. Blachly[1,18], Karilyn T. Larkin[1,18], Bethany Mundy-Bosse[1,18], Andrew J. Carroll[19], Bayard L. Powell[20], Jonathan E. Kolitz[21], Richard M. Stone[22], Cassandra Duarte[23], Diana Abbott[23], Maria L. Amaya[23], Craig T. Jordan[23], Geoffrey L. Uy[24], Wendy Stock[25], Kellie J. Archer[26], Electra D. Paskett[1,27], Monica L. Guzman[11,28], Ross L. Levine[29], Kamal Menghrajani[29], Debyani Chakravarty[29], Michael F. Berger[29], Daniel Bottomly[30], Shannon K. McWeeney[30], Jeffrey W. Tyner[30], John C. Byrd[16], Nathan Salomonis[3], H. Leighton Grimes[3], Elaine R. Mardis[6,7,32] ✉ & Ann-Kathrin Eisfeld[1,4,18,32] ✉

[1]The Ohio State University Comprehensive Cancer Center, Columbus, OH, USA. [2]Department of Cell Biology, Erasmus Medical Center, Rotterdam, the Netherlands. [3]Division of Immunobiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [4]Alliance Statistics and Data Management Center, The Ohio State University Comprehensive Cancer Center, Columbus, OH, USA. [5]Clara D. Bloomfield Center for Leukemia Outcomes Research, The Ohio State University Comprehensive Cancer Center, Columbus, OH, USA. [6]The Steve and Cindy Rasmussen Institute for Genomic Medicine, Abigail Wexner Research Institute at Nationwide Children's Hospital, Columbus, OH, USA. [7]Department of Pediatrics, The Ohio State University College of Medicine, Columbus, OH, USA. [8]Department of Internal Medicine, Montefiore Health System/Albert Einstein College of Medicine, New York, NY, USA. [9]Department of Hematology, Faculty of Basic Medical Science, University of Ibadan, University College Hospital, Ibadan, Nigeria. [10]Department of Haematology, University of Ilorin, Ilorin, Nigeria. [11]Division of Hematology/Oncology, Department of Medicine, Weill Cornell Medicine, New York, NY, USA. [12]School of Pathology, Faculty of Health Sciences, University of the Free State, Bloemfontein, South Africa. [13]National Health Laboratory Service, Universitas Academic Business Unit, Bloemfontein, South Africa. [14]Department of Internal Medicine, University of Texas Southwestern, Dallas, TX, USA. [15]Divison of Biomedical Informatics, Cincinnati Children's Hospital, Cincinnati, OH, USA. [16]Department of Internal Medicine, University of Cincinnati, Cincinnati, OH, USA. [17]Cancer and Blood Diseases Institute, Cincinnati Children's Hospital, Cincinnati, OH, USA. [18]Division of Hematology, Department of Internal Medicine, The Ohio State University Comprehensive Cancer Center, Columbus, OH, USA. [19]Department of Genetics, University of Alabama at Birmingham, Birmingham, AL, USA. [20]Atrium Health Wake Forest Baptist Comprehensive Cancer Center, Winston-Salem, NC, USA. [21]Monter Cancer Center, Hofstra Northwell School of Medicine, Lake Success, NY, USA. [22]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. [23]University of Colorado, Anschutz Medical Campus, College of Medicine, Denver, CO, USA. [24]Washington University School of Medicine, Saint Louis, MO, USA. [25]University of Chicago, College of Medicine, Chicago, IL, USA. [26]Division of Biostatistics, College of Public Health, The Ohio State University, Columbus, OH, USA. [27]Division of Cancer Prevention and Control, Department of Internal Medicine, College of Medicine, The Ohio State University, Columbus, OH, USA. [28]Sandra and Edward Meyer Cancer Center, New York, NY, USA. [29]Memorial Sloan-Kettering Cancer Center, New York, NY, USA. [30]Knight Cancer Institute, Oregon Health & Science University, Portland, OR, USA. [31]These authors contributed equally: Andrew Stiff, Maarten Fornerod, Bailee N. Kain. [32]These authors jointly supervised this work: Elaine R. Mardis, Ann-Kathrin Eisfeld. ✉e-mail: Elaine.Mardis@nationwidechildrens.org; Ann-Kathrin.Eisfeld@osumc.edu

Methods

Our study was undertaken to analyze differences in genomic data between Black and white patients with AML and correlate the results with clinical outcome data to shed light on the origins of survival disparity between these diverse genetic ancestry patient groups. We used the population descriptors Black and white to refer to groups of individuals who have different constitutional genetic backgrounds, typically resulting in observable phenotypic differences. For most cases, the classification of Black or white was determined by SNP array genotyping, but for a subset of patients we relied on self-reported race alone (fully described below). These descriptors were operationalized by using self-reported race to select white and Black patients from the described cohorts of either available samples or available omics data, then validating the genetic ancestry associated with the self-reported race for each patient in the subset of patients for whom existing genotyping data were available, as described below. We note that not having genetic ancestry data on a fraction of patients is a limitation of this study. We acknowledge that additional genetic ancestry groups exist other than white and Black but justify the focus of the current study on only these two groups because they were the largest available genetic ancestry groups in the assessed cohorts; further comparisons are outside the scope of this work. This work complies with all relevant ethical regulations and was approved by the Institutional Review Board (IRB) of the Ohio State University and the Cancer Therapy Evaluation Program of the National Cancer Institute (NCI). The University of Ibadan/University College Hospital (UI/UCH) Research Ethics Committee, which is the equivalent of an IRB, approved the study (no. UI/EC/23/0295). All patients provided written informed consent to participate in the treatment studies before enrollment and for the research use of their specimens in agreement with the 2013 version of the Declaration of Helsinki. Patients provided written informed consent to participate in protocols CALGB 8461 (cytogenetic studies), CALGB 9665 (Leukemia Tissue Bank) and CALGB 20202 (molecular studies), which involved collection of pretreatment BM aspirates and blood samples. No patient was compensated for participation in this study. Study protocols were approved by the IRB at each center.

Patients and treatment (CALGB/Alliance)

In the CALGB/Alliance database, there were 1,660 adults diagnosed with de novo AML (other than acute promyelocytic leukemia) who had molecular genetic data available, including 1,519 self-reported white and 141 self-reported Black patients, of whom 1,149 and 100, respectively, had their genetic ancestry confirmed by genotyping (see below and Extended Data Fig. 5). All but three patients were similarly treated with intensive chemotherapy consisting of cytarabine/daunorubicin-based induction and cytarabine-based consolidation or received autologous HSCT after attaining a CR on the CALGB/Alliance protocols, the details of which are provided in the Supplementary Information. The aforementioned three patients received consolidation with mitoxantrone/diaziquone and cyclophosphamide/etoposide. Patients removed from protocol treatment to undergo allogeneic HSCT in first CR were excluded from the survival analysis. Two patients were treated at the Ohio State University and received cytarabine/daunorubicin-based induction (in one patient together with gemtuzumab ozogamicin) followed by high-dose cytarabine.

Genetic ancestry analyses

Genetic ancestry was inferred for 1,149 of 1,519 self-reported white patients, who had genotyping data available based on 650K Omni arrays using the method in refs. 43,44. For all 1,149 patients, genetic ancestry was confirmed as white European. Additionally, a more refined African ancestry was confirmed using WES-based SNP analysis as detailed below. For more refined genetic ancestry for the self-reported Black patients of the WES cohort, we used the SNVstory workflow[45] (https://github.com/nch-igm/snvstory). SNVstory is a method built on three independent machine learning models for accurately inferring the continental and subcontinental ancestry of individuals. Germline variants from VCF were used as input to the workflow. Genetic ancestry assignment was determined as the output model label with the highest probability; ancestry assignments from each of the three continental models were supplied together. For study inclusion, patients needed to have 40% African ancestry (Genome Aggregation Database).

Gene mutation frequency comparisons versus Beat AML

Mutations and frequencies were generated from the publicly available calls for BeatAML (https://github.com/biodev/beataml2.0_data/blob/main/beataml_wes_wv1to4_mutations_dbgap.txt) from ref. 46, which are a combination of Mutect and VarScan2: (1) the overall cohort consists of 762 samples with 531 of these being from the initial diagnosis samples; (2) we defined the white cohort as having self-reported white race without self-reported Hispanic ethnicity ($n = 323$). In all cases, samples were first limited to those referenced in ref. 46, where each patient was represented by a single sample. All genes found to be recurrently mutated in Black patients were reanalyzed using the analytical workflow and variant curation as outlined for the Black patient cohort.

The genetic ancestry for each patient, inferred from genomic data using the method in ref. 43 was assessed for 177 (55%) of the 323 patients, as previously described by the Beat AML investigators[46], which included all patients of their initial sequencing effort; patients sequenced on the second sequencing effort did not have sufficient data available for the ancestry analyses. There were only two white patients with a discordance between their self-reported and inferred genetic ancestry, as previously reported[46].

PHIP mutation profiling in the African patient cohorts

We collaborate with two investigative sites in Nigeria, the UI/UCH and University of Ilorin Teaching Hospital (UITH). The BM aspirate slides were from samples archived between June 1985 and December 2022. As they were archived samples, all patients were deceased at the time of the analysis. UI/UCH and UITH provided the archived BM aspirate slides and correlated clinical data where available. DNA was extracted centrally using the QIAGEN QIAamp DNA Micro extraction kit. Illumina TruSeq-derived, targeted sequencing data for 46 patients (29 males, 17 females; aged 4–86 years) passed the quality control criteria and are included. Library preparations and subsequent sequencing were performed centrally at the Bloomfield Center at the Ohio State University.

In addition, we partnered with investigators at the University of the Free State in Bloemfontein, South Africa. All research was approved under UFS-HSD2022/0299. Tumor DNA was analyzed centrally at the Bloomfield Center for 20 patients with AML (18 self-reported with African ancestry, two self-reported with mixed ancestry; eight males, 12 females; aged 18–82 years), as detailed above.

Integrated molecular profiling

One hundred Black patients with AML, selected to represent major morphological and cytogenetic subtypes of AML (Table 1), underwent integrated mutational and transcriptional profiling of their pretreatment samples, which was also performed in relapse samples from 18 of these patients. African ancestry for Black patients was confirmed via genotype analyses of germline variants. This included DNA-based tumor-normal comparison by Integrated DNA Technologies xGen Lockdown probe-based whole-exome hybrid capture with chromosomal tiling probes and NGS (250× target coverage of paired-end 151-bp reads) of genomic DNA from leukemic and flow-sorted B and T cells as germline tissue. Secondary analysis was performed using Churchill, which includes alignment to the GRCh38 reference genome, deduplication, germline and somatic SNVs and small insertion and deletion calling, as well as copy number and loss of heterozygosity detection and clonal predictions[47]. Post-variant calling analysis identified AML-associated and hitherto unreported somatic and germline alterations, using a variant allele fraction

cutoff of 2% or more. A gene was considered recurrently mutated if nonsynonymous, protein sequence-altering variants were detected in two or more patients. Ribo-depleted, paired-end 151-bp RNA-seq (60–80 million reads per sample) was performed on RNA extracted from diagnosis and relapse materials, followed by identification of gene fusions and ITDs[27], and gene expression profiling (see Supplementary Information for details of the referenced methods). CH was defined by simultaneous detection of variants in both B and T cell fraction DNA and DNA from leukemic blasts in the absence of other somatic mutations in B and T cell fractions and after filtering known polymorphisms.

scRNA-seq and CITE-seq

scRNA-Seq and CITE-seq was conducted using the 10X genomics platform on samples from randomly selected Black and white patients with *NPM1* mutations, who were not significantly different from the entire Alliance cohorts of Black and white patients with respect to age, sex and OS (see the Supplementary Information for additional details).

Genomic landscape of relapsed or refractory AML among Black patients

Blood and BM samples from 43 relapsed or refractory patients were obtained and sequenced using MSK-IMPACT Heme, a capture-based, NGS assay. This platform targets 400 genes that are known to be involved in the pathobiology of hematological neoplasms. Tumor-specific variants were identified in comparison to a patient-specific or pooled set of normal variants, as described previously[48].

Statistics and reproducibility

Statistical analyses were performed using SAS v.9.4, R v.4 and TIBCO Spotfire S+ 8.2. Alliance data were locked on 31 July 2022. No statistical method was used to predetermine sample size as sample size was limited by the availability of the material. For the two patients who were hypermutated, nonrecurrent mutations were excluded from the frequency comparison analyses. The experiments were not randomized. The investigators were blinded to allocation during the experiments.

Outcome analysis of patients younger than 60 years

Among 141 Black patients in the CALGB/Alliance database, 103 patients were younger than 60 years; among 1,519 white patients, 952 were younger than 60 years. For select survival comparisons with 103 Black patients, a cohort of 206 white patients was matched with regard to age, sex and study date. Clinical and biological characteristics were compared using Fisher's exact and Wilcoxon rank-sum tests for categorical and continuous variables, respectively. For CR, we calculated *P* values using a Fisher's exact test. For the time-to-event analyses, we calculated survival estimates using the Kaplan–Meier method and compared groups using a log-rank test. A limited backward selection technique was used to build the final multivariable models within the 103 Black patients younger than 60 years for the achievement of CR, DFS, OS and EFS. We used logistic regression to model CR and Cox proportional hazards regression to model DFS, OS and EFS for the univariable and multivariable outcome analyses. In our outcome analyses, we used *P* values adjusted to control for the family error rate (probability of a type I error) for all variables considered in the univariable analyses. The families were all variables considered in each outcome analysis and only variables whose likelihood ratio test-adjusted $P < 0.20$ from the univariable models were considered in the multivariable analyses. To identify variables associated with the achievement of CR, DFS, EFS and OS, the following parameters were included in the univariable outcome analyses: age, sex, hemoglobin, platelets, white blood cell count, percentage of blood and BM blasts, extramedullary involvement, CN-AML, CBF-AML, AML with complex karyotype, *FLT3*-ITD and *FLT3*-TKD status, and mutational status of the genes defining myelodysplasia-related AML (that is, *SRSF2*, *SF3B1*, *RUNX1*, *U2AF1*, *ZRSR2*, *ASXL1*, *EZH2*, *BCOR* and *STAG2*) and *DNMT3A*, *IDH1*, *IDH2*, *NPM1* and *NRAS*.

Comparative survival analyses from the Flatiron dataset

The nationwide Flatiron Health electronic health record-derived database is a longitudinal database comprising de-identified patient-level structured and unstructured data, curated using technology-enabled abstraction. During the study period, the de-identified data originated from approximately 280 US cancer clinics (~800 sites of care). Most patients in the database originate from community oncology settings; relative community and academic proportions may vary depending on the study cohort. The data cutoff date was 30 June 2022. The OS was calculated from the date of diagnosis to the last documented follow-up or data cutoff or death. The Flatiron Health dataset consists of 2,245 patients with AML who had *NPM1* molecular testing performed, of whom 140 patients were younger than 60 years and were treated with anthracycline-based induction chemotherapy (seven Black, 133 white).

Definition of the clinical endpoints

For the Alliance data, ED was defined as a patient's death within 30 days of starting therapy, irrespective of the cause. CR required an absolute neutrophil count of 1.5×10^9 per liter or greater, a platelet count greater than 100×10^9 per liter, no leukemic blasts in the blood, cellularity greater than 20% with maturation of all cell lines, no Auer rods, less than 5% BM blast cells and no evidence of extramedullary leukemia, all of which had persisted for at least 1 month. Relapse was defined by 5% or more BM blasts, circulating leukemic blasts or the development of extramedullary leukemia[49]. DFS was measured from the date of CR until the date of relapse or death; patients alive and relapse-free at the last follow-up were censored. EFS was measured from the date of study entry until the date of failure to achieve CR, relapse or death. Patients alive and in CR at the last follow-up were censored. OS was measured from the date on study until the date of death; patients alive at the last follow-up were censored.

Reporting summary

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

Data availability

The RNA-seq data from de novo patients with AML were in part derived from our previously published datasets that are publicly available via the Gene Expression Omnibus under accession nos. GSE137851, GSE63646and GSE216738. Newly generated data are available under accession nos. GSE266099and GSE266498. Genetic variants and raw files are available in the database of Genotypes and Phenotypes (https://www.ncbi.nlm.nih.gov/gap/) under accession no. phs003728.v1.

Code availability

The AltAnalyze v.2.1.4 graphical user interface was used as described for the cellHarmony, principal component and differential expression analyses (https://www.altanalyze.org). Similarly, the graphical user interface of Prism v.10 (GraphPad Software) was used for cell frequency plotting. Accessory analysis scripts for cellHarmony, SoupX, count-scaling and UMAP supervised mapping (approximateUMAP) are provided at https://www.synapse.org/Synapse:syn53222724.

References

43. Zheng, X. & Weir, B. S. Eigenanalysis of SNP data with an identity by descent interpretation. *Theor. Popul. Biol.* **107**, 65–76 (2016).
44. Walker, C. J. et al. Genome-wide association study identifies an acute myeloid leukemia susceptibility locus near *BICRA*. *Leukemia* **33**, 771–775 (2019).
45. Bollas, A. E. et al. SNVstory: A dockerized algorithm for rapid and accurate inference of sub-continental ancestry. Preprint at *bioRxiv* https://doi.org/10.1101/2023.06.02.543369 (2023).

46. Bottomly, D. et al. Integrative analysis of drug response and clinical outcome in acute myeloid leukemia. *Cancer Cell* **40**, 850–864 (2022).
47. Kelly, B. J. et al. Churchill: an ultra-fast, deterministic, highly scalable and balanced parallelization strategy for the discovery of human genetic variation in clinical and population-scale genomics. *Genome Biol.* **16**, 6 (2015).
48. Ptashkin, R. N. et al. Enhanced clinical assessment of hematologic malignancies through routine paired tumor and normal sequencing. *Nat. Commun.* **14**, 6895 (2023).
49. Cheson, B. D. et al. Report of the National Cancer Institute-sponsored workshop on definitions of diagnosis and response in acute myeloid leukemia. *J. Clin. Oncol.* **8**, 813–819 (1990).

Acknowledgements

We thank the patients who consented to participate in these clinical trials and the families who supported them; C. Manring and the CALGB/Alliance Leukemia Tissue Bank at the Ohio State University Comprehensive Cancer Center for sample processing and storage services; and L. J. Sterling for data management. The research reported in this publication was supported by the resources of the Pelotonia Institute for Immuno-Oncology, an allocation of computing resources from the Ohio Supercomputer Center and Shared Resources (Leukemia Tissue Bank). The research reported in this publication was supported in part by the National Cancer Institute (NCI) of the National Institutes of Health (NIH) under award nos. U10CA180821, U10CA180882 and U24CA196171 (to the Alliance for Clinical Trials in Oncology); award nos. UG1CA283338, UG1CA189824, UG1CA233338, U10CA140158, UG1CA233331, U10CA180867, R35CA197734, UG1CA189850 and 5P30CA016058; award nos. R01 CA262496 (to A.-K.E., E.R.M. and A.S.M.), R01CA284595-01, R01CA283574-01, R01 LM013879 and the Coleman Leukemia Research Foundation; the Leukemia & Lymphoma Society (to A.-K.E.); the American Cancer Society (A.-K.E.); the Leukemia Research Foundation (A.-K.E.); an ASCO Young Investigator Award (to A.S.); and an ASH MRI Award (to O.M.). It was jointly sponsored by a Ruth L. Kirschstein National Research Service Award Institutional Predoctoral Training Program in the Neurosciences (R5 T32 CA 117846-15 to B.N.K.) and the D. Warren Brown Foundation. The support provided to the Alliance for Clinical Trials in Oncology and Alliance Foundation Trials programs is listed at https://acknowledgments.alliancefound.org. R.L.L. was supported by an NCI award no. R35CA197594. Studies supported by the MSK core facilities were supported in part by an MSKCC Support Grant/Core Grant no. P30 CA008748 and the Marie-Josée and Henry R. Kravis Center for Molecular Oncology. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Trial registrations: NCT00048958, NCT00899223, NCT00900224.

Author contributions

A.-K.E., E.R.M., A.S., M.F. and D.N. contributed to the study design. A.-K.E., E.R.M., A.S., M.F., D.N., K. Mrózek, C.C.O., K.J.A., E.D.P. and H.L.G. contributed to data interpretation. A.S., B.J.K., I.B., E. Hu, S.W., G.W., A.H., H.D., A.L., E. Hamp, Y.A.-S., H.A., P.K., M.E.J., K.Q., B.W., L.A.M. and B.M.-B. performed the laboratory-based research. M.F., B.N.K., K.E.M., A.B., E.A.R.G., N.M.-T., C.J.W., S.S., J.S.B., D.B., S.K.M., J.W.T. and N.S. performed the genomic analysis. K. Menghrajani and D.C. performed the genomic analysis of relapsed and refractory MSK-IMPACT patients. M.F., D.N., B.J.K., C.J.W. and D.A. performed the statistical analysis. O.M., F.A.F., H.O.O., J.F.K., A.-C.v.M., K. Mrózek, J.B., E. Hertlein, A.S.M., K.T.L., A.J.C., B.L.P., J.E.K., R.M.S., C.D., M.L.A., C.T.J., G.L.U., W.S., M.L.G., M.F.B. and J.C.B. were involved directly or indirectly in the care of patients or sample procurement. A.-K.E., A.S., K. Mrózek, D.N., R.L.L., J.W.T. and E.R.M. wrote the manuscript. All authors read the manuscript and agreed on the final version.

Competing interests

C.J.W. is a consultant for Vigeo Therapeutics and is employed by Karyopharm Therapeutics; he has ownership interest in Karyopharm Therapeutics and Bristol Myers Squibb Co. A.S.M. has served in a consulting or advisory role for AbbVie, Bristol Meyers Squibb, Novartis and Treadwell Therapeutics; he has served on a data monitoring safety committee for Daiichi Sankyo and Foghorn Therapeutics and currently serves as a Senior Medical Director for the Leukemia and Lymphoma Society Beat AML Study. J.S.B. is a consultant and advisory board member for AbbVie, AstraZeneca, Syndax, INNATE and KITE. B.L.P. has received honoraria from Jazz Pharmaceuticals, Novartis and Pfizer. J.F.K. has received honoraria from Gilead, Magellan and Novartis; consulting fees from Gilead, Magellan, Novartis, Pharmacyclics and Seattle Genetics; institutional research funding from Boehringer Ingelheim, Cantex, Erytech and Millennium; and travel support from Gilead, Novartis and Seattle Genetics. R.M.S. has served on independent data safety monitoring committees for trials supported by Celgene, Takeda and Argenx; has consulted for AbbVie, Actinium, Agios, Amgen, Arog, Astellas, AstraZeneca, BioLineRx, Celgene, Daiichi Sankyo, Fujifilm, Janssen, Juno, Macrogenics, Novartis, Ono, Orsenix, Pfizer, Roche, Stemline Therapeutics, Sumitomo, Takeda and Trovagene; and has received research support (to his institution) for clinical trials sponsored by AbbVie, Agios, Arog and Novartis. E.D.P. has received grants from the Merck Foundation and Pfizer. M.L.G. holds research contracts and consults with Bridge Medicines and holds stock options for SeqRx. M.F.B. has a consultancy role with AstraZeneca, Eli Lilly and Paige.AI and has received research support from Boundless Bio. J.C.B. has a consultancy and advisory role with Syndax, Novartis and Vincera; research funding from Pharmacyclics, an AbbVie company, Genentech, Janssen and Acerta; and ownership for Vincera. E.R.M. has received research support (to her institution) from Pfizer, the Merck Foundation, Genentech, Guardant Health and Astra Zeneca; and has served as an advisory board member for GSK and Merck, and as member of the Board of Directors of the Alliance Foundation and American Association for Cancer Research. R.L.L. is on the supervisory board of QIAGEN and is a scientific adviser to Imago, Mission Bio, Syndax, Zentalis Pharmaceuticals, Ajax, Bakx, Auron, Prelude, C4 Therapeutics and Isoplexis, for which he receives equity support. R.L.L. receives research support from Ajax and Abbvie and has consulted for Incyte, Janssen, MorphoSys and Novartis. He has received honoraria from AstraZeneca and Incyte for invited lectures. A.-K.E. has received a research grant from Novartis, and an honorarium from AstraZeneca for serving on their Diversity, Equity and Inclusion Advisory Board; her spouse has ownership interest in Karyopharm Therapeutics. The other authors declare no completing interests.

Additional information

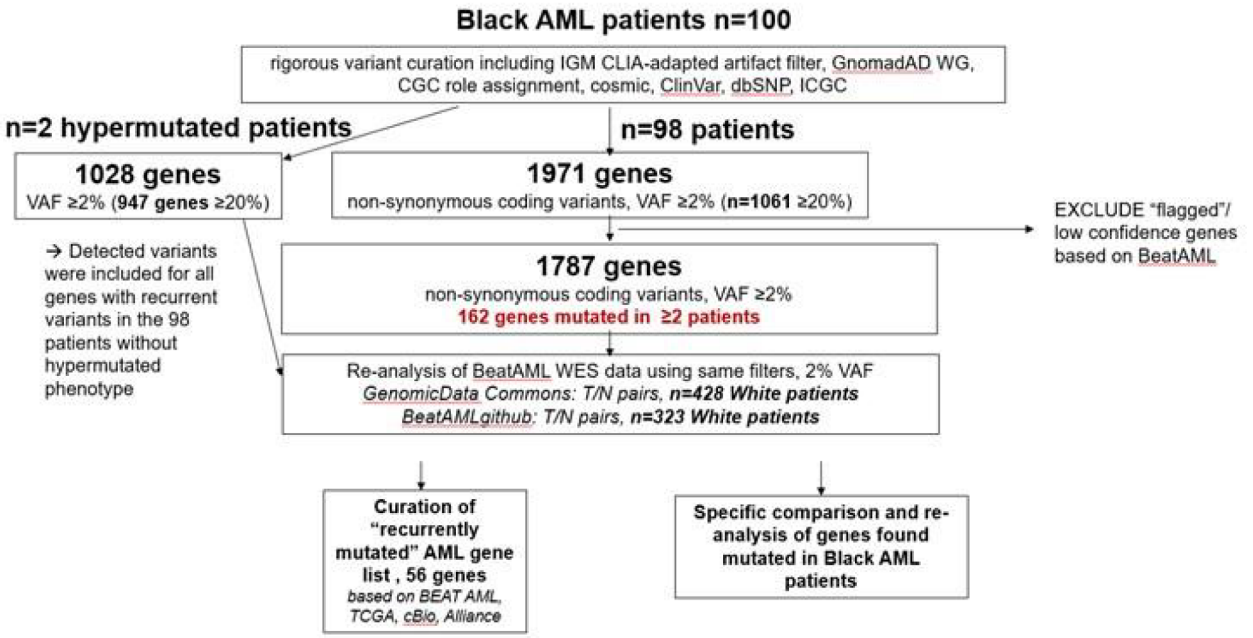
Extended data is available for this paper at https://doi.org/10.1038/s41588-024-01929-x.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-024-01929-x.

Correspondence and requests for materials should be addressed to Elaine R. Mardis or Ann-Kathrin Eisfeld.

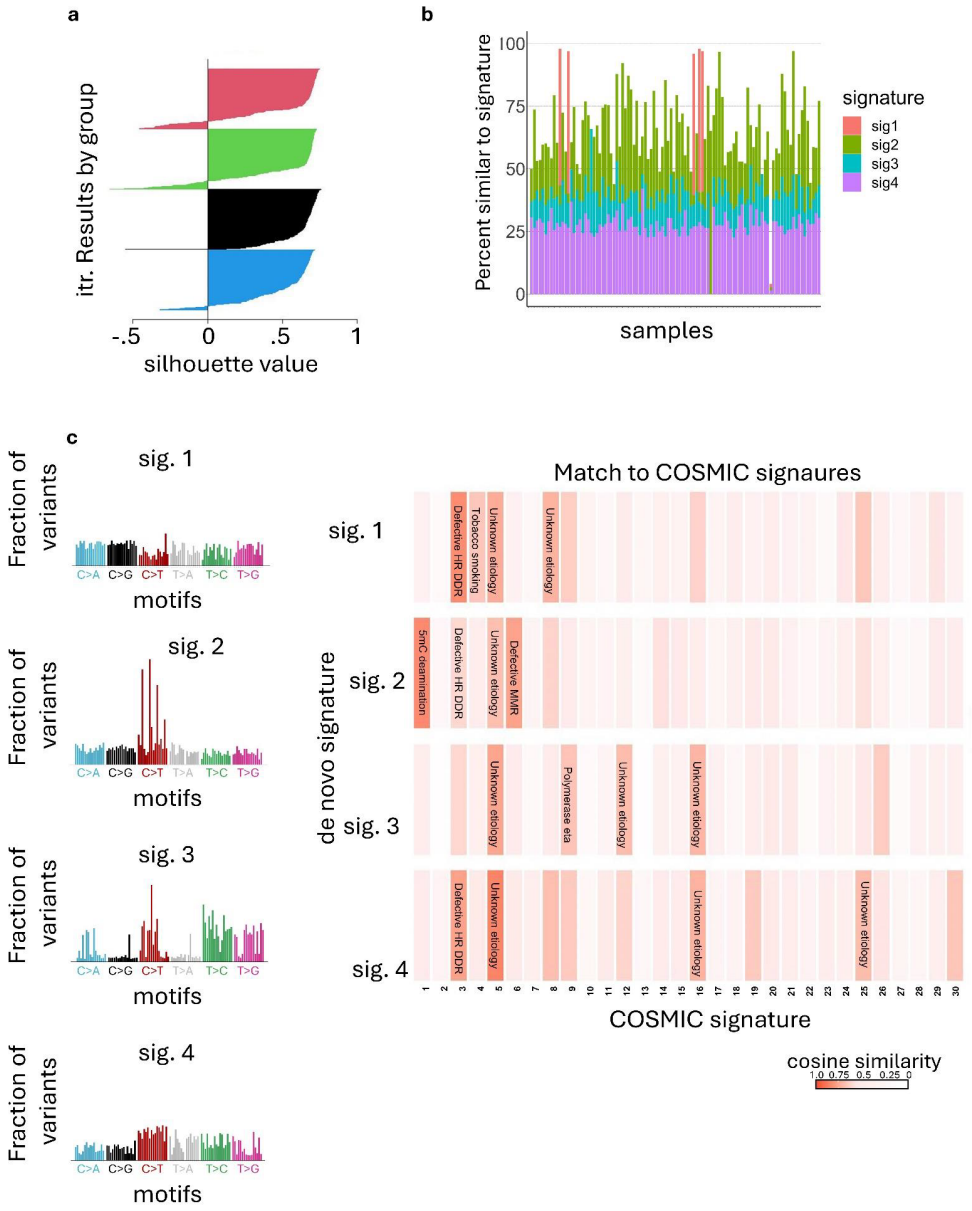
Peer review information *Nature Genetics* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

https://doi.org/10.1038/s41588-024-01929-x



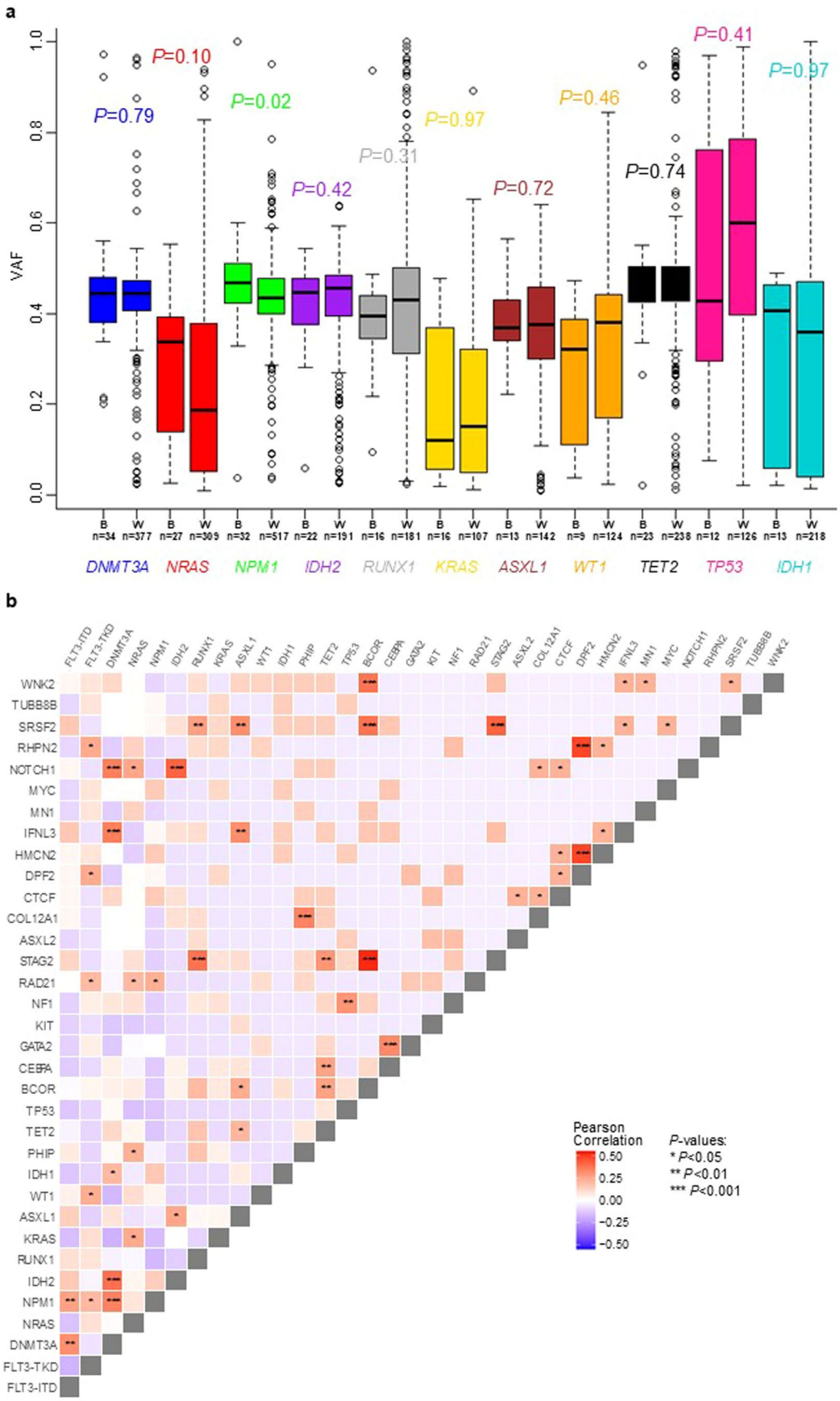
Extended Data Fig. 1 | Variant calling, filtering and analyses workflow for determination of variants identified via paired tumor/normal whole-exome sequencing (WES) of 100 Black patients with AML. Comparisons with White patients were done based on paired tumor-normal WES data of self-identified White patients who were sequenced as part of the BeatAML study and database. To ensure comparability, BeatAML WES data were re-analyzed with identical variant calling workflow and variant allele fraction threshold of ≥2%. Genes recurrently mutated in AML were curated via consensus of major AML databases.

https://doi.org/10.1038/s41588-024-01929-x



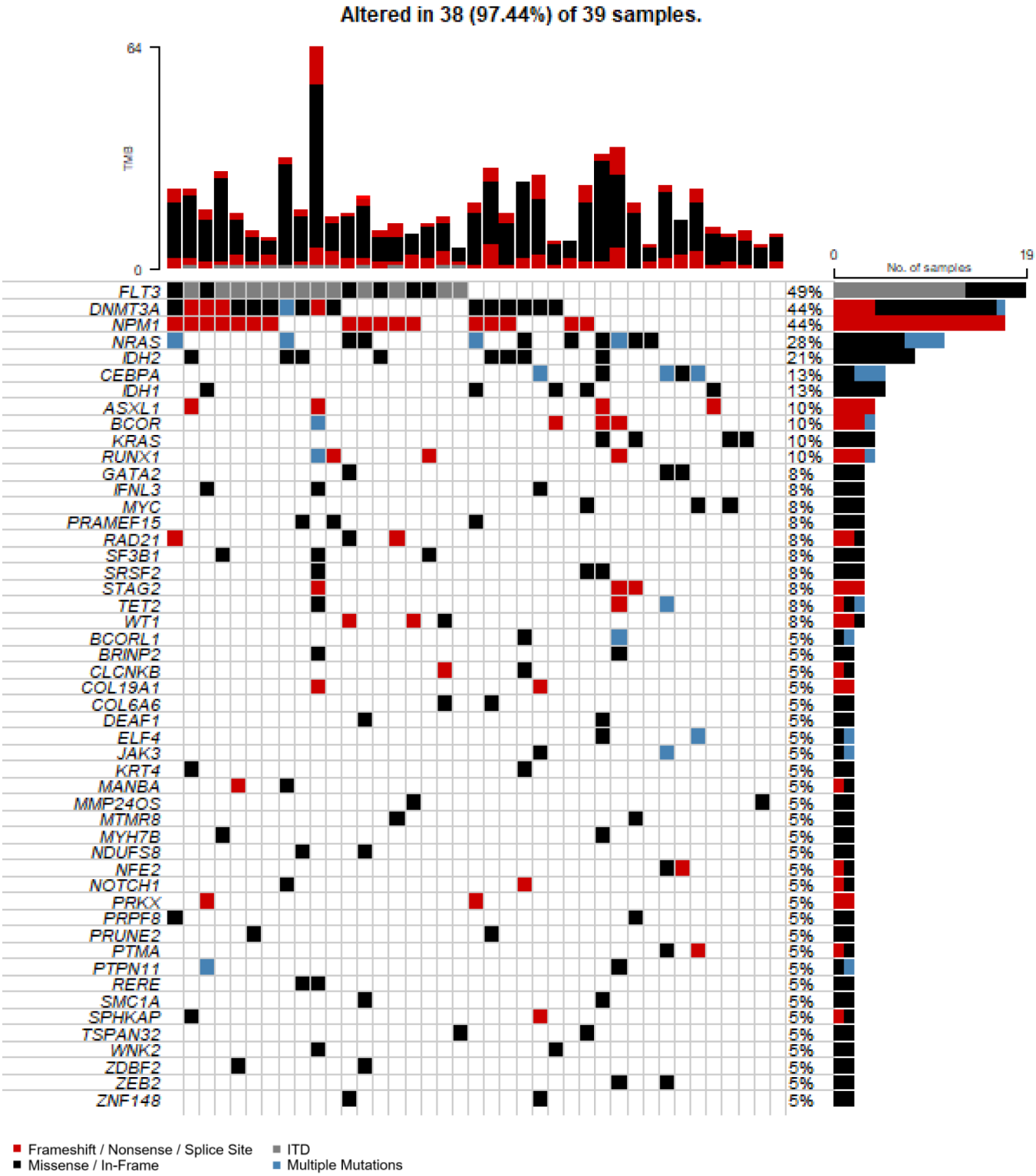
Extended Data Fig. 2 | Mutational signatures identified in Black patients with AML. a, Identification of 4 signatures with optimal performance. **b**, Presence of 4 signatures at the patient level. **c**, Similarity with COSMIC signatures. On the left are the identified mutational signatures, the color-coded bars on the left indicate similarity to known COSMIC mutational signatures. The etiologies of these signatures, if known, are shown for the 4 most similar signatures.

https://doi.org/10.1038/s41588-024-01929-x



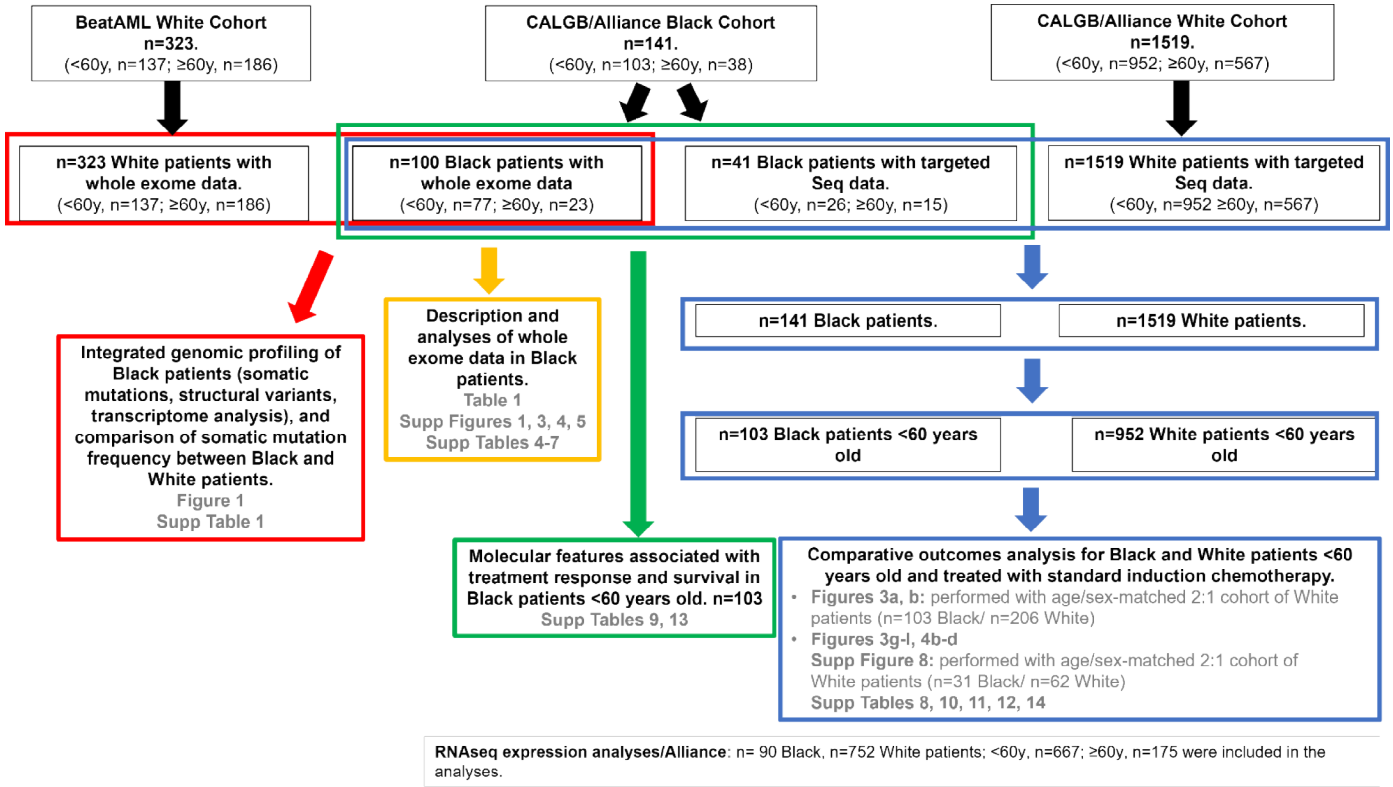
Extended Data Fig. 3 | Driver mutations and mutational co-occurence patters found in Black patients with AML. a, Box plots depicting variant allele fractions of frequent driver mutations, comparing Black and White patients treated on Alliance protocols, *P*-values were calculated using a two-sided Wilcoxon rank sum test (boxplots: centerline, median; box limits, first and third quartiles; whisker, 1.5x interquartile range). **b**, Triangle plot depicting mutational co-occurrence patterns found in Black patients with AML. Included are all genes mutated in at least 4 patients.

https://doi.org/10.1038/s41588-024-01929-x



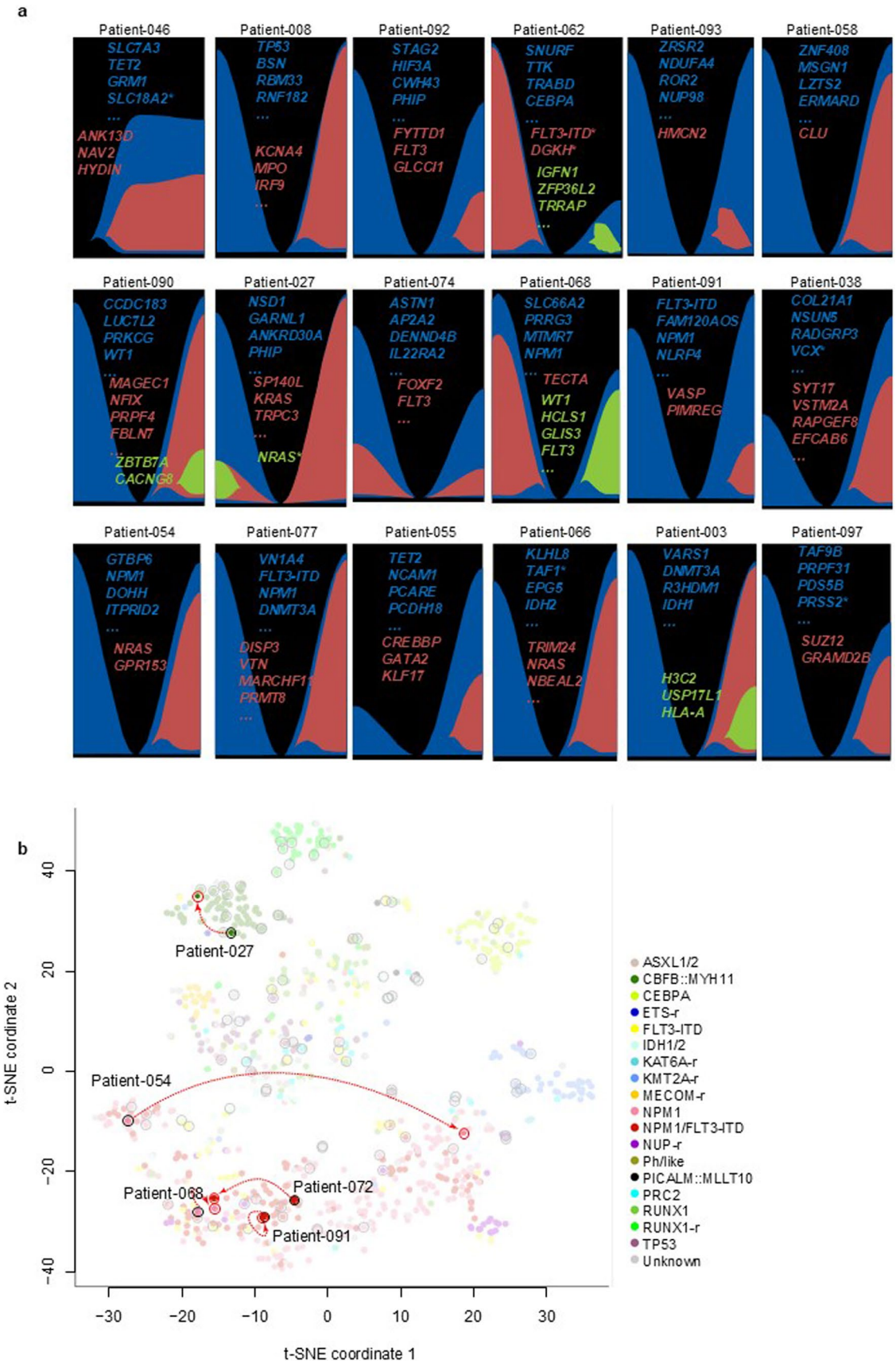
Extended Data Fig. 4 | Oncoprint of cytogenetically normal Black patients with AML. Depicted are all genes with recurrent variants detected in ≥2 patients.

https://doi.org/10.1038/s41588-024-01929-x



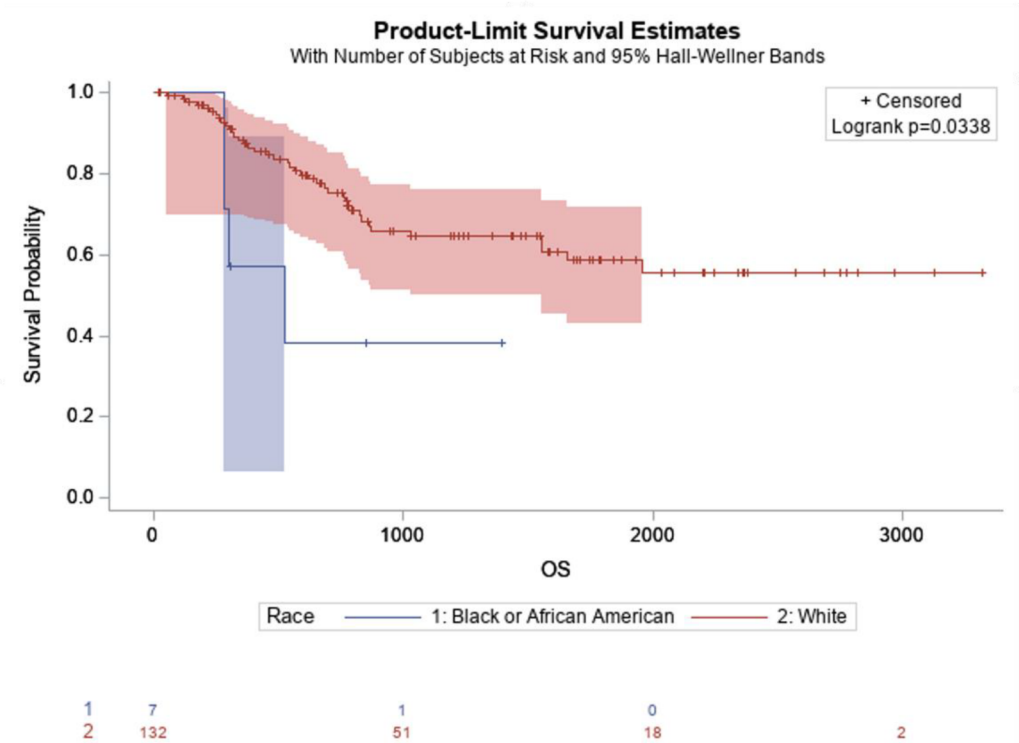
Extended Data Fig. 5 | Consort diagram. Depicted are the different patient cohorts used for comparative analyses.

https://doi.org/10.1038/s41588-024-01929-x



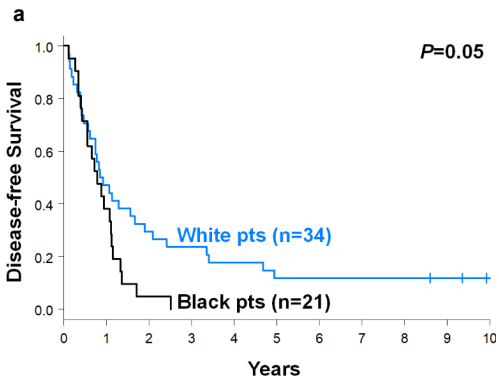
Extended Data Fig. 6 | Clonality analysis of paired diagnosis and relapse samples of 18 patients. a, Transcriptomic changes between diagnosis and relapse. **b**, Changes in gene expression between diagnosis and relapse samples.

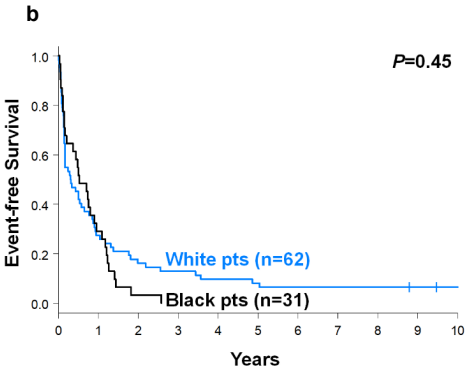
https://doi.org/10.1038/s41588-024-01929-x



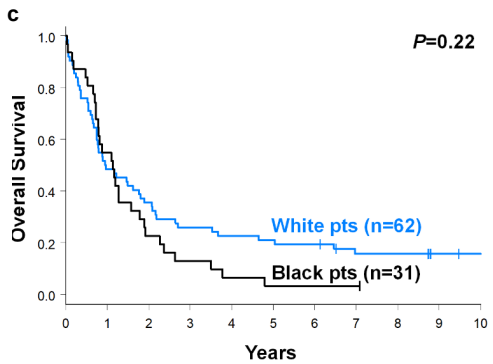
Extended Data Fig. 7 | Overall survival of Black and White patients aged < 60 years with AML and *NPM1* mutations treated with intensive chemotherapy from the Flatiron database. Survival estimates were calculated using the Kaplan-Meier method and compared using the two-sided log-rank test, confidence bands are the 95% Hall-Wellner Bands.

https://doi.org/10.1038/s41588-024-01929-x



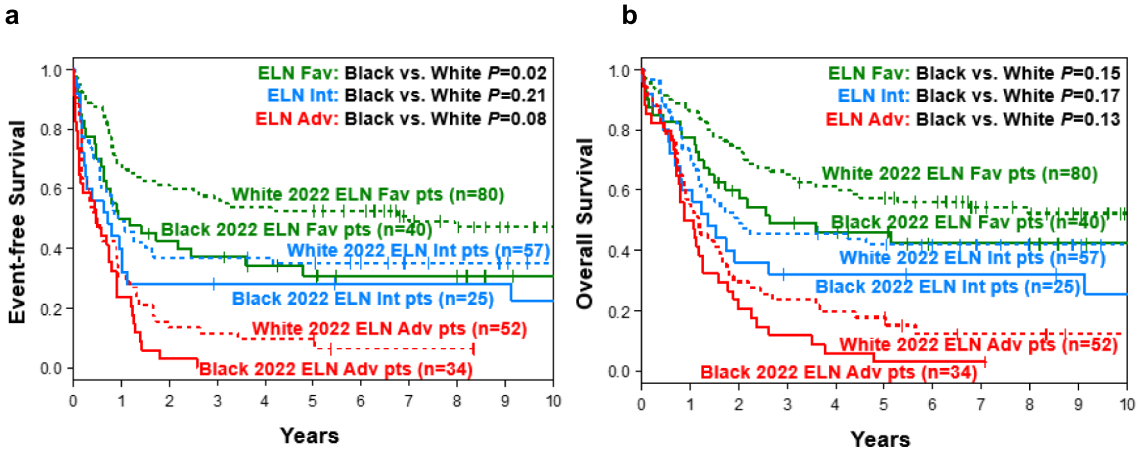


	Black patients (n=31)	White patients (n=62)	P
ED	2 (6%)	5 (8%)	1.00
CR	21 (67%)	34 (55%)	0.27
Relapse	20 (95%)	23 (68%)	0.02



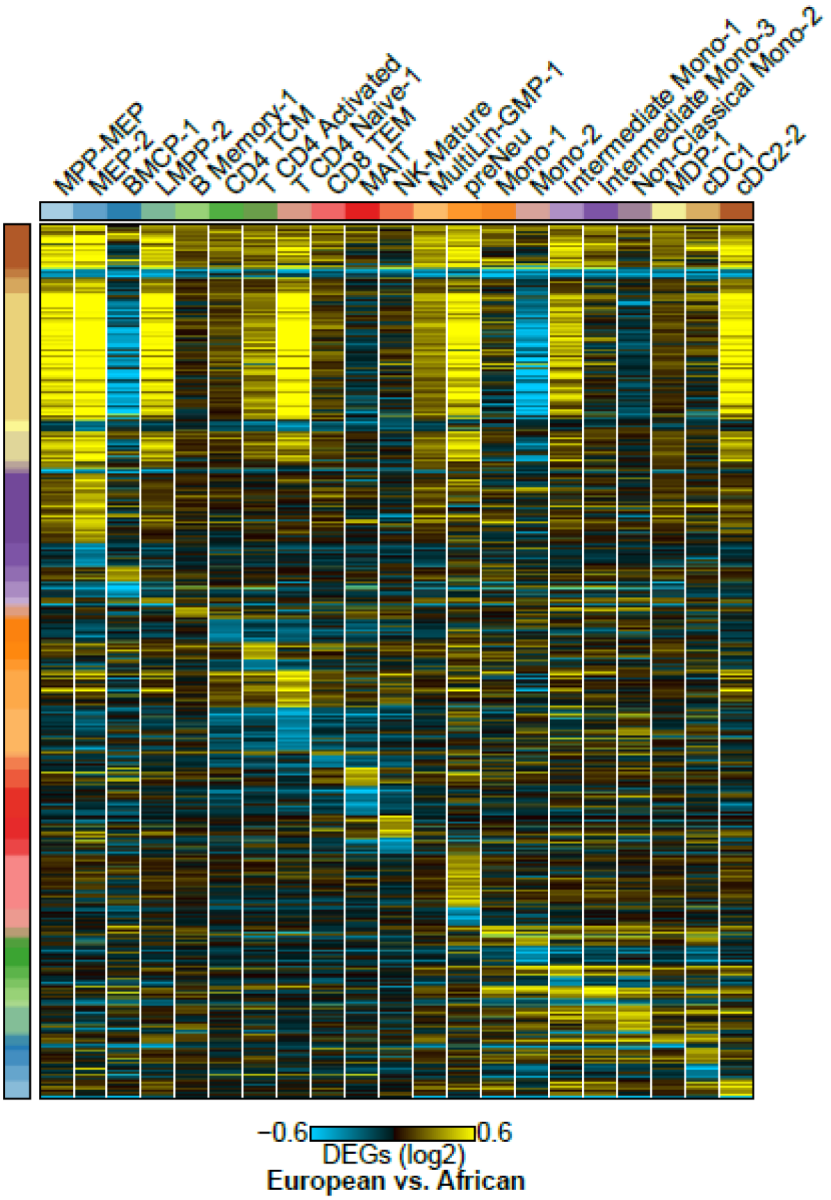
Extended Data Fig. 8 | Outcome of Black and White patients with AML with myelodysplasia related mutations. a, Disease free survival. The table contains the rates of early death (ED), complete remission (CR), and relapse. **b**, Event-free survival. **c**, Overall survival. For ED, CR and relapse rates (**a**), P-values were calculated using two-sided Fisher's exact test. For time-to-event analyses (**a**-**c**), survival estimates were calculated using the Kaplan-Meier method and compared using the two-sided log-rank test.

https://doi.org/10.1038/s41588-024-01929-x



Extended Data Fig. 9 | Outcome of Black and White patients according to 2022 ELN genetic-risk group. a, Event-free survival, **b**, overall survival. Survival estimates were calculated using the Kaplan-Meier method and compared using two-sided log-rank test.

https://doi.org/10.1038/s41588-024-01929-x



Extended Data Fig. 10 | ScRNA-seq analysis. Heatmap depicting cell-state and lineage analyses of scRNA-seq profiles of Black and White patients.

nature portfolio

Corresponding author(s): Ann-Kathrin Eisfeld

Last updated by author(s): August 15, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	CALGB protocol data colletion form, REDCap, SAS files and excel files.
Data analysis	SAS, TIBCO Spotfire S+ and R were used for statistical analyses.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

RNA-seq data from de novo patients with AML was in part derived from our previously published datasets available publicly via GSE266099, GSE137851, GSE63646 and GSE216738. The additional RNAseq data (bulk and single cell) generated in this study is publicly available in Gene Expression Omnibus (GEO) via GSE266498. Genetic variants are deposited in dbGAP (phs 003728.v1).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Self-reported sex was confirmed by karyotype for each patient and included into the description of cohort characteristics as well as included as biologic variable into survival models.
Reporting on race, ethnicity, or other socially relevant groupings	Self-reported race for Black patients from CALGB/Alliance who underwent whole-exome sequencing was confirmed via SNVstory workflow (doi: https://doi.org/10.1101/2023.06.02.543369) (https://github.com/nch-igm/snvstory). For White CALGB/Alliance and White BeatAML patients, self-reported race was inferred via genomic data using the Zheng & Weir method for all patients with available data. For definition of race, the CALGB/Alliance Registration forms follow guidelines from the National Cancer Institute: 1) White: A person having origins in any of the original peoples of Europe, the Middle East, or North Africa. 2) Black or African American: A person having origins in any of the Black racial groups of Africa.
Population characteristics	The patient population of the study includes patients similarly treated with intensive induction chemotherapy on Alliance for Clinical Trials in Oncology (Alliance) protocols. The survival of Black patients was compared to younger 952 White patients treated on the same Alliance protocols. In brief, the patients received intensive cytarabine/daunorubicin-based induction chemotherapy and consolidation with high-dose chemotherapy or autologous hematopoietic stem-cell transplantation (HSCT; details of treatment trials are provided in the supplemental Data). Patients provided written informed consent to participate in protocols CALGB 8461 (cytogenetic studies), CALGB 9665 (leukemia tissue bank) and CALGB 20202 (molecular studies), which involved collection of pretreatment Bone Marrow (BM) aspirates and peripheral blood (PB) samples. Study protocols were approved by the Institutional Review Board at each center.
Recruitment	N/A
Ethics oversight	All patients provided written informed consent to participate in treatment studies before enrollment and for the research use of their specimens in agreement with the Declaration of Helsinki. Study protocols were approved by the Institutional Review Board at each center.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Patients were chosen based on availability of samples and clinical and genetic data to maximize sample size.
Data exclusions	Two patients with hyper-mutated AML were excluded for the purpose of determining recurrently mutated genes in Black patients.
Replication	Given absence of a validation set of Black AML patients at time of diagnosis, for whom treatment and molecular data are available, we utilized the molecular profiles of relapsed/refractory, sef-reported Black AML patients who were treated at Memorial Sloan Kettering Cancer Center as only available validation.
Randomization	Patients were not randomized to comparison groups, comparison groups were based on race.
Blinding	During data collection patients and investigators were blinded per individual CALGB treatment protocol guidelines and rules.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	Patients provided written informed consent to participate in protocols CALGB 8461 (cytogenetic studies), CALGB 9665 (leukemia tissue bank) and CALGB 20202 (molecular studies), which involved collection of pretreatment BM aspirates and PB samples. Study protocols were approved by the Institutional Review Board at each center.
Study protocol	CALGB 8461 (cytogenetic studies), CALGB 9665 (leukemia tissue bank) and CALGB 20202 (molecular studies)
Data collection	CALGB/Alliance patient data was collect using treatment protocol specific collection forms by CALGB between 1986-2022. The Ohio State University patient data was collected using REDCap.
Outcomes	For Alliance data, early death was defined as patient's death within 30 days of starting therapy, irrespective of cause. Complete Remission (CR) required an absolute neutrophil count ≥1.5 x 10^9/l, a platelet count >100 x 10^9/l, no leukemic blasts in the blood or BM, cellularity greater than 20% with maturation of all cell lines, no Auer rods, less than 5% BM blast cells, and no evidence of extramedullary leukemia, all of which had persisted for at least one month. Relapse was defined by ≥5% BM blasts, circulating leukemic blasts, or the development of extramedullary leukemia. Disease-free survival (DFS) was measured from the date of CR until the date of relapse or death; patients alive and relapse-free at last follow-up were censored. Overall survival (OS) was measured from the date on study until the date of death, and patients alive at last follow-up were censored. In our outcome analyses, we used P-values adjusted to control for per family error rate (probability of a Type I error) for all variables considered in univariable analyses. The families were all variables considered in each outcome analysis and only the variables whose likelihood ratio test adjusted P-value was <0.20 from the univariable models were considered in the multivariable analysis.

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A