1 **Hierarchical Lineage Tracing Reveals Diverse Pathways of AML Treatment Resistance**

2

3 Rachel Saxe[1,2], Hannah Stuart[1,3], Abigail Marshall[1,2], Fahiima Abdullahi[5], Zoë Chen[4], Francesco

4 Emiliani[1,2], Aaron McKenna[1,4]

5

6 [1]Molecular and Systems Biology, Dartmouth College, Hanover, NH
7 [2]Molecular and Cellular Biology Program, Dartmouth College, Hanover, NH
8 [3]Quantitative Biomedical Science Program, Dartmouth College, Lebanon, NH
9 [4]Dartmouth Cancer Center, Dartmouth College, Lebanon, NH
10 [5]The Dartmouth MD-PhD Undergraduate Summer Fellowship Program, Lebanon, NH

11 Abstract

12 Cancer cells adapt to treatment, leading to the emergence of clones that are more aggressive

13 and resistant to anti-cancer therapies. We have a limited understanding of the development of

14 treatment resistance as we lack technologies to map the evolution of cancer under the selective

15 pressure of treatment. To address this, we developed a hierarchical, dynamic lineage tracing

16 method called FLARE (Following Lineage Adaptation and Resistance Evolution). We use this

17 technique to track the progression of acute myeloid leukemia (AML) cell lines through exposure

18 to Cytarabine (AraC), a front-line treatment in AML, in vitro and in vivo. We map distinct cellular

19 lineages in murine and human AML cell lines predisposed to AraC persistence and/or

20 resistance via the upregulation of cell adhesion and motility pathways. Additionally, we highlight

21 the heritable expression of immunoproteasome 11S regulatory cap subunits as a potential

22 mechanism aiding AML cell survival, proliferation, and immune escape in vivo. Finally, we

23 validate the clinical relevance of these signatures in the TARGET-AML cohort, with a bisected

24 response in blood and bone marrow. Our findings reveal a broad spectrum of resistance

25 signatures attributed to significant cell transcriptional changes. To our knowledge, this is the first

26 application of dynamic lineage tracing to unravel treatment response and resistance in cancer,

27 and we expect FLARE to be a valuable tool in dissecting the evolution of resistance in a wide

28 range of tumor types.

Introduction

Acute myeloid leukemia (AML) is a heterogeneous hematological malignancy that has served as a model of clonal evolution in response to treatment[1–4]. AML subtypes and risk categories have long been primarily defined by cytogenetics, mutation status, and structural genomic variations. However, with recent advances in RNA sequencing, transcriptome profiling has begun to uncover further leukemic disease states, dependencies, and prognostic signatures[5–7]. Upon diagnosis, most patients are treated with induction therapy, including an aggressive cycle of Cytarabine (AraC). While ~80% of patients reach remission following induction, most will relapse[8]. Single-cell RNA sequencing (scRNA-seq) of matched pre- and post-treatment patient bone marrow samples show patients have dramatic disease landscape changes following AraC treatment[2].

It remains unclear how resistance to AraC evolves in many AML patients. Are AraC-resistant subclones pre-existing, or does resistance emerge de novo in rare subpopulations? Does the resistant phenotype pass through multiple intermediate states or emerge in a punctuated evolution? Previous studies have investigated the origin of chemoresistant cell populations in various cancer types[9–12]. Shaffer et al. concluded that while rare subpopulations of cells exhibit pre-resistant-like phenotypes, melanoma cells could transiently shift to a resistant phenotype upon treatment in vitro[11]. In contrast, Yan et al. identified a subpopulation of TP53-mutated AML cells in treatment-naïve samples believed to be the source of future AraC-resistant populations[9]. These studies highlight that treatment failure and resistance are likely driven by diverse disease, drug, and likely, patient-specific mechanisms. While individual patients may display unique lineage developments and gene expression profiles, pathway alterations that arise frequently can be exploited for treatment response prediction, patient stratification, and development of targeted therapies.

We set out to develop a cellular lineage tracing system to record the evolutionary steps of AraC treatment resistance in AML. Here, we present FLARE (Following Lineage Adaptation and Resistance Evolution), a multi-tiered lineage tracing approach, which we use to identify clonal patterns of drug response and expansion in vitro and in vivo. Using advances in CRISPR lineage tracing, we track the evolution of cell fate independently of cell state, recording the relationship between cells in small, integrated recorders, which can be used to recover cellular 'phylogenetic' trees[13,14]. This lineage information alone can uncover treatment response

62   patterns, but when paired with transcriptomic data, it acts as a window into disease evolution.

63   CRISPR lineage recording has been previously used to characterize clonal evolution through

64   tumorigenesis and metastasis, yet the effect of treatment has yet to be explored[14,15].

65

66   We use FLARE to uncover multiple heritable and lineage-specific gene modules associated with

67   drug response with matched outcomes in pediatric AML patient survival. FLARE lineage

68   patterns uncovered a highly heritable gene signature in treatment-naive AML cells involved in

69   cell-adhesion-mediated drug resistance (CAM-DR) pathways, the expression of which leads to

70   lineage branch survival and expansion under AraC exposure. In addition, we identified an AML

71   lineage-dependent upregulation of immunoproteasome subunits, with further upregulation

72   leading to improved expansion in vivo. Finally, we validated the clinical relevance of our findings

73   using RNA sequencing data from the pediatric TARGET AML cohort and discover bone marrow

74   and peripheral blood-specific expression profiles significantly associated with patient outcomes.

Results

75

Hierarchical lineage recording in AML

76

77   We set out to develop a dynamic cellular lineage tracing system to model the fitness landscape

78   of AML in vitro and in vivo. Expanding on our previous macsGESTALT lineage tracing system[14],

79   we created a hierarchical system, capturing lineage information at three levels (**Figure 1A-B**).

80   At the top level, FLARE encodes founder populations using a unique combination of static tags

81   accompanying the lineage recorders when integrated into the genome. Within each labeled

82   founder population, individual clones are divided into sub-populations with a second round of

83   tags integrated with the Cas9 cassette. This Cas9 integration also begins the dynamic lineage

84   recording, creating insertions and deletions (indels) within the lineage recorders, capturing sub-

85   clonal lineage information. We can then use computational methods to generate a phylogenetic

86   tree of cell relationships within these founder and clonal populations from these indels. This

87   three-level structure (founder, clone, and subclonal lineage information) allows FLARE to track

88   dynamics across time and to compare sister populations in both the control and treatment arms,

89   in vitro and in vivo.

90

We first engineered this approach in the C1498 cell line, a syngeneic mouse model of AML compatible with transplantation into immune-competent C57BL/6 mice. Despite being long-passaged, C1498 cells are known to be heterogeneous, making them an optimal model for investigating the diversity of clonal treatment response[16]. We established 11 founder populations from individually isolated cells through successive engineering rounds (**Figure S1A-B**). Sequencing revealed a unique set of founder tags for each population, varying from 19 to 30 integrations (**Figure S1C**). These isolated founders are heterogeneous, with a wide range of growth kinetics (**Figure S1D**). We found no putative driver mutation differences between individual founders using whole-genome sequencing that would explain this heterogeneity.
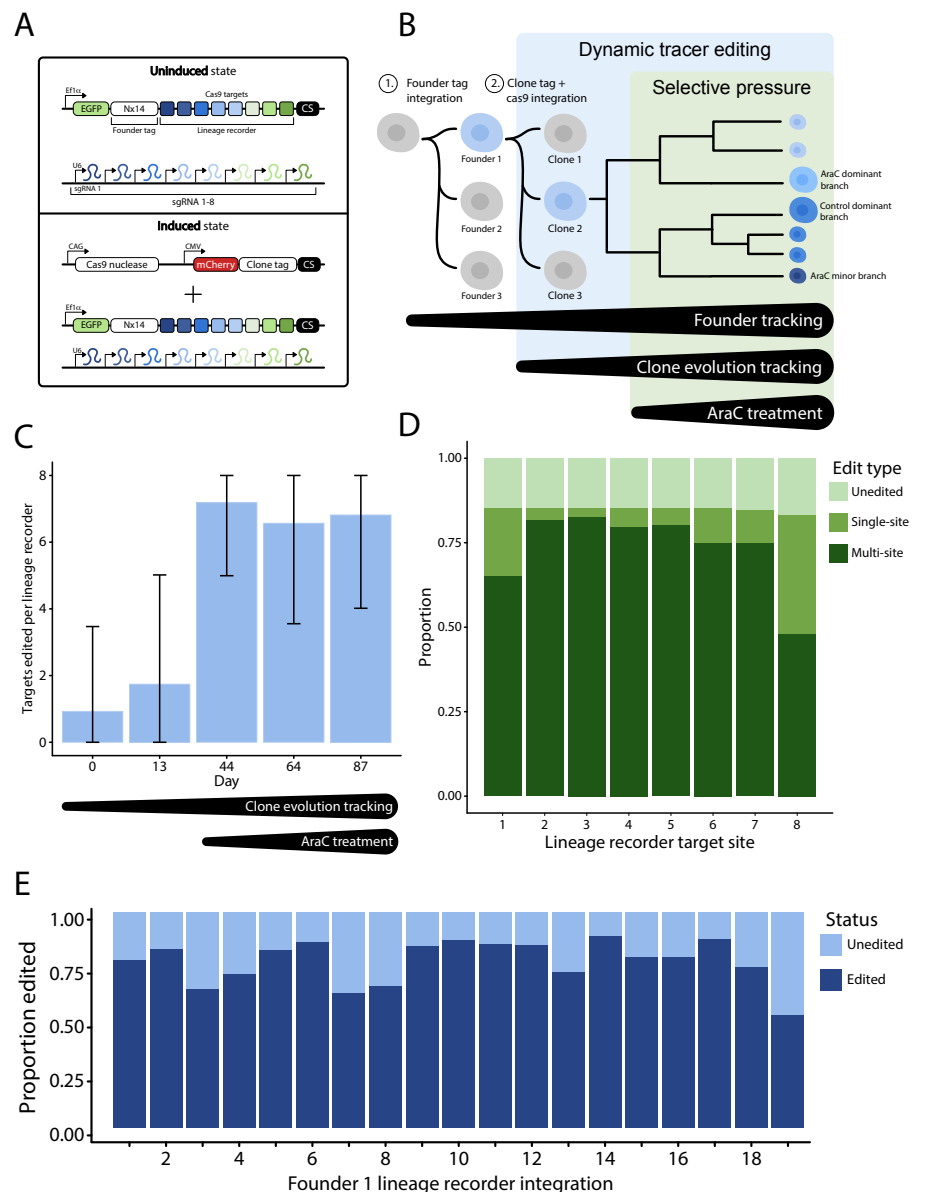

Figure 1: FLARE lineage tracing system allows for hierarchical tracking of cell populations through time. (**A**) Constructs required for FLARE lineage tracing. The top panel includes FLARE-miniV10-10X-UMI (founder tag + lineage recorder) and FLARE-pbV10g-neo (Cas9 guide sequences). We then electroporate the FLARE-pbCas9-tag-mCherry into cells (bottom panel), tagging each clone and activating CRISPR lineage recording. (**B**) The resulting cells record lineage information at 3 levels, (1) founder, (2) clonal, and (3) subclonal (CRISPR recordings). AraC treatment selective pressure is introduced following the initiation of lineage recorder editing. (**C**) Lineage recorders are readily edited following the introduction of Cas9. Editing rates plateau after around 44 days of Cas9 exposure. (**D**) Following 87 days of Cas9 exposure lineage recorder edits are primarily composed of multi-site editing events, with a smaller proportion of edits impacting single sites. Around 15% of lineage recorder sites remain unedited through 87 days of editing. (**E**) Editing rates are consistent across lineage recorder integrations within the C1498-FLARE Founder 1 population.

Characterizing lineage recording

Next, we set out to test the lineage recording capacity of our engineered cells. We activated CRISPR recording by integrating the Cas9 construct into the genome (**Figure 1A-B**). After electroporation and sorting, we sampled Cas9-positive C1498-FLARE cell populations at days 13, 44, 64, and 87 and sequenced the lineage recorders to characterize Cas9 editing (**Figure 1C**). Recording proportions suggest the bulk of lineage recorder editing occurs between day 13 (22.61% of recorders contain >= 1 edit) and 44 (99.97% >= 1 edit). Editing rates remained high through at least day 87 (85.93% >= 1 edit), with continuing fluctuations in editing patterns due to population drift. Editing patterns were comparable between lineage recorder target sites (**Figure 1D)**, and most edits span multiple target sites with a smaller proportion of edits affecting individual targets. Approximately 15% of sequenced lineage recorder targets remained unedited, indicating our ability to trace lineage over at least 87 days without saturation.

We further characterized the response of individual recorder integrations with a single founder group, founder 1 (F1), containing 19 unique integrations. All integrations were active recorders displaying diverse editing outcomes (**Figure 1E**). For all founder populations assessed, individual recording events tended to be large, spanning 3.075 targets on average, and each recorder contained an average of 2.57 individual recording events. While edits consisting of large deletions can lead to loss of lineage information, we could still reliably reconstruct lineage trees with an average depth of 5.566 nodes over relevant timescales. We attribute this to this system's high lineage recording capacity, with engineered cells containing between 152 and 240 CRISPR target sites per cell.

Clonal treatment response and tree generation in vitro and in vivo

Confident in our ability to record and recover cellular lineage information, we set out to characterize C1498-FLARE population dynamics in response to the chemotherapy AraC in vitro and in vivo (**Figure 2A)**. To recreate the heterogeneity observed in leukemia patients, we pooled four C1498 founder groups for our starting population, which was then used for both in vitro and in vivo AraC treatment.

153   By day 43 of in vitro AraC treatment we observed two distinct growth phenotypes. Our AraC-

154   resistant flask displayed rapid growth and a 10-fold higher ED50 than control cell populations

155   (**Figure S2A**). We also observed a separate flask phenotype, where cells appeared to be

156   "persistent", displaying minimal proliferation and cell numbers too low for ED50 determination.

157   We sequenced 27,237 single cells from the untreated, resistant, and persistent flasks. Treated

158   cells clustered into distinct groups predominantly separated by flask phenotype (**Figure 2B**). We

159   successfully captured founder tags and lineage recorders from 100% of single cells and clone

160   tags from 31.9%. After filtering, quality control, and lineage assignment, 25,489 single cells were
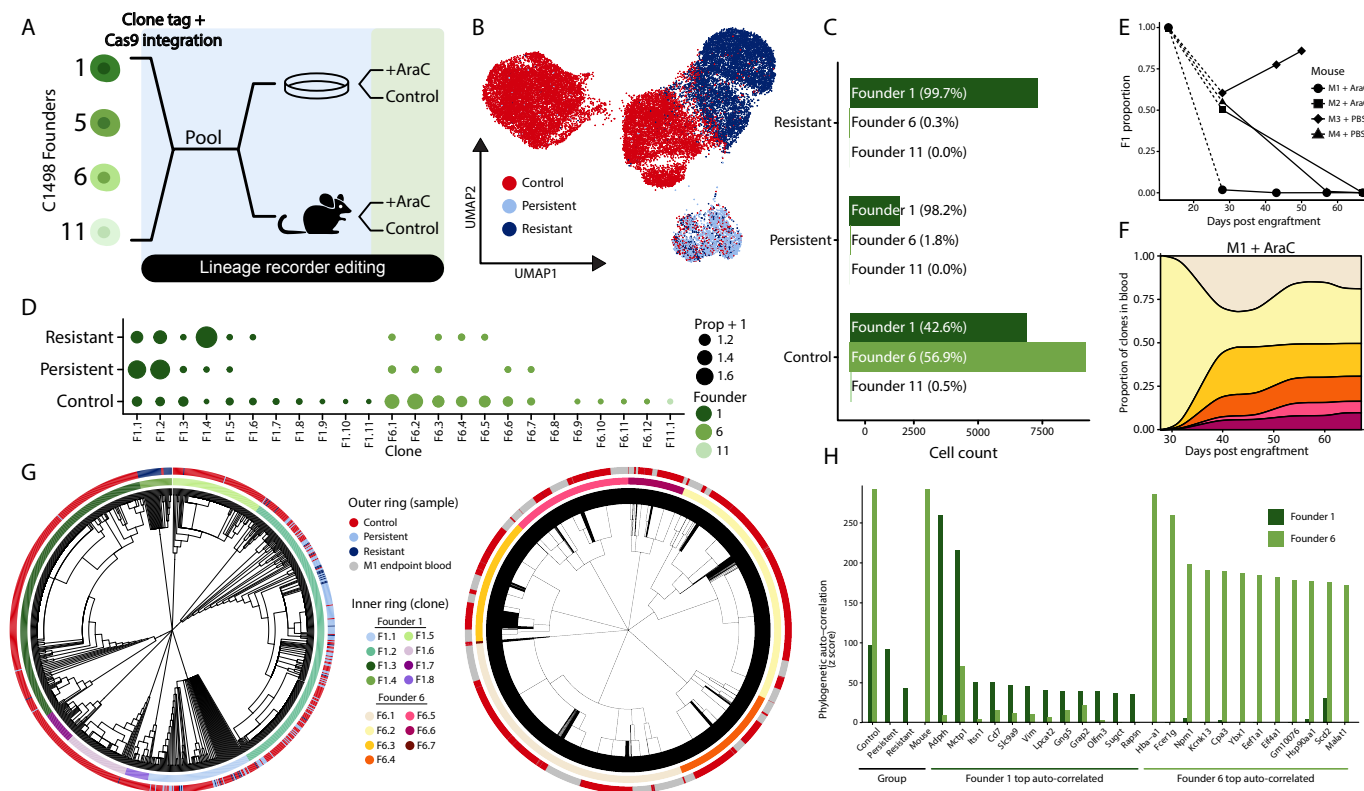

Figure 2: FLARE lineage tracing tracks founder and clone-level evolution in response to AraC treatment in vitro and in vivo. (**A**) Experimental design of lineage tracing C1498-FLARE AML cells in vitro and in vivo. (**B**) Single-cell sequencing of C1498-FLARE cells following 43 days of in vitro AraC treatment. Cells cluster into distinct groups associated with their sample (red - no AraC; dark blue - AraC resistant flask; light blue - AraC persistent flask). (**C**) Founder proportions in each sample display increased survival of F1 cells under AraC treatment in both the resistant and persistent populations. (**D**) Clone populations detected within each sample, ordered by founder and clone size. (**E**) Founder and clone tags detected in blood draws from mice engrafted with C1498-FLARE cells were used to calculate the proportion of circulating leukemic cells belonging to F1 compared to F6. Day 13 proportions were computed using founder tag sequencing and are connected to day 28 with dashed lines. All other timepoint proportions were calculated using clone tag sequencing, normalized for differences in the number of integrations. The F1 proportion was defined as the F1 average tag count divided by the sum of F1 and F6 average tag counts. Untreated mouse M3 was euthanized on day 50 due to disease progression. The remaining mice were euthanized on day 67. (**F**) Bulk sequencing of clone tags detected in blood draws from mouse M1 following AraC treatment on day 13. Clone tag reads were used to determine the clonal proportions of leukemic burden in the blood from day 28 to day 67. (**G**) Phylogenetic reconstruction of founders F1 (left) and F6 (right) generated from single-cell sequencing of in vitro C1498-FLARE lineage tracing and single-cell sequencing of endpoint blood and bone marrow from treated mouse M1. (**H**) Heritability of both sample features and genes for founder 1 and founder 6 computed with PATH.

161   assigned to founder groups, of which 5,468 were successfully assigned to clones. Founder tag

162   sequencing revealed in vitro fitness differences: in untreated flasks, founder 1 (F1) and founder

163   6 (F6) dominated (43% and 57% of cells, respectively), and few founder 11 (F11) cells were

164   detected (<0.5% of cells). F1 and F6 control cells clustered separately, suggesting distinct

165   founder-specific transcriptomes (**Figure S2B)**. In contrast, treated populations were primarily

166   composed of F1 with few F6 (<0.7%) and zero F11 cells. F5 was outcompeted in all conditions

167   (**Figure 2C)**.

168

169   We performed differential gene expression analysis between founder groups, focusing on F1

170   and F6 in the control setting. F1 lineages were characterized by high expression of Nkain3,

171   Wnt6, and Chl1 (**Table S1**). In contrast, F6 lineages were characterized by high expression of

172   Sox2ot, Tnni1, and Tnnt1 (**Table S2**). GSEA results indicated enrichment of multiple Hallmark

173   pathways in F6 compared to F1, including MYC targets V1, oxidative phosphorylation, and

174   MTORC1 signaling (**Figure S2C**). Enriching these oncogenic pathways likely contributed to the

175   success of F6 in the absence of treatment. Since these pathways have well-established roles in

176   chemoresistance, we expected F6 to display a survival advantage under AraC treatment[17–20].

177   Unexpectedly, F6 was almost fully out-competed by F1 in the presence of AraC (**Figure 2C**).

178

179   We then explored founder F1's competitive advantage over F6 and the other founders during

180   treatment. Within F1 populations, there was a distinct and divergent selection of specific clones

181   in AraC-treated and control flasks (**Figure 2D**). The highest proportion of persistent cells

182   belonged to clones F1.1 and F1.2, while the highest proportion of resistant cells belonged to

183   F1.4. F1 cells have a clear survival advantage through in vitro AraC treatment, but the

184   outgrowth of different clones in treated replicates suggests clone-specific responses that

185   contribute to population growth dynamics.

186

187   To identify F1 clone-specific treatment responses, we analyzed differentially expressed genes

188   between control, resistant, and persistent-dominant clones (**Table S3**). Our dominant clone in

189   the resistant population, F1.4 (**Figure 2D**), expressed high levels of Igll1 and Evac1 compared

190   to other resistant cells. Clones F1.1 and F1.2 were successful in both persistent and resistant

191   populations; when we compared their control gene expression to all other F1 clones, a

192   transcriptional signature containing Grap2, Trat1, and Ldlrad4 emerged (**Figure S2D**).   Trat1 is

193   involved in the regulation of T cell receptors and expression is associated with immune cell

194   infiltration and improved survival in breast, ovarian, and lung cancer [21–23], Ldlrad4 is a negative

195   regulator of TGF-beta signaling [24,25], and expression promotes the proliferation and migration of

196   hepatic cancer cells and is downregulated in CD34+ cells from patients with MDS [26]. Consistent

197   with the upregulation of Ldlrad4 expression in F1.1 and F1.2 control cells, sister cells of both

198   clones in the persistent population display high Ldlrad4. We also identified clones with pre-

199   existing lineage-specific expression patterns. Cells belonging to clone F1.3 tended to colocalize

200   and shared high expression of Adprh and Mctp1 (**Figure S2E-F**). These genes highlighted by

201   differential gene expression analysis do not have well-established roles in AML.

202

203   In parallel to our in vitro treatment assay, we injected immune-competent C57BL/6 KH2[27] mice

204   with 25k of the pooled founder starting population to induce leukemia with lineage tracing

205   capacity (**Figure 2A**). Mice were then treated with AraC to a dose comparable to that given to

206   patients in one round of 7+3 induction chemotherapy[9]. We used bulk lineage sequencing to

207   estimate founder and clone dynamics from blood draws of circulating leukemia cells. Contrary to

208   our in vitro results, F6 outcompeted other founder populations in vivo in all but one mouse

209   (**Figure 2E**). This bottlenecking occurred sometime between day 28, when all founders were

210   detectable in control and treated mice (**Figure S3A),** and euthanasia on day 67. A selective

211   sweep was nearly complete in treated mouse M1 (>99.9% F6 by day 43), treated mouse M2

212   (100% F6 by day 67), and untreated mouse M4 (>99.9% F6 by day 57). Control mouse M3's

213   leukemic burden was composed of >77.5% F1 by day 43, and this increased to >85.9% on day

214   50 when M3 was euthanized due to symptomatic disease.

215

216   We then focused on the within-founder clone tag sequences seen in blood draws. Interestingly,

217   M1 displayed sustained clonal diversity split amongst F6.1, F6.2, F6.3, F6.4, and F6.5,

218   indicating these clones survived AraC treatment and had comparable fitness during post-

219   treatment outgrowth. In contrast, blood samples from M3 were primarily dominated by F1's

220   clone F1.2 by day 43 (**Figure S3C)**. Blood samples from M2 and M4 did not contain enough

221   overlap with our known clone tags to infer population dynamics, likely representing diversity not

222   captured in our single-cell sequencing at end-point.

223

224   To further examine the lineage identities and transcriptomes of cells surviving in vivo treatment,

225   we sequenced 10,473 and 6,344 cells from blood and bone marrow, respectively, of treated

226   mouse M1 displaying sustained F6 clonal diversity. Non-leukemic cells clustered by cell type,

227   and leukemic cells were identified by founder and clone tags (**Figure S3D-E**). Of the leukemic

228   cells, 6 and 6,053 belonged to founders F1 and F6, respectively. Clone groups were assigned to

229   2,002 single cells, 702 belonging to clone F6.1. The remaining clone assignments were split

230   between F6.2, F6.3, F6.4, and F6.5 (**Figures 2F, S3B**). We also detected low frequencies of

231   cells belonging to F6.6 and F6.8. These single-cell sequencing results matched the population

232   detected in bulk sequencing, confirming our ability to reconstruct founder and clone patterns

233   from limited samples, such as blood draws, using FLARE lineage tracing.

234

235   We then created phylogenetic trees for F1 and F6 using FLARE lineage recorder indels

236   captured with our in vitro and in vivo scRNA-seq (**Figure 2G**). To identify heritable gene

237   expression patterns associated with the development of AraC resistance, we conducted lineage

238   autocorrelation analysis on founder trees using PATH (**Figure 2H**)[28]. PATH highlighted genes

239   previously identified by differential gene expression analysis as highly cross-correlated with

240   resistant and/or persistent cell identities in the F1 tree (**Figure S3F**). These genes include

241   Olfm3, a gene associated with anoikis resistance in AML cells[29], and Cd7, the expression of

242   which in AML is a prognostic risk factor[30–32]; both genes were significantly upregulated in AraC

243   persistent cells. Genes such as Adprh and Mctp1, associated with the F1.3 clone branch, were

244   also strongly heritable but expressed primarily in a subset of control cells.

245

246   These initial experiments with C1498-FLARE cells allowed us to optimize the lineage tracing

247   system and test our ability to track leukemic lineage in immune-competent mice. With FLARE,

248   we successfully identified highly heritable genes selected for through AraC treatment that

249   contribute to drug persistence and resistance.

250   ## Clonal treatment response and tree generation in human AML cells

251   We next engineered FLARE lineage recording human HL60 cells, a model of FAB-M2 AML[33].

252   We seeded 500 HL60-FLARE founder cells, expanded, and split into parallel cultures (**Figure**

253   **3A**). This starting pool expands on the transcriptional landscape sampled in C1498-FLARE cells

254   while maintaining sister founder populations in the control and treatment arms. Similarly to

255   C1498-FLARE cells, our resulting HL60-FLARE cells clustered in distinct groups based on AraC

256   treatment in vitro (**Figure 3B**). However, unlike C1498-FLARE cells, AraC-treated HL60-FLARE

257   cells in different flasks clustered together despite different regrowth rates (**Figure S4A**). AraC-

258   treated flask A reached cell concentrations comparable to untreated HL60-FLARE cells by day

259   80 of AraC treatment, so we posited these cells would align transcriptionally with C1498-FLARE

260   resistant cells. Surprisingly, differential expression analysis uncovered the upregulation of genes

261   shared with AraC-persistent C1498 F1 cells such as VIM, PSD3, CRIP1, and S100A4 (**Table**

262   **S4**). AraC-treated flask B displayed slower growth dynamics but shared flask A's transcriptional
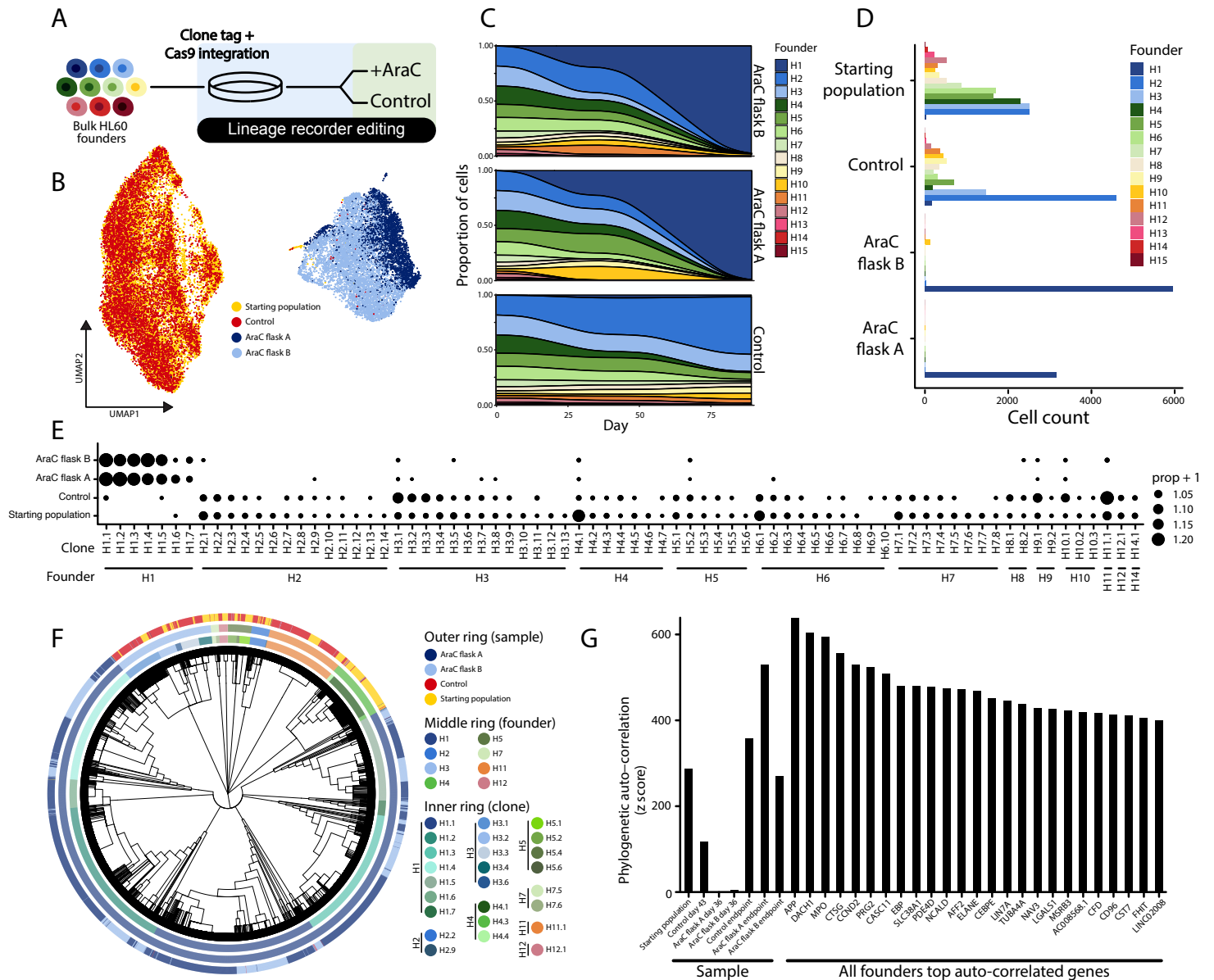
263   profile (**Table S5**).

264

Figure 3: Founder and clone-level selection of founder H1 in response to AraC treatment in human HL60 AML cells. (**A**) Experimental design of HL60-FLARE AML lineage tracing in vitro. (**B**) Single cell sequencing of HL60-FLARE cells following 89 days of in vitro AraC treatment. Cells cluster into distinct groups associated with their sample (yellow - starting population; red - no AraC; dark blue - AraC flask A; light blue - AraC flask B). (**C**) Clone population fluctuations for each flask throughout the experiment. (**D**) Founder proportions in each sample display increased survival of founder H1 cells under AraC treatment in both flask A and B. (**E**) Clone populations for each sample at experimental endpoint, ordered by founder and clone size. (**F**) Phylogenetic reconstruction of HL60-FLARE cells generated from single cell sequencing of in vitro lineage tracing. (**G**) Heritability of both sample features and genes for all HL60-FLARE founders computed via PATH.

265 Single-cell sequencing of founder and clone tags was used to cluster cells into 15 founder

266 populations (H1-H15), individually composed of 1-14 clones. Sequencing from cells collected at

267 the experiment start, midpoint, and endpoint were used to determine founder proportions over

268 time for each flask. All 15 HL60 founders were detected in the experimental starting population

269 and the control flask at endpoint (**Figure 3C-D**). There was an extensive bottleneck in all

270 cultured populations, regardless of treatment. Untreated endpoint cells were most likely to

271 descend from H2, while both AraC-treated flasks consisted primarily of cells belonging to H1.

272 Both treated flasks maintained diversity within H1 on the clonal level, with most cells being

273 evenly split between clones H1.1, H1.2, H1.3, H1.4, and H1.5 (**Figure 3E**).

274

275 The success of H1 in both treated flasks suggests a founder-specific propensity for treatment

276 resistance prior to treatment onset. This is also supported by the comparable fitness of H1

277 clonal populations, indicating the lack of clone-specific resistance evolution. We used differential

278 gene expression analysis to explore H1-specific expression patterns in the presence and

279 absence of treatment to determine whether H1 cells were consistently differentiated from other

280 founders or were primed for reprogramming following the introduction of AraC. Pre-experiment

281 and endpoint control H1 cells shared the upregulation of 104 genes (**Table S6**). Many of these

282 genes were also upregulated in AraC-resistant H1 cells compared to other founders, including

283 MKX, APBA2, LAMP5, CAMK2D, and NAV3.

284

285 We turned to phylogenetic reconstruction and heritability analysis to further dissect the H1's

286 resistance mechanisms. We used lineage recorder indel patterns to reconstruct phylogenetic

287 trees for each clone population. These clonal trees were merged to create founder-level lineage

288 trees, and founder-level trees were merged to create the HL60 population tree (**Figure 3F**).

289 PATH analysis on the population tree revealed many highly heritable genes associated with

290 evolution with and without AraC, including genes previously implicated in AML prognosis and

291 treatment response, such as MPO (**Figure 3G**). This analysis also indicated significant

292 heritability of genes previously identified as upregulated in treatment-naive H1 cells.

Identification of pro-leukemic transcriptional programs associated with pre-

294 AraC-resistant lineages

295 To better understand broad transcriptomic changes over AraC-induced clonal evolution, we

296 generated lineage-dependent gene modules for C1498-FLARE F1 and F6 trees, and the HL60-

297   FLARE population tree using Hotspot (**Figure 4A-B, Table S3, Table S7**)[34]. Jaccard indices

298   revealed significant similarity between a number of HL60 and C1498 modules, indicating shared

299   mechanisms across our two experimental lines (**Figure 4C, S4B-C**). Moreover, expression of

300   these modules within the lineage structure showed founder and clone-specific enrichment
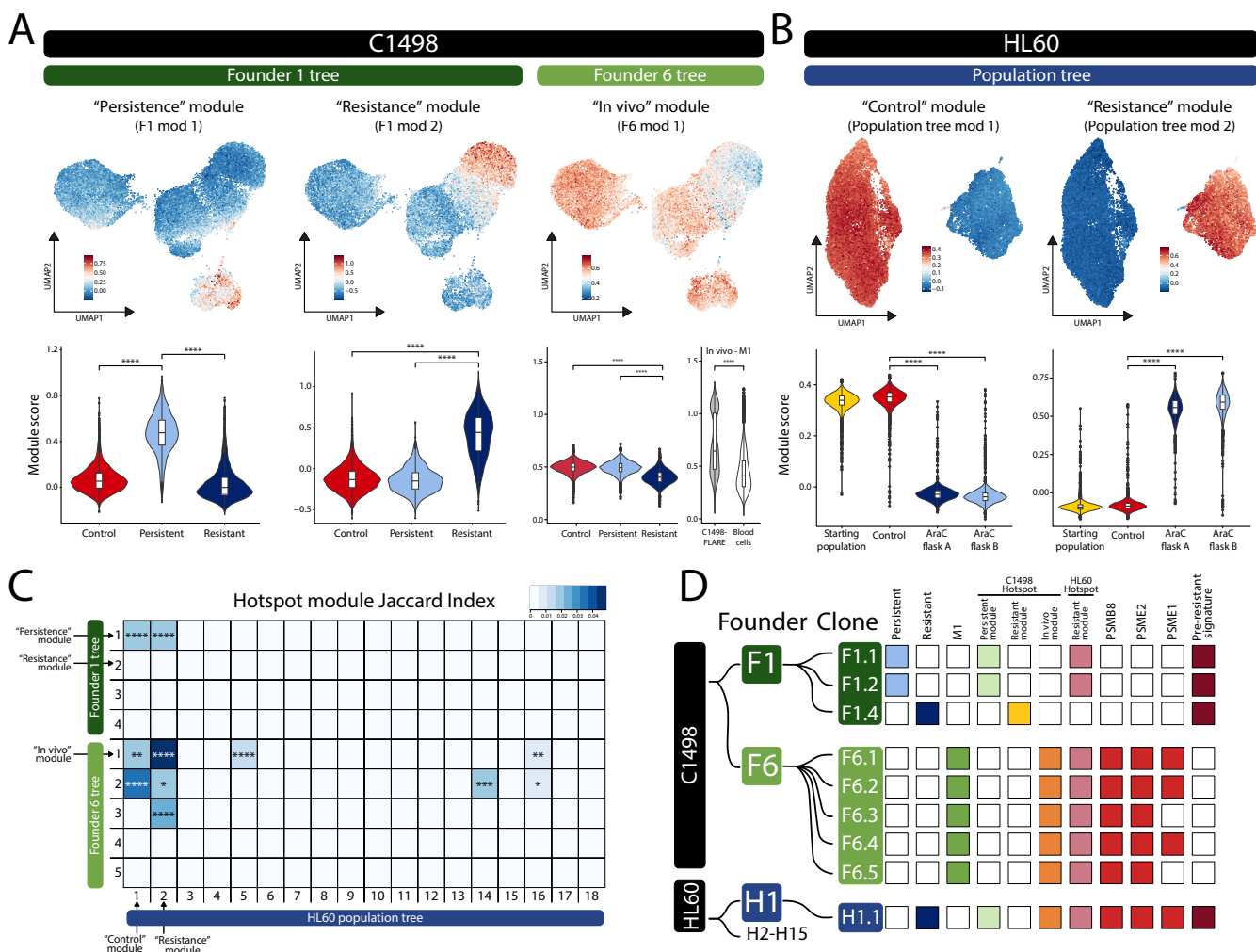
301   (**Figure 4D**).

302

Figure 4: Hotspot gene modules reveal heritable AraC response mechanisms common to mouse and human AML cells. (**A**) Feature and violin plots for selected C1498-FLARE Hotspot modules. Significance determined by wilcoxon rank sum test with Bonferroni correction. (**B**) Feature and violin plots for selected HL60-FLARE Hotspot modules. Significance determined by wilcoxon rank sum test with Bonferroni correction. (**C**) Jaccard Index similarity comparing C1498-FLARE and HL60-FLARE Hotspot modules. Square color indicates Jaccard Index, asterisks represent BH-corrected P-value significance levels. (**D**) Schematic showing the lineage relationships of C1498-FLARE and HL60-FLARE clones. The first three columns to the right of the clone labels indicate the setting in which each clone had a survival advantage, either in the Ara-C persistent population (a growth phenotype we only saw in C1498-FLARE populations), resistant population, or in AraC-treated mouse M1's peripheral blood and bone marrow. The next four columns to the right indicate modules upregulated in each clone. The next three columns indicate whether immunoproteasome gene PSMB8 and 11S regulatory cap genes PSME1 and PSME2 were significantly upregulated in the AraC treated populations. The rightmost column indicates the upregulated expression of the pre-resistant signature in AraC-treated populations.

303   We first focused on the overlap between HL60-FLARE "resistance" module 2 and C1498-

304   FLARE "persistence" F1 module 1 (**Table S9**), the most heritable modules enriched in treatment

305   conditions. The HL60-FLARE "resistance" module was distinctly enriched in all AraC-treated

306   cells (**Figure 4B**), while the C1498-FLARE "persistence" module was associated with AraC-

307   persistent populations (**Figure 4A**). The HL60 resistance module was primarily expressed in the

308   H1 founder population and contained many genes found to be distinctly upregulated in

309   treatment-naive H1 cells (**Table S6-7**). This suggests that the "resistance" module describes a

310   lineage-specific profile that can identify cells most likely to become refractory to AraC treatment.

311

312   The HL60-FLARE "resistance" module shared 9 genes with the C1498 "persistence" module:

313   ITSN1, VIM, CRIP1, S100A11, ANXA2, MYL6, S100A4, PSD3, and DOCK10. This set of genes

314   is highly active in cytoskeleton maintenance and remodeling. Such pathways are implicated in

315   the epithelial-to-mesenchymal transition (EMT), a type of transdifferentiation involved in cancer

316   cell migration, invasion, and drug resistance[35]. While EMT is less well studied in hematological

317   malignancies than solid tumors, there is increasing evidence of its relevance in diseases like

318   AML[36–40]. There was no significant correlation of module expression with any tested EMT gene

319   sets, suggesting that F1 module 1 and HL60 module 2 describe cell adhesion and motility

320   processes largely independent from canonical EMT (**Figure S4D**). The co-upregulation and

321   significant heritability of these genes in both cell lines under AraC treatment supports the pro-

322   survival role for cytoskeletal dynamics separate from EMT processes, potentially constituting a

323   form of CAM-DR. This is further supported by cell adhesion molecule genes present in the HL60

324   "resistance" module, including ITGB1, a primary mediator of CAM-DR[41] and LGALS1, implicated

325   in CAM-DR in multiple myeloma [42].

326

327   To identify whether heritable upregulation of CAM-DR genes in leukemic lineages prior to

328   treatment was associated with the generation of AraC resistance, we examined related

329   pathways in pre-experiment and control H1 cells. Sixty-five genes significantly upregulated in

330   treatment-naive H1 cells were also found in the HL60 population tree "resistance" module,

331   constituting an H1-specific pre-resistant state (**Figure S4E**).The majority of these genes are

332   also involved in cell motility and adhesion, such as NCAM1[43], PRICKLE1[44], BMP2[45], CPNE8[46],

333   and NAV3[47]. Also included are ANXA2 and PSD3, genes from the C1498 "persistence" module.

334   While H1 cells uniquely express this pre-resistant signature prior to AraC exposure, treatment

335   increased H1 population expression in a time-dependent manner (**Figure S4F**). In the control

336 setting, this signature does not clearly contribute to cell fitness, but under AraC, it confers a

337 large survival benefit.

Identification of pro-leukemic transcriptional programs associated with in vivo immune escape

340 The mouse M1 "in vivo" module was the most heritable module over the F6 phylogenetic tree; it

341 showed significant enrichment in cells collected from M1 and may represent an in vivo

342 mechanism of survival through AraC treatment (**Figure 4A, S4G**). To determine shared

343 mechanisms with our HL60-FLARE modules, we again compared modules and found a strong

344 Jaccard index score between the M1 "in vivo" and our HL60-FLARE "resistance" modules

345 (**Table S10**, **Figures 4C, S4B-C**). These two modules share immunoproteasome genes

346 PSME1, PSME2, PSMB8, and PSMB9, immunoproteasome components involved in antigen

347 processing and presentation for MHC class I-mediated immune response. PSMB8 and PSMB9

348 encode two of the three beta core subunits of the 20S immunoproteasome core complex[48]. In a

349 variation of the immunoproteasome, the 20S core can complex with the 11S (or PA28ab)

350 regulatory cap, the primary subunits of which are encoded by the PSME1 (PA28a) and PSME2

351 (PA28b) genes[49]. There is evidence that peptide fragments produced by the 20S/11S

352 immunoproteasome may be less conducive to MHC class I presentation and lead to tumor cell

353 immune escape[50–52].

354

355 We investigated Psme1 and Psme2 expression patterns in F6 cells that survived engraftment

356 and treatment in our in vivo model. Differential gene expression analysis comparing F1 and F6

357 in vitro revealed universal F6-specific upregulation of 11S regulatory cap genes Psme1 and

358 Psme2 (**Table S2**). Additionally, Psmb8, Psme1, and Psme2 were significantly upregulated in

359 F6 clones in vivo compared to their in vitro counterparts proportional to clone success in vivo

360 (**Figures 4D, S4H**). F6 clones 1-5 upregulated Psme2 and Psmb8 in the in vivo setting,

361 representing a founder-specific enrichment. Expression of Psme1 differentiated F6 clones

362 further; significant in vivo upregulation was unique to clones F6.1, F6.2, and F6.4, the three

363 most successful clones in the blood (**Figures 4D, S3B**). While Psmb8 was significantly heritable

364 for F6 clones 1-5, Psme2 was only heritable in clones 1-3. In contrast, Psme1 expression was

365 only heritable over the lineage tree within clone F6.5. The evolutionary plasticity of Psme1

366 expression in the F6.1, F6.2, and F6.4 lineages may provide a competitive advantage within the

367 broader F6 immunoproteasome resistance mechanism.

Patient analysis

Phylogenetic reconstruction of in vitro and in vivo AraC-treated C1498-FLARE and HL60-FLARE cells revealed heritable gene modules associated with generating and maintaining AraC resistance. To validate their clinical relevance, we used gene expression data from the TARGET AML cohort (2,007 patients, ages 0-29), the most extensive, publicly available AML patient dataset[53]. The prognostic impact of module expression in bone marrow samples, collected at initial diagnosis, showed significantly increased hazard ratios associated with the expression of HL60-FLARE "resistance" module (HR 4.99), C1498-FLARE "persistence" module (HR 2.69), and "in vivo" module (HR 3.18) (**Figure 5A**). Kaplan Meier curves confirmed that heightened expression of these modules was associated with shorter overall survival (**Figure 5B-D**). These results support the ability of FLARE lineage tracing to generate heritability patterns relevant to AML patient survival.
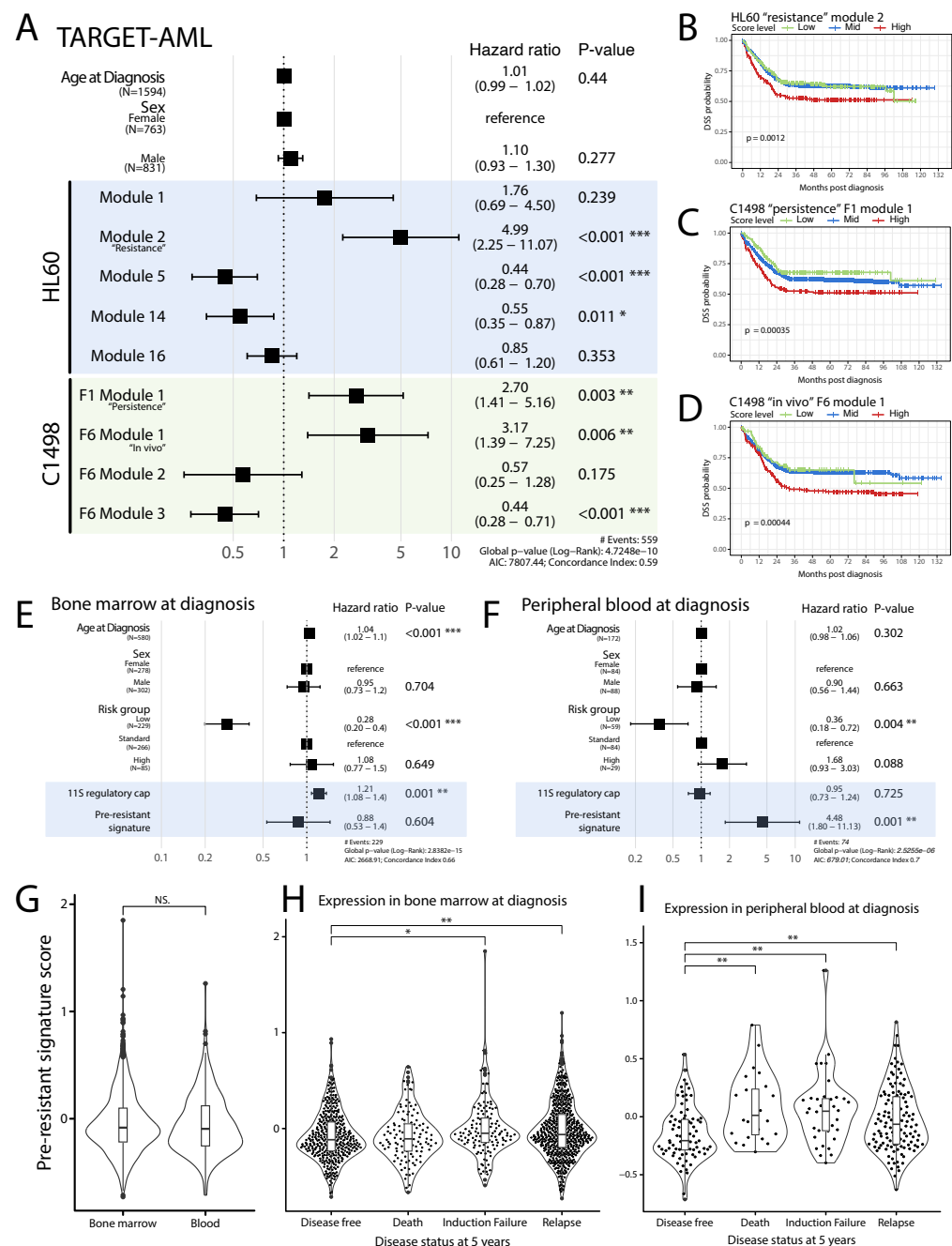
Figure 5: Expression levels of HL60-FLARE modules, C1498-FLARE modules, 11S regulatory cap genes, and pre-resistant signature genes are significantly associated with survival in pediatric AML patients. (**A**) Cox proportional hazard ratios for HL60-FLARE and C1498-FLARE Hotspot module scores calculated from bone marrow samples taken at diagnosis in the TARGET-AML cohort. (**B**) Kaplan Meier survival plots for the TARGET-AML cohort stratified into low, mid, and high expression of HL60-FLARE "resistance" module 2. Corresponding risk table can be found in supplementary table S11. (**C**) Kaplan Meier survival plots for the TARGET-AML cohort stratified into low, mid, and high expression of C1498-FLARE "persistence" F1 module 1. Corresponding risk table can be found in supplementary table S11. (**D**) Kaplan Meier survival plots for the TARGET-AML cohort stratified into low, mid, and high expression of C1498-FLARE "in vivo" F6 module 1. Corresponding risk table can be found in supplementary table S11. (**E**) Cox proportional hazard ratios for 11S regulatory cap and pre-resistant scores calculated from bone marrow samples taken at diagnosis in the TARGET-AML cohort. (**F**) Cox proportional hazard ratios for 11S regulatory cap and pre-resistant scores calculated from peripheral blood samples taken at diagnosis in the TARGET-AML cohort. (**G**) Violin plot showing pre-resistant signature scores calculated from samples taken at diagnosis. Bone marrow and blood samples do not significantly differ in signature scores. (**H**) Violin plot showing pre-resistant signature scores calculated from bone marrow samples taken at diagnosis. Mean scores for patients who experience induction failure or relapse within 5 years post diagnosis are minimally but significantly higher than patients who are disease free 5 years post diagnosis. Significance was calculated with Wilcoxon rank sum tests and Bonferroni corrected. (**I**) Violin plot showing pre-resistant signature scores calculated from peripheral blood samples taken at diagnosis. Mean scores for patients who experience death, induction failure, or relapse within 5 years post diagnosis are significantly higher than patients who are disease free 5 years post diagnosis. Significance was calculated with Wilcoxon rank sum tests and Bonferroni corrected.

402   We then focused on our pre-resistant/CAM-DR sub-module and 11S cap subunits of the

403   20S/11S immunoproteasome variant to explore their roles in treatment response and survival.

404   We generated Cox proportional hazard ratios, including patient risk groups as covariates to

405   account for interactions between risk group and signature expression. Expression levels of both

406   signatures were associated with the risk of death independent of predefined risk groups.

407   Interestingly, signature-associated risk of death differed drastically when patient expression

408   scores were calculated from bone marrow or peripheral blood taken at diagnosis. In bone

409   marrow samples, expression of 11S regulatory cap genes, but not the pre-resistant signature,

410   was significantly associated with the risk of death (HR 1.21, p-value < 0.001, **Figure 5E**).

411   Conversely, in peripheral blood, expression of the pre-resistant signature, but not the 11S

412   regulatory cap genes, was significantly associated with a greater risk of death (HR 4.48, p-value

413   < 0.001, **Figure 5F**). While 11S regulatory cap genes were expressed slightly more in bone

414   marrow samples, potentially explaining why PSME1 and PSME2 expression have higher

415   prognostic impact in bone marrow, there was no significant difference in expression of the pre-

416   resistant sub-module between sample types (**Figure S5A, 5G**).

417

418   We then assessed the relationship between signature expression in pre-treatment diagnosis

419   samples and patient status at 5 years post-inclusion. Specifically, we compared expression

420   levels between patients who experienced death, induction failure, or relapse within 5 years

421   compared to patients who remained disease-free for 5 years. Patients who experienced

422   induction failure or relapse, but not death, tended to have higher pre-resistant sub-module

423   scores in their blood and bone marrow samples at diagnosis (**Figure 5H-I**). Importantly, these

424   trends were more dramatic when comparing blood samples at diagnosis; patients experiencing

425   induction failure, relapse, and death expressed significantly higher levels of the pre-resistant

426   signature in blood compared to patients disease-free at 5 years. There was a slight negative

427   relationship between pre-resistant sub-module score and peripheral blast percentage at

428   diagnosis, suggesting this pattern does not reflect peripheral leukemic burden (**Figure S5B-C**).

429   Discussion

430   Chemotherapy resistance likely arises from the selection of pre-existing resistant populations

431   and the reprogramming of previously sensitive cells. There is an unmet need for tools to dissect

432   this evolution over time. We developed FLARE, a CRISPR-based dynamic lineage tracing

433   system capable of tracking founder, clone, and subclone lineage in tandem with single-cell

434   transcriptomics in response to chemotherapy. We demonstrated how our hierarchical approach

435   can record cancer transcriptional evolution over an extended treatment time course, capturing

436   both clonal evolution and selection of pre-existing cellular states. We used this technology to

437   dissect the transcriptional evolution of AML, leveraging both in vitro and vivo systems and

438   mouse and human models.

439

440   Using FLARE we uncovered the (1) heritable patterns of cell adhesion-associated AraC-

441   tolerance and the (2) upregulation of immunoproteasome 11S regulatory cap subunits

442   contributing to clone-specific in vivo fitness. The expression of both signatures in AML patients

443   is significantly associated with shorter overall survival. High expression of PSME1 and PSME2,

444   the major subunits of the 11S regulatory cap, in bone marrow but not peripheral blood, is

445   associated with poorer prognosis. In contrast, high expression of our pre-resistant signature,

446   which consists of various genes involved in cytoskeletal and cell adhesion pathways, is more

447   informative when measured in blood samples taken at diagnosis. Both the 11S regulatory cap

448   and pre-resistant signatures are composed of many genes not previously implicated in AML.

449

450   Our pre-resistant signature comprises many cell adhesion-mediated drug resistance (CAM-DR)

451   genes. This mechanism has long been studied in the context of solid tumors, but its importance

452   in hematological malignancies is growing[54,55]. Leukemic CAM-DR is thought to result from

453   interactions between AML blasts and bone marrow stromal cells via various integrin

454   molecules[56]. These interactions may protect leukemia cells, leading to minimum residual

455   disease persistence and treatment failure upon disease recurrence[55]. This increased survival is

456   thought to result from anoikis resistance, in which cells avoid apoptosis in response to becoming

457 detached from the extracellular matrix[57]. Interestingly, the emergence of CAM-DR in our work

458 was primarily seen in cells treated with AraC in standard in vitro cell culture. This may indicate

459 that treatment alone can lead to such resistance outside of the bone marrow niche.

460 Furthermore, survival analysis of the TARGET AML cohort revealed that expression of our

461 potential CAM-DR-associated pre-resistant signature had a significantly higher prognostic

462 impact when measured from peripheral blood samples taken at diagnosis. Our survival analysis

463 also suggests that the presence of leukemic blasts with enrichment of cytoskeletal, cell

464 adhesion, and cell motility processes in the blood before induction therapy is associated with

465 induction failure, death and relapse within 5 years of diagnosis. Therefore, patients with "pre-

466 resistant" lineages prior to treatment are less likely to benefit from induction chemotherapy, with

467 cells already displaying CAM-DR that can survive treatment and expand to outcompete other

468 leukemic lineages. This mechanism needs further exploration, but it suggests a secondary role

469 of CAM-DR in AML outside of the bone marrow niche.

470

471 While the exact pro-leukemic mechanism of our 11S signature cap is still unclear, there are data

472 to suggest the 11S regulatory cap leads to the generation of fewer peptides conducive to MHC

473 class I presentation and can interfere with tumor-specific epitope production[51,52,58]. 11S subunit

474 expression levels in pediatric AML patient bone marrow samples taken at diagnosis are

475 significantly associated with an increased risk of death and shorter overall survival. PSME1 and

476 PSME2 expression levels in bone marrow samples differed negligibly between patients with

477 adverse outcomes within 5 years compared to those that remained disease-free. This indicates

478 that while PSME1 and PSME2 expression in bone marrow is associated with a higher risk of

479 death, it is not highly predictive of induction failure or relapse within 5 years of diagnosis. It may

480 instead indicate the presence of leukemic lineages that are better able to evade immune

481 detection, resulting in a more aggressive and "colder" immune microenvironment.

482

483 In the future, the FLARE approach could be used to dissect the transcriptional evolution of other

484 tumor types under selective treatment pressure. Given the ability to sample tumor evolution

485 from tagged cell populations in circulating blood, this can be multiplexed to explore many more

486 initial cell states and treatment spaces. Moreover, this approach can be coupled to mouse

487 lineage-tracing models such as DARLIN[59] or the MARC1[60] mouse to capture the interplay

488 between engrafted tumor and immune cell populations in the context of checkpoint inhibitors

489 and other immunotherapies.

Methods

Cell Lines and drugs

The murine C1498 cell line (Cat# TIB-49) was purchased from ATCC in 2021. The HL60 AML cell line was a gift from Dr. Bonnie Lau's lab. C1498 cells were cultured in DMEM, supplemented with 10% FBS, 100 u/mL penicillin, and 100 ug/mL streptomycin. HL60 cells were cultured in Iscove's Modified Dulbecco's Medium, supplemented with 20% FBS, 100u/mL penicillin, and 100ug/mL streptomycin. All cell lines were expanded, aliquoted, and cryopreserved in culture media plus 5% DMSO and stored in liquid nitrogen upon receival. Active cell cultures were stored at 37°C with 5% CO_2. All cultured lines were regularly checked for the presence of mycoplasma using the Universal Mycoplasma Detection Kit (ATCC, 30-1012K). Cytarabine (AraC) powder was purchased from Selleck Chemicals (Cat # S1648). AraC was resuspended in PBS fresh for each in vitro and in vivo treatment.

Mice

KH2/iCas9 mice (Jackson Laboratory strain #029415) were a gift of the Huang lab at Dartmouth College. All mice were maintained in pathogen-free housing at the Center for Comparative Medicine and Research in Dartmouth Hitchcock Medical Center. All experimental protocols were approved by and performed in accordance with the guidelines of the Institutional Animal Care and Use Committee at Dartmouth College. To attenuate the immune response to Cas9, all injected KH2/iCas9 were exposed to doxycycline during in utero development through irradiated feed containing 625ppm doxycycline (ScottPharma, Inc. Marlborough, MA).

Molecular cloning

GESTALT lineage tracing system engineering

All digested vectors and fragments and PCR products were purified with the Monarch Gel Purification Kit (Cat# T1120S) or the Zymo Clean and Concentrate-5 Kit (Cat# D4003). All Gibson assembly reactions were conducted with NEB HiFi DNA Assembly 2x Master Mix (Cat# E2621S), following the ssDNA HiFi assembly protocol. Golden Gate assemblies were conducted following the Rice University Esp3I protocol. Chemical bacterial transformations were done with DH5alpha competent cells following the NEB Transformation Protocol with a 2-hour recovery at 37°C. Electroporated transformations were conducted with NEB Electrocompetent E. coli and were electroporated with the following conditions: 2kV, 200Ohms, 25uFd. All

520   bacterial cultures were grown in a 37°C shaker at 220-290rpm overnight. Vectors were purified

521   using either the Qiagen Miniprep Kit or the Zymo Midiprep II Kit. Correct cloning of all final

522   constructs was confirmed via Circuit-seq[61].

523   **FLARE-miniV10-10x-UMI**

524   Eight Cas9 target sequences were designed for high efficiency (>69% Doench) and low off

525   target scores. The eight-target array was purchased as a gene block from IDT, with upstream

526   Esp3I sites for founder tag insertion and NdeI + BamHI sites for cloning purposes. The target

527   gBlock and pLARRY-EGFP vector (Addgene #140025) were digested with NdeI+BamHI and

528   ligated together with Quick Ligase (Cat# M2200s). Resulting construct is referred to as

529   pLARRY-targets.

530

531   The PB-CMV-MCS-EF1-puro vector (Systems Bio cat #PB510B-1) was digested with SpeI and

532   HincII, removing the puromycin resistance gene, and then circularized with a Gibson assembly

533   reaction with a short single stranded joining oligo. The resulting construct is referred to as

534   miniPB.

535

536   The FLARE-miniV10-10x vector was then constructed by amplifying EGFP and the target array

537   from pLARRY-targets. The PCR primers used KpnI and EcoRI sites. The amplified

538   EGFP+target array and miniPB were digested with KpnI and EcoRI, gel purified, and ligated

539   together using T7 ligase. The 10X capture sequence 1 was incorporated via Gibson assembly.

540   The FLARE-miniV10-10x construct is available at Addgene (Plasmid #234492).

541

542   Founder tag oligos consisting of 14 degenerate bases (the founder tag) and Esp3I sites were

543   purchased from IDT. One cycle of PCR was used to anneal and extend the oligos. Founder tag

544   oligos were then inserted into FLARE-miniV10 via a Golden Gate Assembly reaction; the

545   resulting product is FLARE-miniV10-10X-UMI.

546   **FLARE-pbV10g-neo**

547   Individual Cas9 guides complementary to the target sequences in FLARE-miniV10 were

548   designed and cloned into the pX330-mCherry (Addgene #98750) vector using the Zheng Lab

549   pX330 cloning protocol. Guides, along with their individual U6 promoters, were amplified from

550   pX330-mCherry vectors. A secondary PCR reaction added flanking Esp3I sites to each U6-

551   guide fragment. Esp3I sites were also added to the PB-CMV-MCS-EF1-puro vector, while the

552   puro resistance gene was removed. U6/guide sequences were joined and inserted into PB-

553   CMV-MCS-EF1-puro via a 9-part Golden Gate Assembly reaction. The neomycin resistance

554   gene was amplified from the hCas9 vector and inserted into the U6/guide array vector via

555   Golden Gate Assembly. The FLARE-pbV10g-neo construct is available at Addgene (Plasmid

556   #234493).

557

558   In earlier versions of FLARE constructs, Cas9 guides were encoded into a single construct

559   along with the lineage recorder Cas9 target array. For reasons not yet fully understood, splitting

560   this into two vectors (FLARE-miniV10-10X and FLARE-pbV10g-neo) led to significantly higher

561   editing rates across targets.

562   **FLARE-pbCas9-tag-mCherry**

563   The miniPB construct was digested with KpnI for a Gibson Assembly insertion of a landing pad

564   containing a NotI and a new KpnI site. The resulting vector is referred to as miniPB-LP. The

565   pX330 construct was digested with KpnI and NotI to isolate the Cas9 transgene, which was then

566   ligated into the landing pad miniPB-LP vector. The resulting construct is referred to as miniPB-

567   Cas9.

568   The pX330-mCherry construct was digested with BamHI and HindIII for the insertion of a clone

569   tag landing pad containing type IIs restriction sites in the 3' UTR of the mCherry transgene using

570   Gibson Assembly. The mCherry transgene, including the clone tag landing pad, was amplified

571   and extended to include Gibson Assembly adaptors complementary to the insertion region in

572   miniPB-Cas9. The miniPB-Cas9 construct was then digested with NotI and the mCherry and

573   clone tag landing pad were inserted using Quick Ligase. The resulting construct is referred to as

574   FLARE-pbCas9-mCherry, and is available at Addgene (Plasmid #234494).

575   Lastly, clone tag oligos with Esp3I sites were purchased from IDT. For C1498-FLARE cells, the

576   clone tag consisted of 14 degenerate bases followed by 10x capture sequence 2. For HL60-

577   FLARE cells, the clone tag consisted of 38 patterned bases

578   (NNNWSNNNWSNNNWSNNNWSNNNWSNNNWSNNNWSNNN, pattern from Johnson et al [62])

579   followed by 10x capture sequence 1. One cycle of PCR was used to anneal and extend the

580   oligos, and the resulting product was inserted into FLARE-pbCas9-mCherry via Golden Gate

581   Assembly. Capture sequence 2 was replaced by capture sequence 1 for HL60-FLARE cells due

582   to the planned discontinuation of support for capture sequence 2 announced by 10x.

583   Cell engineering and lineage tracing system testing

584   We used the Neon Electroporation System to integrate all piggyBac constructs into cells to
585   generate stable cell lines. Electroporations were carried out with the following protocols: C1498
586   cells were resuspended at 2x10^7 cells/mL and electroporated with 1080V, 50ms, and 1 pulse;
587   HL-60 cells were resuspended at 2x10^7 cells/mL and electroporated with 1450V, 35ms, and 1
588   pulse. All electroporations were done with the Neon 100ul tip system. Transposase:transposon
589   ratios for piggyBac integration ranged from 1:1 to 1:8. All FACS was conducted on a BD
590   FACSAria II or BD FACSAria Fusion. For antibiotic selection cells were incubated with selection
591   media for 1-3 weeks, and dead cells were removed from culture 2-3 times/week using a Ficoll
592   density gradient.
593
594   GESTALT constructs were introduced to cell lines in the following order: first low passage cells
595   were electroporated with FLARE-miniV10-10X-UMI and FLARE-pbV10g-neo. EGFP-positive
596   and neomycin resistant cells were isolated via FACS and 1ug/mL G418 selection. This step was
597   repeated two to three more times, gating for higher EGFP expression during FACS in each
598   subsequent round. For the third round of FACS for the C1498 cell line, individual cells were
599   sorted into wells of a 96-well plate to initiate individual clonal populations.
600
601   Prior to lineage tracing experiments, engineered lineage tracing cells were electroporated with
602   FLARE-pbCas9-tag-mCherry and mCherry+ cells were isolated via FACS. Cells were sorted
603   into bulk populations and 96-well plates at 100 cells/well.
604
605   Lineage recorder editing levels in C1498-FLARE cells over time were calculated using single
606   cell and bulk lineage barcode data. Day 0 editing was calculated from bulk (not a subset of
607   founders) C1498-FLARE cells prior to FLARE-pbCas9-tag-mCherry integration; day 13 editing
608   was calculated from this same population 13 days following integration post FACS. The
609   remaining editing rate timepoints were measured using the cell population used in C1498-
610   FLARE founder pool AraC treatment experiment and lineage recorder reads were subset for
611   reads containing valid founder tags associated with founders 1, 5, 6, and 11. Day 44 editing was
612   calculated from single cell sequencing of the founder pool subset for F1 and F6 prior to AraC
613   treatment. Day 64 editing was calculated from single cell sequencing of one untreated flask
614   cultured from the same population as used in day 44, and was subset for valid cells identified
615   via GEX analysis. Day 87 was also measured from one control flask from this population -
616   editing was calculated from single cell lineage endpoint data.

C1498-FLARE founder isolation and characterization

617

618 To screen individual C1498-FLARE founder populations for copies of FLARE-miniV10-10X-UMI

619 integrations gDNA was collected and founder tags were amplified and prepared into sequencing

620 libraries. Founder tag libraries for individual populations were multiplexed and sequenced on an

621 Illumina MiniSeq with a Mid-Ouput 300-cycle kit. Founder tags for each clone were determined

622 as having > 4,300 reads. Cells from each clone were expanded, aliquoted, and cryopreserved.

623 **Growth curves**

624 To begin growth curves for individual founders, cells were resuspended at 50k cells/mL and

625 plated in T25 flasks in triplicate. From day 0-7, two 10ul aliquots from each flask were mixed

626 with 10ul Trypan Blue and counted on the Luna II Cell Counter daily. Growth curve data were

627 analyzed and plotting using the Growthcurver R package [63].

628 **AraC dose-response curves**

629 Cells were plated at 15,000 cells/well in a 96-well plate in media containing various AraC

630 dosages in triplicate (0.0001mg/ml-10mg/ml). 100ul of CellTiter 96 AQueous One Solution (Cat#

631 G3582) was added to each well and after incubation for 1-3 hours absorbance was measured at

632 490nm on a SpectraMax i3x. Absorbance from wells containing media without cells or drug

633 were averaged to calculate background absorbance. Corrected absorbance for each well was

634 calculated as (well absorbance - background absorbance). Corrected absorbances were then

635 loaded into the drc R package for generating dose-response curves, ED50 values, and EC50

636 values[64].

637 **Whole genome sequencing**

638 High molecular weight (HMW) gDNA was isolated from parental C1498 cells and individually

639 engineered clones using the Monarch High Molecular Weight gDNA Kit for Cells (Cat# T3050S).

640 HMW gDNA samples were multiplexed and prepared into a sequencing library following the

641 Oxford Nanopore SQK-NBD114.24 kit. The library was then sequenced on a PromethION flow

642 cell with R10.4.1 chemistry. Sequencing data were basecalled and demultiplexed using Guppy.

643 Variants were called using PEPPER-Margin-DeepVariant, SNPs were isolated and filtered for a

644 minimum quality of 20 and mean depth between 10 and 80. To remove population variants, the

645 GRCm39 dbSNP was annotated using Ensembl VEP and variants with moderate or high impact

646 were removed. The resulting dbSNP was subtracted from filtered SNPs using Bedtools [65].

647 Remaining variants were annotated with Ensembl VEP and intersected with datasets of known

648 AML mutations.

AraC response and resistance experiments

C1498-FLARE founder pool treatment response and resistance

C1498-FLARE founders F1, F5, F6, and F11 were sorted by FACS on day 13 following electroporation with the FLARE-pbCas9-tag-mCherry vector. Cells were sorted into 96-well plates at 100 cells/well. In vitro AraC treatment began on day 44 (31 days post FACS). Cells were allowed to grow for a longer period following sorting due to growth rate differences between founders. On day 44, 1 well of sorted cells for each clone was pooled at equal numbers and split into four flasks for in vitro treatment with an aliquot set aside for mouse injections.

In vitro

Two flasks were maintained in normal culture media, the remaining two flasks were maintained in media containing 0.05ug/ml AraC. Flasks were maintained for 43 days to allow for the development of AraC resistance. Media was replaced 2-3 times/week and split when needed. Dose-response curves were calculated for each flask on day 15 of treatment, and a final dose-response curve was calculated for flask displaying resistance on day 36 of treatment. On day 43 of treatment, cells were prepared for MULTIseq tagging and single cell capture following MULTIseq and 10x protocols.

In vivo

Pooled C1498 clone cells were washed twice with PBS, resuspended at 250,000 cells/mL and filtered through a 40uM filter. Mice were injected IV with 100ul of cell suspension (25,000 cells/mouse). On day 14 mice were randomly assigned to treatment groups; on days 14-16 treatment mice received 100mg/kg AraC IP and control mice received an equal volume of PBS. Blood was collected via the retro-orbital sinus on days 13, 28, 43, 50, and 57. Prior to blood draws mice were anesthetized with Isoflurane. Microhematocrit capillary tubes were placed in 1.7mL tubes containing 50mM EDTA and then used to collect 50-200ul of blood from the retro-orbital sinus into EDTA coated tubes. Blood samples were then kept on ice until further processing. Mice displaying symptoms of high AML burden were euthanized. All surviving mice were euthanized on day 67 post cell injections. Upon euthanasia whole blood was collected via cardiac puncture and femoral bone marrow was harvested. gDNA was isolated from 20ul of whole blood, and remaining blood was separated using a Ficoll-paque density gradient to isolate PBMCs. To isolate bone marrow, 18g needles were used to puncture the bottom of 0.5mL tubes, femur bones were placed in the 0.5mL tubes and nested in 1.7mL tubes, and

681   tubes were spun at > 10,000g for 20 seconds. Bone marrow pellets and PBMCs were methanol

682   fixed and permeabilized following 10x protocols.

683   **Human AML cell line in vitro treatment response**

684   Engineered lineage tracing HL60-FLARE cells were sorted into 96-well plates at 100 cells/well

685   following electroporation with FLARE-pbCas9-tag-mCherry to establish GESTALT lineage

686   tracing populations, as described above. Sorted well populations were expanded for 23 days.

687   On day 23 post FACS, five expanded wells were combined into an initial cell pool. The pool was

688   then split for 10x scRNA-seq preparation, methanol fixation, cryopreservation, and in vitro AraC

689   treatment. For in vitro AraC treatment cell pools were split into control (no AraC) and AraC

690   conditions (0.01ug/mL AraC). All treatment wells were cultured in duplicates. AraC solutions and

691   media were made fresh for media changes twice per week, and wells were split as needed. Cell

692   concentration for each well was measured once per week in duplicates using the Luna II Cell

693   Counter. Cells were regularly cryopreserved through the experiment for later analysis. In vitro

694   treatment continued for 89 days, after which cells were prepared for 10x scRNA-seq preparation

695   with MULTIseq multiplexing.

696   ## Lineage library preparation and sequencing

697   All PCR reactions were conducted using Q5 HotStart Polymerase (Cat # M0494L) and Sybr

698   Green dye. DNA was purified from PCR reactions using Omega TotalPure NGS magnetic

699   beads with 0.8-1.5x ratios based on amplicon size.

700   **Bulk DNA library preparation and sequencing**

701   gDNA was collected from cultured cells, PBMCs, and bone marrow samples using the Qiagen

702   DNeasy Blood and Tissue Kit (Cat# 69504). For library preparation, lineage recorders (including

703   founder tags and clone tags) were first tagged with sequencing UMIs prior to amplification to

704   control for PCR bias, with the exception of libraries constructed from blood collections. Blood

705   collection libraries were treated as sensitive samples, as C1498-FLARE cells were sparse in the

706   blood. As a result, these libraries were amplified without sequencing UMIs and initial PCRs were

707   set up in a PCR hood to prevent contamination. For all libraries, TruSeq/Nextera indices and

708   Illumina adaptors were subsequently attached with PCR. Final sequencing libraries were

709   quantified with the Qubit fluorometer and the NEB NextQuant Library Quant kit (Cat # E7630L).

710   Libraries were then diluted to the appropriate loading concentration for sequencing.

711   **Single cell library preparation and sequencing**

712 Cells were captured for single cell sequencing following 10x guidelines for 3' scRNA-seq with

713 feature barcoding. For multiplexed single cell experiments, cells were first tagged following

714 MULTIseq protocols (Millipore Sigma, LMO001)[66]. Gene expression and MULTIseq libraries

715 were constructed and quantified following 10x and MULTIseq guidelines with one alteration - a

716 portion of cDNA was set aside following initial cDNA amplification but prior to SPRIselect

717 cleanup to be used for FLARE lineage library construction. This was done to avoid splitting

718 FLARE transcripts between GEX and feature barcoding fractions.

719 The previously set aside cDNA was purified using a 2x bead cleanup. For each sample, cDNA

720 was split to generate a library of dynamic lineage recorders (which includes founder tags) and a

721 library of clone tags. All single cell FLARE libraries were initially amplified in 9-16 replicate wells

722 and then combined into 3-4 separate sub-libraries. TruSeq/Nextera indices and Illumina

723 adaptors were subsequently attached with PCR. Final sequencing libraries were quantified with

724 the Qubit fluorometer and the NEB NextQuant Library Quant kit (Cat # E7630L). Libraries were

725 then diluted to the appropriate loading concentration for sequencing.

Computational analysis

726

Sequencing data pre-processing

727

728 For bulk DNA sequencing raw data were converted to fastq format using bcl2fastq2 or BCL

729 Convert (Illumina). For single cell RNA sequencing raw data were converted to fastq format

730 using CellRanger mkfastq. Cells were then called using CellRanger count for non-multiplexed

731 10x runs or CellRanger multi for MULTIseq multiplexed 10x runs[67].

Single cell RNA-seq analysis

732

733 Cellranger output filtered matrices were analyzed with Seurat 5.1.0[68]. Low quality cells were

734 filtered out based on the number of RNA features and percentage of mitochondrial reads. For

735 non-multiplexed samples doublets were identified and removed using the scDblFinder R

736 package[69]. For multiplexed samples, MULTIseq barcode filtered matrices were added as Assay

737 Objects to Seurat objects. Cells were demultiplexed and doublets were identified and removed

738 with the *MULTIseqDemux()* function. For PBMC and bone marrow samples cell type

739 annotations were added using the SingleR R package with the MouseRNAseqData reference

740 set from the celldex R package[70]. All single cell Seurat objects were normalized with the

741 SCTransform R package for data visualization and integration with other objects[71,72]. Differential

742   gene expression analysis was conducted using the *FindMarkers()* and *FindAllMarkers()* Seurat

743   functions.

Gene set enrichment analysis

745   C1498-FLARE founder-specific gene set enrichment analysis was conducted with the fGSEA R

746   package[73]. Differentially expressed genes were identified for each founder by Wilcoxon rank

747   sum test and input into the fGSEA function with Mus musculus Hallmark gene sets from

748   mSigDB.

Single cell lineage analysis and tree reconstruction

750   *Founder and clone tag clustering*

751   To map founder tags to founder tag groups (or founders) and clone tags to clone tag groups (or

752   clones), we used a clustering strategy based on per cell tag set similarity. Briefly, given a

753   mapping of cells to a set of tags, a graph was constructed with cells as nodes and Jaccard

754   similarity as edges. Jaccard similarity was calculated pairwise as the intersection between per

755   cell tag sets divided by the union of per cell tag sets. Any pairs of cells with a Jaccard similarity

756   of 0 (indicating no overlapping tags) shared no edges. Cells were assigned to clusters using the

757   Louvain algorithm for community detection which returned the partition that maximized

758   modularity and with resolution = 1.

759   *Founder assignment*

760   To extract lineage information from scRNA-seq-captured dynamic lineage recorders, we first

761   used SingleCellLineage, as previously described[13,14]. Briefly, sequences are collapsed by 10x

762   UMI and aligned to the reference to obtain a pattern of insertions and deletions across a target

763   array per cell. Founder tags, located immediately upstream of the lineage recorders, were

764   extracted per collapsed merged read. We filtered cell barcodes against the 10x feature

765   barcoding Cell ID whitelist. Merged reads with less than 85% identity to the reference sequence

766   were removed from the dataset. To ensure we were using high quality barcodes to build single-

767   cell lineage trees, we leveraged tools from the Cassiopeia pipeline[74] to perform additional data

768   cleaning. Founder tags were corrected for PCR errors via Starcode[75] with Levenshtein distance

769   = 1. UMIs were error corrected using Cassiopeia function *error_correct_umi()*. Finally, we

770   filtered our error corrected founder tags and target arrays using *filter_molecule_table()* to

771   remove low read count UMIs and cells with too few UMIs. Additionally, this step ensures there is

772   only one target array per 10x cell ID-founder tag combination.

773

774   Founders were assigned by the combination of founder tag integrations in cells. For C1498-

775   FLARE founder populations, founder tag assignments were previously determined via sub-

776   founder culturing and bulk gDNA sequencing. Therefore, if cells had reads mapping to more

777   than one founder based on the founder tags found in its lineage data, it was assigned to a

778   founder based on the total sum of reads and UMIs associated with a certain founder group.

779   Only cells with 90% identity to a certain founder were kept.

780

781   Unlike the C1498-FLARE founders, HL60-FLARE founders had no ground truth mapping of

782   founder tags to founder groups. Subsequently, we took additional steps to ensure we had a

783   confident list of founder tags per cell and group. To remove within-cell background noise, we

784   removed integrations that made up less than 1.5% of reads per cell. This threshold was based

785   on qualitative inspection of the distribution of integration read counts per integration per cell.

786   The founder tags were further filtered by removing tags found in fewer than 5 cells across the

787   dataset. The cleaned set of tags were clustered into founders as described previously (*Founder*

788   *and clone tag clustering*). Finally, to refine the final set of founder tags defining a founder, we

789   filtered for tags with greater than 50% identity to a single founder group. This threshold was

790   based on visual inspection of the distribution of tag identity to founders, determined by the total

791   number of cells with a tag in a founder divided by the total number of cells with a tag. This

792   additionally ensures that each founder tag is found in one founder.

793   *Clone assignment*

794   Clone tags were extracted from sequencing reads using the SingleCellLineage pipeline. For

795   HL60-FLARE clone tag reads, only tags matching the expected 38bp IUPAC pattern were kept.

796   Clone tag reads with more than one missing base within the tag sequence window were

797   removed from the dataset, then clone tags were error corrected using Starcode with

798   Levenshtein distance = 1 for 14 bp tags or Levenshtein distance = 2 for 38 bp tags. These clone

799   tags were then filtered for read count > 1 and number of UMIs > 1. Tags with nucleotide repeats

800   of 5 or longer were also removed (i.e. AAAAA, TTTTT, GGGGG, CCCCC). Background tags,

801   defined as contributing to fewer than 10% of all reads in the cell, were filtered out. Finally, clone

802   tags were discretized to founders by keeping tags with 90% identity (C1498s, 2 founders) or

803   50% identity (HL60s, 15 founders) to a given founder based on the total proportion of reads per

804   clone tags. C1498-FLARE and HL60-FLARE cells were assigned to clones as previously

805   described (*Founder and clone tag clustering*) given the final set of clone tags.

806   *Building Trees*

807   Trees were reconstructed per clone using Cassiopeia's[74] greedy vanilla solver from the pattern

808   of indels per target array. The method was chosen for its scalability given the size of our trees

809   ranges from tens to thousands of cells. Briefly, a tree is built in a top-down fashion by splitting

810   cells into two groups based on the most frequent mutation in the population. Cell subsets are

811   repeatedly split until only one cell remains per subset. Clones belonging to the same founder

812   population were combined into founder trees using a custom python script to attach trees at the

813   root, due to the hierarchical nature of the lineage tracing tag system. All founder trees were

814   combined with the same method to generate a population tree. Cells with unedited/wildtype

815   target arrays were removed from the analysis prior to building the trees.

816   *PATH analysis*

817   To identify genes that were significantly heritable in our clonal population we used the recently

818   published tool, PATH [28]. PATH takes as input a phylogenetic tree along with either discrete

819   annotations like cluster assignments or continuous data like gene expression or module scores.

820   We considered the top 3000 highly variable features per tree to identify genes with strong

821   autocorrelations, suggesting high heritability.

822   *Hotspot module generation*

823   To identify informative gene expression modules based on our single-cell lineage trees and

824   gene expression data, we leveraged the tool Hotspot[34] as described in the following vignette

825   https://hotspot.readthedocs.io/en/latest/Lineage_Tutorial.html. Briefly, we ran Hotspot on our

826   log-normalized data via the 'normal' model filtered for genes in a minimum of 10 cells. Then we

827   generated the KNN graph with 10 neighbors. Autocorrelation scores are calculated per gene to

828   perform feature selection. These genes are then correlated to generate a gene-by-gene matrix,

829   which is used to identify modules using a hierarchical clustering approach with

830   min_gene_threshold = 10 to control the size of modules being joined.

831   *Module overlap analysis*

832   C1498-FLARE module mouse genes were converted to human orthologs for analysis. C1498-

833   FLARE and HL60-FLARE modules were assessed for significant overlap as measured by

834   Jaccard indices and odds ratios compared to genomic background using the GeneOverlap R

835   package[76].

836   **Survival analysis**

837 TARGET-AML RNAseq, clinical, and sample data were accessed through the cBioPortal.

838 Expression datasets were split by sample type where applicable. Sample expression scores for

839 modules and gene sets were calculated by averaging the FPKM or RPKM scores. Cox

840 proportional hazard ratios were generated using the survival R package[77]. To stratify patients for

841 Kaplan-Meier analysis, we first determined the expression score mean and standard deviation

842 within the samples of interest. Patients with scores within one standard deviation of the mean

843 were assigned to the Mid-expression group. Patients with scores above and below one standard

844 deviation of the mean were assigned to the High and Low expression groups, respectively. We

845 excluded patient samples missing information for tissue type.

Acknowledgments

Data availability

865 All sequencing data associated with this publication have been deposited in NCBI's Gene

866 Expression Omnibus[78] and are accessible through GEO Series accession numbers GSE289454

867 and GSE289459.

References

1. Boyd, A. L. *et al.* Identification of Chemotherapy-Induced Leukemic-Regenerating Cells Reveals a Transient Vulnerability of Human AML Recurrence. *Cancer Cell* **34**, 483-498.e5 (2018).

2. Stetson, L. C. *et al.* Single cell RNA sequencing of AML initiating cells reveals RNA-based evolution during disease progression. *Leukemia* **35**, 2799–2812 (2021).

3. Morita, K. *et al.* Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics. *Nature Communications 2020 11:1* **11**, 1–17 (2020).

4. Ojamies, P. N. *et al.* Monitoring therapy responses at the leukemic subclone level by ultra-deep amplicon resequencing in acute myeloid leukemia. *Leukemia 2017 31:5* **31**, 1048–1058 (2016).

5. Docking, T. R. *et al.* A clinical transcriptome approach to patient stratification and therapy selection in acute myeloid leukemia. *Nat. Commun.* **12**, 2474 (2021).

6. Severens, J. F. *et al.* Mapping AML heterogeneity - multi-cohort transcriptomic analysis identifies novel clusters and divergent ex-vivo drug responses. *Leukemia* **38**, 751–761 (2024).

7. Stratmann, S. *et al.* Transcriptomic analysis reveals proinflammatory signatures associated with acute myeloid leukemia progression. *Blood Adv.* **6**, 152–164 (2022).

8. Thol, F. & Ganser, A. Treatment of Relapsed Acute Myeloid Leukemia. *Current Treatment Options in Oncology* vol. 21 Preprint at https://doi.org/10.1007/s11864-020-00765-5 (2020).

9. Yan, B. *et al.* Low-frequency TP53 hotspot mutation contributes to chemoresistance through clonal expansion in acute myeloid leukemia. *Leukemia* **34**, 1816–1827 (2020).

10. Sharma, S. V. *et al.* A Chromatin-Mediated Reversible Drug-Tolerant State in Cancer Cell Subpopulations. *Cell* **141**, 69–80 (2010).

11. Shaffer, S. M. *et al.* Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**, 431–435 (2017).

12. Caiado, F. *et al.* Lineage tracing of acute myeloid leukemia reveals the impact of hypomethylating agents on chemoresistance selection. *Nat. Commun.* **10**, 1–15 (2019).

13. McKenna, A. *et al.* Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).

14. Simeonov, K., Mckenna, A., Shendure, J. & Legner, C. Single-cell lineage and transcriptome reconstruction of metastatic cancer reveals selection of aggressive hybrid EMT states. 1–36 (2020).

15. Quinn, J. J. *et al.* Continuous lineage recording reveals rapid, multidirectional metastasis in a lung cancer xenograft model in mouse. *bioRxiv* 2020.04.16.045245 (2020).

16. Driss, V. *et al.* Sub-clonal analysis of the murine C1498 acute myeloid leukaemia cell line reveals genomic and immunogenic diversity. *Immunol. Lett.* **192**, 27–34 (2017).

17. de la Cruz López, K. G., Toledo Guzmán, M. E., Sánchez, E. O. & García Carrancá, A. MTORC1 as a regulator of mitochondrial functions and a therapeutic target in cancer. *Front. Oncol.* **9**, 1373 (2019).

18. Zeng, Z. *et al.* Targeting of mTORC1/2 by the mTOR kinase inhibitor PP242 induces apoptosis in AML cells under conditions mimicking the bone marrow microenvironment. *Blood* **120**, 2679–2689 (2012).

19. Schulze, A., Oshi, M., Endo, I. & Takabe, K. MYC targets scores are associated with cancer aggressiveness and poor survival in ER-positive primary and metastatic breast cancer. *Int. J. Mol. Sci.* **21**, 8127 (2020).

20. Donati, G. *et al.* Targeting mitochondrial respiration and the BCL2 family in high-grade MYC-associated B-cell lymphoma. *Mol. Oncol.* **16**, 1132–1152 (2022).

21. Quan, Q., Xiong, X., Wu, S. & Yu, M. Identification of immune-related key genes in Ovarian Cancer based on WGCNA. *Front. Genet.* **12**, 760225 (2021).

22. Xiao, X.-Y. *et al.* TRAT1 overexpression delays cancer progression and is associated with immune infiltration in lung adenocarcinoma. *Front. Oncol.* **12**, 960866 (2022).

23. Lee, H. *et al.* A novel immune prognostic index for stratification of high-risk patients with early breast cancer. *Sci. Rep.* **11**, 128 (2021).

24. Xie, W. *et al.* Identification of low-density lipoprotein receptor class A domain containing 4 (LDLRAD4) as a prognostic indicator in primary gastrointestinal stromal tumors. *Curr. Probl. Cancer* **44**, 100593 (2020).

25. Liu, Z. *et al.* Low density lipoprotein receptor class A domain containing 4 (LDLRAD4) promotes tumorigenesis of hepatic cancer cells. *Exp. Cell Res.* **360**, 189–198 (2017).

26. Ou, R. *et al.* Transcriptome analysis of CD34+ cells from myelodysplastic syndrome patients. *Leuk. Res.* **62**, 40–50 (2017).

27. Bowling, S. *et al.* An engineered CRISPR/Cas9 mouse line for simultaneous readout of lineage histories and gene expression profiles in single cells. *Cell* **181**, 1410 (2020).

28. Schiffman, J. S. *et al.* Defining heritability, plasticity, and transition dynamics of cellular phenotypes in somatic evolution. *Nat. Genet.* **56**, 2174–2184 (2024).

29. Park, S. *et al.* Genome-wide genotype-based risk model for survival in core binding factor acute myeloid leukemia patients. *Ann. Hematol.* **97**, 955–965 (2018).

30. Venditti, A. *et al.* Prognostic relevance of the expression of Tdt and CD7 in 335 cases of acute myeloid leukemia. *Leukemia* **12**, 1056–1063 (1998).

31. Saxena, A. *et al.* Biologic and clinical significance of CD7 expression in acute myeloid leukemia. *Am. J. Hematol.* **58**, 278–284 (1998).

32. Ogata, K. *et al.* Reappraisal of the clinical significance of CD7 expression in association with cytogenetics in de novo acute myeloid leukaemia: Short Report. *Br. J. Haematol.* **115**, 612–615 (2001).

33. Dalton, W. T., Jr *et al.* HL-60 cell line was derived from a patient with FAB-M2 and not FAB-M3. *Blood* **71**, 242–247 (1988).

34. DeTomaso, D. & Yosef, N. Hotspot identifies informative gene modules across modalities of single-cell genomics. *Cell Syst.* **12**, 446-456.e9 (2021).

35. Ribatti, D., Tamma, R. & Annese, T. Epithelial-mesenchymal transition in cancer: A historical overview. *Transl. Oncol.* **13**, 100773 (2020).

36. Wu, S., Du, Y., Beckford, J. & Alachkar, H. Upregulation of the EMT marker vimentin is associated with poor clinical outcome in acute myeloid leukemia. *J. Transl. Med.* **16**, 170 (2018).

37. Varisli, L. & Vlahopoulos, S. Epithelial-mesenchymal transition in acute leukemias. *Int. J. Mol. Sci.* **25**, 2173 (2024).

38. Greaves, D. & Calle, Y. Epithelial mesenchymal transition (EMT) and associated invasive adhesions in solid and haematological tumours. *Cells* **11**, 649 (2022).

39. Cuevas, D. *et al.* The role of epithelial-to-mesenchymal transition transcription factors (EMT-TFs) in acute myeloid leukemia progression. *Biomedicines* **12**, 1915 (2024).

40. Qi, J. *et al.* Five EMT-Related Gene Signatures Predict Acute Myeloid Leukemia Patient Outcome. *Disease Markers* **2022**, 1–8 (2022).

41. Damiano, J. S., Hazlehurst, L. A. & Dalton, W. S. Cell adhesion-mediated drug resistance (CAM-DR) protects the K562 chronic myelogenous leukemia cell line from apoptosis induced by BCR/ABL inhibition, cytotoxic drugs, and gamma-irradiation. *Leukemia* **15**, 1232–1239 (2001).

42. Hu, Q., Wang, M., Wang, J., Tao, Y. & Niu, T. Development of a cell adhesion-based prognostic model for multiple myeloma: Insights into chemotherapy response and potential reversal of adhesion effects. *Oncol. Res.* **32**, 753–768 (2024).

43. Sasca, D. *et al.* NCAM1 (CD56) promotes leukemogenesis and confers drug resistance in AML. *Blood* **133**, 2305–2319 (2019).

44. Lim, B. C. *et al.* Prickle1 promotes focal adhesion disassembly in cooperation with the CLASP-LL5β complex in migrating cells. *J. Cell Sci.* **129**, 3115–3129 (2016).

45. Jin, H. *et al.* BMP2 promotes migration and invasion of breast cancer cells via cytoskeletal reorganization and adhesion decrease: an AFM investigation. *Appl. Microbiol. Biotechnol.* **93**, 1715–1723 (2012).

46. Zhang, P. *et al.* CPNE8 promotes gastric cancer metastasis by modulating focal adhesion pathway and tumor microenvironment. *Int. J. Biol. Sci.* **18**, 4932–4949 (2022).

47. Škarková, A. *et al.* Microtubule-associated NAV3 regulates invasive phenotypes in glioblastoma cells. *Brain Pathol.* **35**, e13294 (2025).

48. Csizmar, C. M., Kim, D.-H. & Sachs, Z. The role of the proteasome in AML. *Blood Cancer J.* **6**, e503 (2016).

49. Groettrup, M. *et al.* The interferon-gamma-inducible 11 S regulator (PA28) and the LMP2/LMP7 subunits govern the peptide production by the 20 S proteasome in vitro. *J. Biol. Chem.* **270**, 23808–23815 (1995).

50. Seliger, B., Cabrera, T., Garrido, F. & Ferrone, S. HLA class I antigen abnormalities and immune escape by malignant cells. *Semin. Cancer Biol.* **12**, 3–13 (2002).

51. Kurnick, J. T. *et al.* A novel autocrine pathway of tumor escape from immune recognition: melanoma cell lines produce a soluble protein that diminishes expression of the gene encoding the melanocyte lineage melan-A/MART-1 antigen through down-modulation of its promoter. *J. Immunol.* **167**, 1204–1211 (2001).

52. Raule, M. *et al.* PA28αβ reduces size and increases hydrophilicity of 20S immunoproteasome peptide products. *Chem. Biol.* **21**, 470–480 (2014).

53. Farrar, J. E. *et al.* Genomic profiling of pediatric acute myeloid leukemia reveals a changing mutational landscape from disease diagnosis to relapse. *Cancer Res.* **76**, 2197–2205 (2016).

54. Kim, H. N., Ruan, Y., Ogana, H. & Kim, Y.-M. Cadherins, selectins, and integrins in CAM-DR in leukemia. *Front. Oncol.* **10**, 592733 (2020).

55. Chen, W.-C., Hu, G. & Hazlehurst, L. A. Contribution of the bone marrow stromal cells in mediating drug resistance in hematopoietic tumors. *Curr. Opin. Pharmacol.* **54**, 36–43 (2020).

56. Kuželová, K., Obr, A., Marková, J. & Gašová, Z. Integrin expression and adhesivity to fibronectin in primary acute myeloid leukemia cells: Impact of NPM1 and FLT3 mutations. *Eur. J. Haematol.* **105**, 578–587 (2020).

57. Estrugo, D. *et al.* Ligand bound beta1 integrins inhibit procaspase-8 for mediating cell adhesion-mediated drug and radiation resistance in human leukemia cells. *PLoS One* **2**, e269 (2007).

58. Keller, M. *et al.* The proteasome immunosubunits, PA28 and ER-aminopeptidase 1 protect melanoma cells from efficient MART-126-35 -specific T-cell recognition: Antigen processing. *Eur. J. Immunol.* **45**, 3257–3268 (2015).

59. Li, L. *et al.* A mouse model with high clonal barcode diversity for joint lineage, transcriptomic, and epigenomic profiling in single cells. *Cell* **186**, 5183-5199.e22 (2023).

60. Kalhor, R. *et al.* Developmental barcoding of whole mouse via homing CRISPR. *Science* **361**, eaat9804 (2018).

61. Emiliani, F. E., Hsu, I. & McKenna, A. Multiplexed assembly and annotation of synthetic biology constructs using long-read nanopore sequencing. *ACS Synth. Biol.* **11**, 2238–2246 (2022).

62. Johnson, M. S., Venkataram, S. & Kryazhimskiy, S. Best Practices in Designing, Sequencing, and Identifying Random DNA Barcodes. *J. Mol. Evol.* **91**, 263–280 (2023).

63. Sprouffske, K. Simple Metrics to Summarize Growth Curves [R package growthcurver version 0.3.1]. (2020).

64. Ritz, C., Baty, F., Streibig, J. C. & Gerhard, D. Dose-Response Analysis Using R. *PLoS One* **10**, e0146021 (2015).

65. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

66. McGinnis, C. S. *et al.* MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nature Methods 2019 16:7* **16**, 619–626 (2019).

67. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).

68. Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).

69. Germain, P.-L., Lun, A., Garcia Meixide, C., Macnair, W. & Robinson, M. D. Doublet identification in single-cell sequencing data using *scDblFinder*. *F1000Res.* **10**, 979 (2021).

70. Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).

71. Choudhary, S. & Satija, R. Comparison and evaluation of statistical error models for scRNA-seq. *Genome Biol.* **23**, 27 (2022).
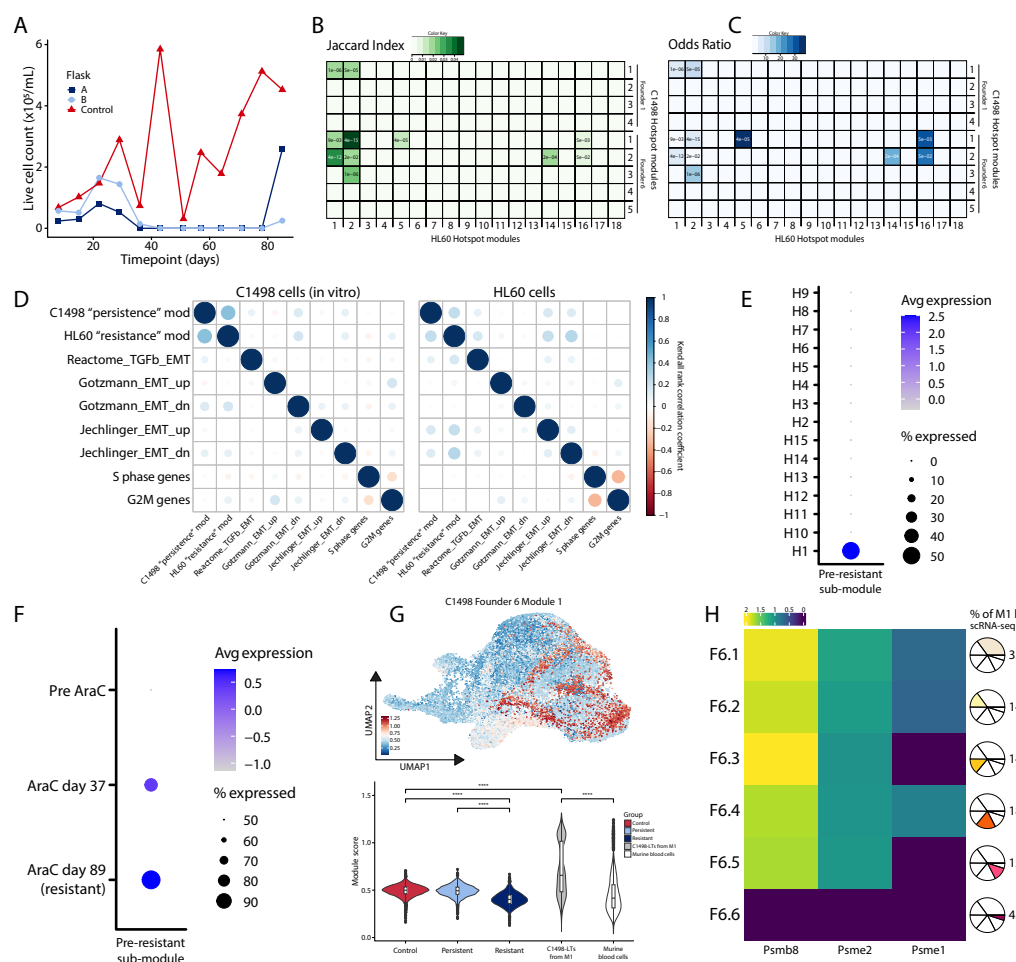
1034   72. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq

1035   data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).

1036   73. Korotkevich, G. *et al.* Fast gene set enrichment analysis. *bioRxiv* 060012 (2016)

1037   doi:10.1101/060012.

1038   74. Jones, M. G. *et al.* Inference of single-cell phylogenies from lineage tracing data using

1039   Cassiopeia. *Genome Biol.* **21**, 1–27 (2020).

1040   75. Zorita, E., Cuscó, P. & Filion, G. J. Starcode: sequence clustering based on all-pairs

1041   search. *Bioinformatics* **31**, 1913–1919 (2015).

1042   76. *GeneOverlap: R Package for Testing and Visualizing Gene List Overlaps*. (Github).

1043   77. Therneau, T. M. Survival Analysis [R package survival version 3.8-3]. (2024).

1044   78. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene

1045   expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).

1046
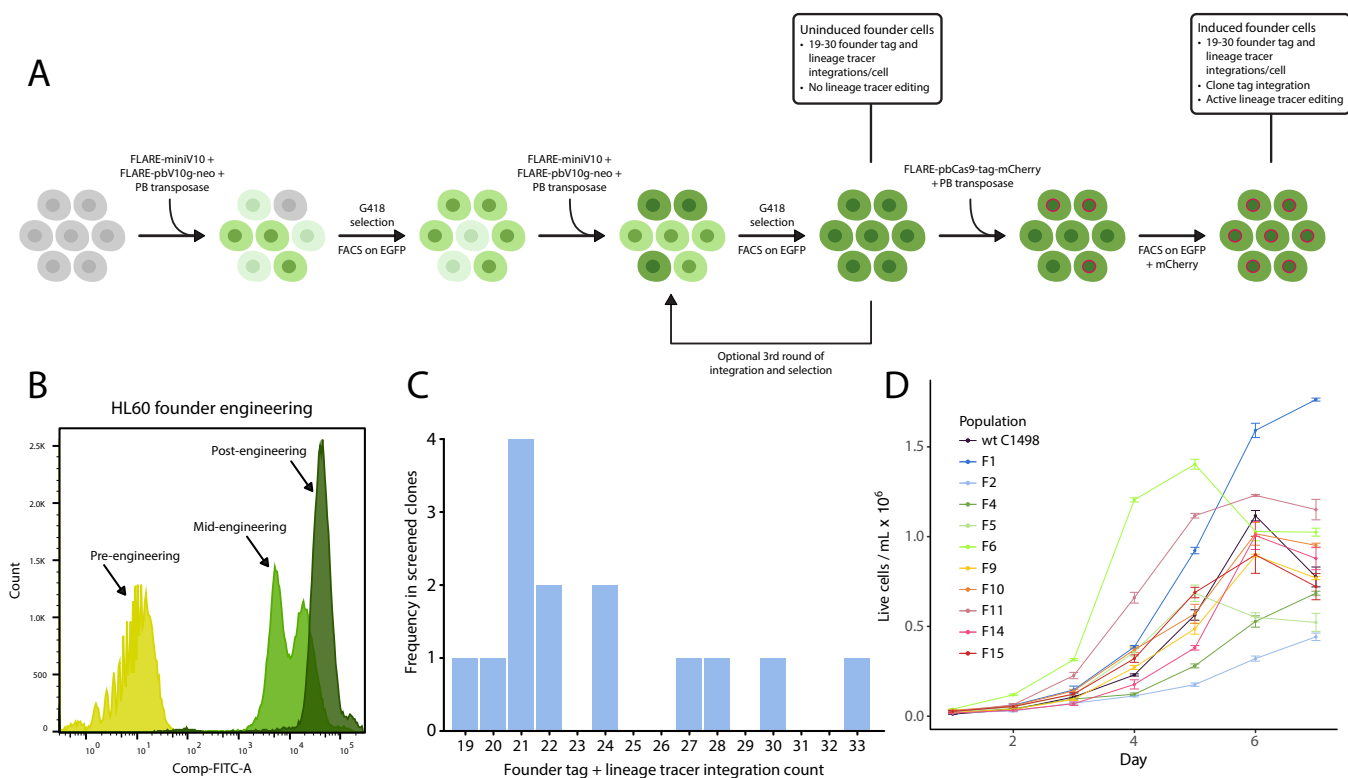
Supplemental figure 1: FLARE cell line engineering allows for the generation of founder populations maintaining intra-founder heterogeneity. (**A**) Schematic of engineering lineage tracing AML cells. Wildtype cells are first electroporated with FLARE-miniV10-10X-UMI + FLARE-pbV10g-neo + PB-transposase. Successfully integrated cells are selected via G418 treatment (FLARE-pbV10g-neo) and FACS for EGFP (FLARE-miniV10-10X-UMI). This process is repeated 1-2 more times, successively selecting for cells with high numbers of integrations. Once the desired number of integrations is reached, cells are considered uninduced founder cells. Prior to the start of a lineage tracing experiment, uninduced founder cells are electroporated with FLARE-pbCas9-tag-mCherry and PB-transposase and sorted on EGFP (FLARE-miniV10-10X-UMI) and mCherry (FLARE-pbCas9-tag-mCherry). Resulting cells are referred to as induced founder cells and contain all the necessary integrations for founder and clonal tracking. (**B**) FACS plots showing the increase in EGFP expression in HL60 cells through 2 rounds of electroporation with FLARE-miniV10-10X-UMI + FLARE-pbV10g-neo + PB transposase. Post-engineering uninduced HL60-FLARE founder cells uniformly express high levels of EGFP. (**C**) Uninduced C1498-FLARE founder populations were subcultured to determine founder tag and lineage recorder integration counts. Founders contained between 19 and 33 integrations. (**D**) Subcultured founder populations display a range of growth rates and carrying capacities, indicating sustained heterogeneity following the engineering process.
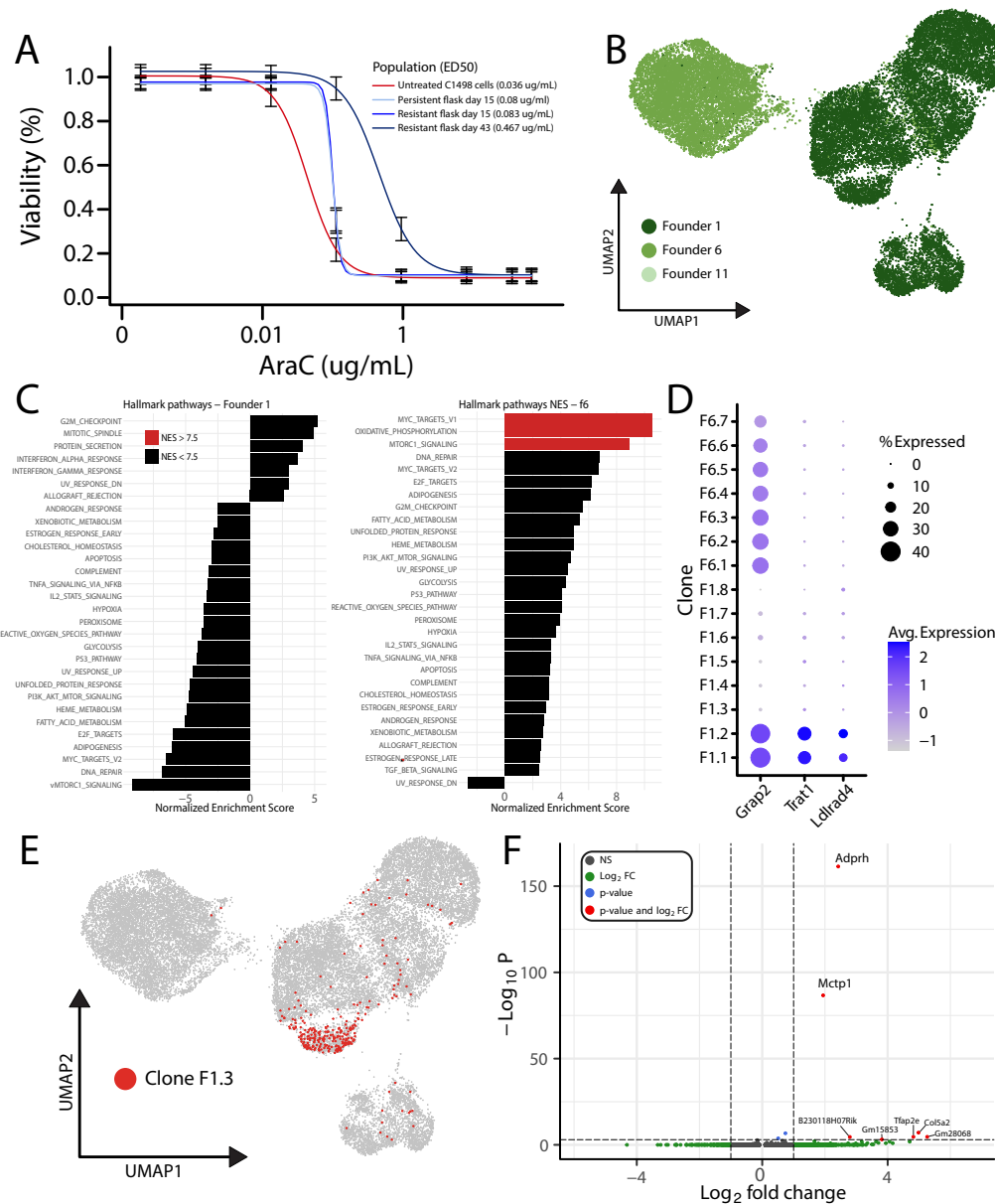
Supplemental figure 2: AraC treatment of C1498-FLARE cells in vitro leads to the generation of AraC-insensitive populations. FLARE recorder analysis reveals founder and clone-level transcriptomic differences. (**A**) AraC dose-response curves for C1498-FLARE cells through in vitro treatment. Cells in the resistant flask displayed an increase in ED50 values throughout treatment, indicating the emergence of an AraC-resistant population. Cells in the persistent flask also displayed an increase in ED50 values on day 15 of AraC treatment, but did not recover enough for another dose-response curve on day 43, suggesting higher AraC sensitivity compared to the resistant flask. (**B**) Captured C1498-FLARE cells colored by founder population. (**C**) GSEA analysis on founder 1 and founder 6 control cells to identify basal founder-specific differences. Analysis was conducted using the fGSEA R package and Hallmark gene sets from the mSigDB. (**D**) Dot plot showing expression of Grap2, Trat1, and Ldlrad4 in C1498-FLARE clone populations. (**E**) UMAP of C1498-FLARE cells with clone F1.3 cells highlighted in red. (**F**) Differential gene expression analysis on clone F1.3 cells compared to all founder 1 control cells uncovers clone-specific increased expression of Adprh and Mctp1.

Supplemental figure 3: AraC treatment of C1498-FLARE cells in vivo leads to outgrowth of founder 6 cells. (**A**) Founder tag sequencing of blood draws taken 28 days after mice were injected IV with C1498-FLARE cells. (**B**) Clone populations for each sample, including in vivo sequencing from M1, first ordered by founder then by clone size. (**C**) Bulk sequencing of clone tags detected in blood draws from mouse M3. Clone tag reads were used to determine the clonal proportions of leukemic burden in the blood from day 30 to euthanasia on day 50. Founder proportions differ from figure 2E due to subsetting for clone tags detected in single cell data. (**D**) Captured single cells from endpoint blood and bone marrow collected from mouse M1. (**E**) Captured single cells from endpoint blood and bone marrow collected from mouse M1. Predicted C1498-FLARE cells were defined as having sufficient founder and clone tag reads to assign lineage identities, highlighted in red. (**F**) Cross-correlation heatmap of the top 35 auto-correlated genes over the Founder 1 tree.

1092
1093 **Supplemental figure 4: AraC treatment of HL60-FLARE cells in vitro leads to the generation of**
1094 **chemoresistance. Resistance-associated heritable Hotspot modules in C1498-FLARE and**
1095 **HL60-FLARE cells suggest the existence of drug response pathways common to murine and**
1096 **human AML cells.** (**A**) Live cell counts for HL60-FLARE cells undergoing in vitro AraC treatment.
1097 Each flask was counted in duplicates at each timepoint, and counts were averaged. Control flask
1098 counts vary due to culture splitting in the days prior to counting. Treated flasks A and B did not have
1099 enough cells to be detected between day 40 and 80 and were not split during that time frame. (**B**)
1100 Jaccard Index similarity comparing C1498-FLARE and HL60-FLARE Hotspot modules. Square color
1101 is the Jaccard Index; values indicate BH-corrected P-values. (**C**) The odds ratio of gene overlaps
1102 comparing C1498-FLARE and HL60-FLARE Hotspot modules. Square color determined by odds
1103 ratio; values indicate BH-corrected P-values. (**D**) Expression scores for C1498 "persistence"
1104 Founder 1 module 1, HL60 "resistance" module 2, various EMT gene sets, and cell cycle gene sets
1105 are minimally correlated as measured by Kendall ranked correlation coefficient in C1498-FLARE and
1106 HL60-FLARE cells in vitro. (**E**) Dot plot showing expression of the pre-resistant sub-module in pre-
1107 experiment HL60-FLARE cells by founder. Founder H1 shows distinct upregulation of the signature.
1108 (**F**) Dot plot showing expression of the pre-resistant sub-module in AraC-treated HL60-FLARE H1
1109 cells pre-experiment, at midpoint, and at endpoint. Pre-resistant signature expression increases
1110 through AraC treatment. (**G**) UMAP of founder 6 module 1 expression in the in vitro C1498-FLARE
1111 cells integrated with M1 endpoint sequencing. Below is expression quantification by cell group and
1112 shows module expression enrichment in cells collected from M1. (**H**) Log2 fold change values of
1113 Psmb8, Psme1, Psme2 genes for C1498-FLARE founder 6 clones comparing in vitro to in vivo clone
1114 cells.
1115
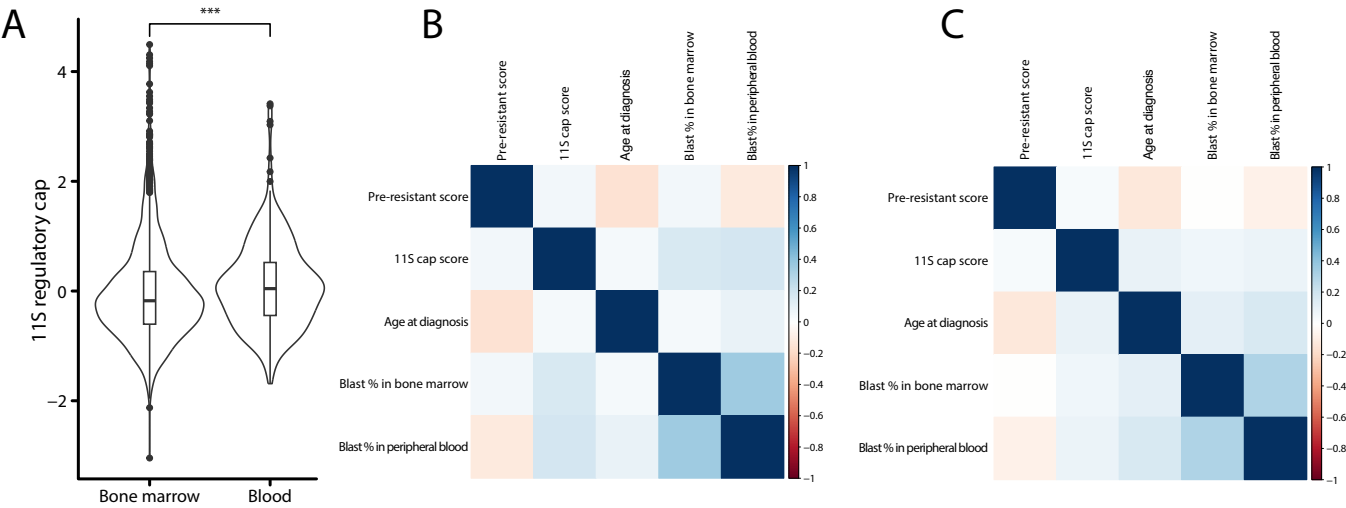
Supplemental figure 5: Expression of the 11S cap and pre-resistant signature are not strongly correlated with bone marrow or peripheral blood blast percentage at diagnosis. (**A**) Violin plot showing 11S regulatory cap gene set scores calculated from TARGET-AML patient samples taken at diagnosis. Mean expression in blood is significantly higher than in bone marrow. (**B**) TARGET-AML expression scores for 11S regulatory cap and pre-resistant signature in the bone marrow at diagnosis, age at diagnosis, and percentage of leukemic blasts in bone marrow and peripheral blood are minimally correlated as measured by Kendall ranked correlation coefficient. (**C**) TARGET-AML expression scores for 11S regulatory cap and pre-resistant signature in the peripheral blood at diagnosis, age at diagnosis, and percentage of leukemic blasts in bone marrow and peripheral blood are minimally correlated as measured by Kendall ranked correlation coefficient.