

1 **Functional investigation of inherited noncoding genetic variation impacting the**
2 **pharmacogenomics of childhood acute lymphoblastic leukemia treatment**

3
4 Kashi Raj Bhattarai, PhD^{1,2†}, Robert J. Mobley, PhD^{1,2†}, Kelly R. Barnett, PhD^{1,2}, Daniel C.
5 Ferguson, PhD^{1,2}, Baranda S. Hansen, MS^{3,4}, Jonathan D. Diedrich, PhD^{1,2}, Brennan P.
6 Bergeron, PhD^{1,2,5}, Wenjian Yang, PhD^{1,2}, Kristine R. Crews, PharmD^{1,2}, Christopher S.
7 Manring, MBA⁶, Elias Jabbour, MD⁷, Elisabeth Paietta, PhD⁸, Mark R. Litzow, MD⁹, Steven M.
8 Kornblau, MD⁷, Wendy Stock, MD¹⁰, Hiroto Inaba, MD, PhD^{1,11}, Sima Jeha, MD^{1,11}, Ching-Hon
9 Pui, MD^{1,11}, Cheng Cheng, PhD¹², Shondra M. Pruett-Miller, PhD^{3,4}, Mary V. Relling, PharmD^{1,2},
10 Jun J. Yang, PhD^{1,2,5,13}, William E. Evans, PharmD^{1,2} and Daniel Savic, PhD^{1,2,5,13,*}

11
12
13
14 ¹Hematological Malignancies Program, St. Jude Children's Research Hospital, Memphis, TN
15 ²Department of Pharmacy and Pharmaceutical Sciences, St. Jude Children's Research Hospital,
16 Memphis, TN
17 ³Center for Advanced Genome Engineering, St. Jude Children's Research Hospital, Memphis, TN 38105,
18 USA.
19 ⁴Department of Cell and Molecular Biology, St. Jude Children's Research Hospital, Memphis, TN 38105,
20 USA.
21 ⁵Graduate School of Biomedical Sciences, St. Jude Children's Research Hospital, Memphis, TN
22 ⁶Alliance Hematologic Malignancy Biorepository; Clara D. Bloomfield Center for Leukemia
23 Outcomes Research, Columbus, OH 43210, USA
24 ⁷Department of Leukemia, The University of Texas MD Anderson Cancer Center, Houston, TX
25 ⁸Albert Einstein College of Medicine, New York, NY
26 ⁹Division of Hematology, Department of Medicine, Mayo Clinic, Rochester, MN 55905, USA.
27 ¹⁰Comprehensive Cancer Center, University of Chicago Medicine, Chicago, IL
28 ¹¹Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN
29 ¹²Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN
30 ¹³Integrated Biomedical Sciences Program, University of Tennessee Health Science Center, Memphis,
31 TN

32
33
34 †Authors contributed equally to this work

35 *Corresponding author: Daniel Savic, PhD
36 Division of Pharmaceutical Sciences
37 Department of Pharmacy and Pharmaceutical Sciences
38 St. Jude Children's Research Hospital
39 262 Danny Thomas Place
40 Memphis, TN, 38105
41 daniel.savic@stjude.org

42 **ABSTRACT (155/160 words max)**

43 Although acute lymphoblastic leukemia (ALL) is the most common childhood cancer, there is
44 limited understanding of the contribution of inherited genetic variation on inter-individual
45 differences in chemotherapy response. Defining genetic factors impacting therapy failure can help
46 better predict response and identify drug resistance mechanisms. We therefore mapped inherited
47 noncoding variants associated with chemotherapeutic drug resistance and/or treatment outcome
48 to ALL *cis*-regulatory elements and investigated their gene regulatory potential and genomic
49 connectivity using massively parallel reporter assays and promoter capture Hi-C, respectively.
50 We identified 53 variants with reproducible allele-specific effects on transcription and high-
51 confidence gene targets. Subsequent functional interrogation of the top variant (rs1247117)
52 determined that it disrupted a PU.1 consensus motif and PU.1 binding affinity. Importantly,
53 deletion of the genomic interval containing rs1247117 sensitized ALL cells to vincristine.
54 Together, these data demonstrate that noncoding regulatory variation associated with diverse
55 pharmacological traits harbor significant effects on allele-specific transcriptional activity and
56 impact sensitivity to chemotherapeutic agents in ALL.

57 INTRODUCTION

58 Due to continual advances in treatment protocol over the last 60 years, the survival rate
59 of the most common malignancy in children, acute lymphoblastic leukemia (ALL), has dramatically
60 improved to over 90% in the high-income countries (1). Despite these advances, survival rates of
61 pediatric patients experiencing refractory or relapsed ALL were only 30-50% and those of adults
62 were especially low (~10%) (2). Thus, improving the understanding of the underlying genetic risk
63 factors impacting response to ALL chemotherapy is a major step in improving outcomes for
64 patients with refractory or relapsed ALL.

65 Genome-wide association studies (GWAS) have identified numerous inherited DNA
66 sequence variants associated with treatment outcome in childhood ALL from clinical trials carried
67 out by St. Jude Children's Research Hospital and the Children's Oncology group (3-5). This
68 includes GWAS analyses that identified inherited genetic contributors associated with patient
69 relapse (4, 5) and persistence of minimal residual disease (MRD) after induction chemotherapy
70 (3), which is an early indicator of treatment failure (6-9). In addition, *ex vivo* chemotherapeutic
71 drug sensitivity testing using primary ALL cells from patients serves as an informative
72 pharmacological phenotype (10). When integrated with genotype profiling for GWAS, these
73 analyses identify variants contributing to antileukemic drug resistance that reflects *in vivo* and *ex*
74 *vivo* resistance and is therefore predictive of treatment outcome in patients (10-22).

75 Because most GWAS variants, including pharmacogenomic variants (23, 24), lie in
76 noncoding sequences in the human genome, their connection to gene regulation and cellular
77 biology has yet to be established. Moreover, given that dozens of variants are typically in strong
78 linkage disequilibrium (LD) with an associated sentinel variant, pinpointing causal variants at
79 GWAS loci has been challenging. Noncoding GWAS variants have been consistently linked to
80 disruption of *cis*-regulatory element (CRE) activity and gene regulation (25). As a result, the
81 functional evaluation of these regulatory variants involves an examination of their allele-specific
82 activities on transcriptional output which has traditionally been a low-throughput endeavor.

83 Therefore, the functional investigation of all associated regulatory variation at GWAS loci (sentinel
84 and LD) proved to be an intractable hurdle to investigators. Recent technological advances
85 however have ameliorated these challenges through the advent of massively parallel reporter
86 assays (MPRAs) where the reporter aspect is often a self-transcribed barcode in the 3' UTR of a
87 reporter gene that is detected using next-generation sequencing. MPRAs allow the simultaneous,
88 rapid and robust detection of differences in transcriptional output from a library of *cis*-regulatory
89 sequences of interest (26-28). MPRAs have since been applied to the study of regulatory variation
90 at GWAS loci through an examination of allele-specific effects on reporter gene expression (29-
91 35).

92 Another challenge is connecting regulatory variation at promoter-distal CREs to a target
93 gene, as the closest gene may not be the target gene (36). To circumvent these challenges,
94 regulatory variation can be coupled to transcriptomics to identify variants impacting the expression
95 of a candidate target gene through expression quantitative trait locus (eQTL) mapping (36, 37).
96 Functional genomics offer additional solutions through the mapping of three-dimensional (3D)
97 genomic interactions (38). Because promoter-distal CREs (e.g., enhancers) regulate gene
98 expression through long-range 3D looping to the promoters of target genes (25), an attractive
99 assay to identify gene targets of promoter-distal regulatory variants is promoter capture Hi-C
100 (promoter CHiC) (39). Promoter capture Hi-C and related 3D chromatin interaction assays have
101 been implemented at multiple GWAS loci to identify gene targets of promoter-distal regulatory
102 variation (29, 36, 40-43).

103 To better understand the underlying genetic and gene regulatory factors that impact
104 diverse pharmacological traits in ALL, we performed a comprehensive functional interrogation of
105 GWAS regulatory variants that map to ALL accessible chromatin sites and that are associated
106 with *ex vivo* chemotherapeutic drug resistance in primary ALL cells from patients and/or ALL
107 treatment outcome (i.e., relapse and persistence of MRD) in patients using MPRA. We coupled
108 these results with promoter CHiC to identify candidate target genes of functional regulatory

109 variants with significant allele-specific effects on reporter gene expression. Finally, we functionally
110 investigated the impact of the top regulatory variant on transcription factor binding, neighboring
111 gene expression and chemotherapeutic drug resistance in ALL cell lines. To our knowledge, this
112 study represents the largest functional investigation of regulatory variants impacting the
113 pharmacogenomics of chemotherapy treatment and fills an unmet need for large-scale functional
114 examinations of regulatory GWAS variants associated with pharmacological traits.

115

116 **RESULTS**

117

118 **Identification of regulatory variants impacting the pharmacogenomics of ALL treatment**

119 Single nucleotide variants (SNVs) impacting diverse pharmacological traits in ALL were identified
120 for functional interrogation. We chose SNVs associated with relapse or persistence of MRD after
121 induction chemotherapy in childhood ALL patients to investigate the role of inherited regulatory
122 variants impacting clinical phenotypes (i.e., treatment outcome). These SNVs were identified from
123 published GWAS of ALL patients enrolled in St. Jude Children's Research Hospital and the
124 Children's Oncology Group clinical protocols (3-5) (see **Methods** for SNV selection criteria).
125 Variant selection included additional prioritization for SNVs associated with drug resistance
126 phenotypes in primary ALL cells to enrich for variation impacting ALL cell biology. These treatment
127 outcome-associated variants, as well as all variants in high LD ($r^2 > 0.8$) with the sentinel GWAS
128 variants, were further evaluated (**Fig 1A-B**).

129 We also identified variants directly associated with *ex vivo* chemotherapeutic drug
130 resistance in primary ALL cells from patients by performing GWAS analyses using SNV genotype
131 information and *ex vivo* drug resistance assay results for six antileukemic agents (prednisolone,
132 dexamethasone, vincristine, L-asparaginase, 6-mercaptopurine and 6-thioguanine) in primary
133 ALL cells from 312-344 patients enrolled in the Total Therapy XVI clinical protocol at St. Jude
134 Children's Research Hospital (see **Methods**). We further prioritized for functional *ex vivo* drug

135 resistance SNVs by examining if they were eQTLs in primary ALL cells or in EBV-transformed
136 lymphocytes from the Genotype-Tissue Expression (GTEx) consortium (37). All *ex vivo* drug
137 resistance-associated eQTL variants, as well as variants in high LD ($r^2>0.8$) with these sentinel
138 GWAS variants, were further evaluated (**Fig 1A-B**).

139 GWAS have also been performed for childhood ALL disease susceptibility and identified
140 several GWAS loci harboring variants with genome-wide significance (44-50). Several follow-up
141 studies of these GWAS loci have identified candidate causal noncoding regulatory variants and
142 mechanisms involving gene regulatory disruptions (51-53). As a result, we used ALL disease
143 susceptibility variants, as well as variants in high LD ($r^2>0.8$) with them, as positive controls in our
144 study (**Fig 1A-B**).

145 Because most of these variants map to noncoding portions of the human genome, these
146 data point to disruptions in gene regulation as the underlying mechanism of how these variants
147 impact ALL cell biology. We therefore utilized assay for transposase-accessible chromatin with
148 high-throughput sequencing (ATAC-seq) (54) chromatin accessibility data in 158 ALL cell models,
149 comprised of primary ALL cells (cryopreserved, n=24 (55); fresh, n=120), ALL cell lines (n=14)
150 and ALL patient-derived xenografts (PDXs, n=3), to uncover which variants map to putative CREs
151 in ALL cells (56) (i.e., regulatory variants; **Fig 1C**). ATAC-seq data from primary ALL, ALL cell
152 lines and PDXs were combined and identified 1696 regulatory variants at accessible chromatin
153 sites in ALL cells for functional investigation (**Fig 1C** and **Sup File 1**).

154

155 **Assessing the impact of regulatory variation on transcriptional output using MPRA**

156 To examine the functional effects of these 1696 regulatory variants on transcriptional output in a
157 high-throughput manner we utilized a barcode-based MPRA platform (29, 32) to measure
158 differences in allele-specific transcriptional output (**Fig 2A**). Oligonucleotides containing 175-bp
159 of genomic sequence centered on each reference (ref) or alternative (alt) variant allele, a
160 restriction site, and a unique 10-bp barcode sequence were cloned into plasmids. An open

161 reading frame containing a minimal promoter driving GFP was then inserted at the restriction site
162 between the alleles of interest and their unique barcodes (**Fig 2A**). We utilized 28 unique 3'UTR
163 DNA barcodes per variant allele (56 barcodes per regulatory variant), and variants near
164 bidirectional promoters (47 total variants) were tested using both sequence orientations. In total,
165 97,608 variant-harboring oligonucleotides were evaluated for allele-specific differences in gene
166 regulatory activity (**Fig 2A**).

167 Following transfection into 7 different B-cell precursor ALL (B-ALL; 697, BALL1, Nalm6,
168 REH, RS411, SEM, SUPB15) and 3 T-cell ALL (T-ALL; CEM, Jurkat, P12-Ichikawa) human cell
169 lines (n=4 transfections per cell line; 40 total), the transcriptional activity of each allele variant was
170 measured by high-throughput sequencing to determine the barcode representation in reporter
171 mRNA and compared to DNA counts obtained from high-throughput sequencing of the MPRA
172 plasmid pool (**Fig 2A**). MPRA detected 4633 instances of significant differential activity between
173 alleles across 91% (1538/1696) of regulatory variants across the 10 ALL cell lines tested (**Fig 2B-**
174 **C, Sup File 2**). The 10 ALL cell lines showed substantial differences in the total number of
175 regulatory variants harboring significant allele-specific activity (**Fig 2C**). Importantly, when
176 comparing changes in allele-specific MPRA activity for each regulatory variant we found that
177 significant changes in activity (adj. $p < 0.05$) were highly correlated between ALL cell lines, with
178 87% concordance in allelic-specific activity, suggesting that significant MPRA hits were likely to
179 be robust and reproducible between cell lines (**Fig 2D**). Allele-specific MPRA activities were also
180 correlated using all pairwise cell line comparisons for each regulatory variant, irrespective of
181 significance (**Sup Fig 1**). Importantly, 31 of the 35 positive control variants (i.e., ALL disease
182 susceptibility-associated variants and variants in high LD) showed significant allelic effects in at
183 least 1 cell line, and 10 showed significant and concordant allelic effects in at least 3 ALL cell
184 lines, including two variants (rs3824662 at *GATA3* locus and rs75777619 at 8q24.21) directly
185 associated with ALL susceptibility (44, 49, 52). The risk A allele at rs3824662 was associated with
186 higher *GATA3* expression and chromatin accessibility and demonstrated significantly higher

187 allele-specific activity in our MPRA (44, 52), thereby demonstrating that the MPRA could detect
188 allelic effects identified by others. Overall, these data suggest that the chemotherapeutic drug
189 sensitivity and patient treatment outcome SNVs tested were heavily enriched for functional
190 regulatory variants with the potential to impact gene regulation.

191
192 **Identification of functional regulatory variants showing reproducible and concordant**
193 **changes in allele-specific gene expression**

194 To further focus on regulatory variants most likely to broadly impact gene regulation in ALL cells,
195 we prioritized 556 variants with significant and concordant allele-specific activities in at least 3
196 ALL cell lines (i.e., functional regulatory variants; **Fig 3A-B, Sup File 3**). Most of these functional
197 regulatory variants (318/556) mapped to accessible chromatin found only in primary ALL cell
198 samples, underscoring the importance of incorporating chromatin architecture from primary ALL
199 cells, and 54 functional regulatory variants mapped to transcription factor footprints in primary
200 ALL cells (**Sup Fig 2**). However, because further functional investigation of variants in primary
201 ALL cells is currently intractable, we focused on 210 functional regulatory variants that reside in
202 open chromatin in an ALL cell line, and most of these variants (159/210; 76%) were also found in
203 accessible chromatin in PDX and/or in primary ALL cells from patients (**Fig 3B-C**).

204 For validation using traditional luciferase reporter assays, we prioritized these 210
205 functional regulatory variants based on allele-specific effect size and selected high-ranking SNVs
206 with known eQTL status. Dual-luciferase reporter assays showed similar allele-specific changes
207 in activity to that which was detected by MPRA, in validation of our MPRA analysis (**Fig 3D, Sup**
208 **Fig 3**). In fact, a significant positive correlation ($p=0.0017$) was observed between the allelic
209 effects detected by MPRA and luciferase reporter assays (**Fig 3D**). Together these analyses
210 validated the robustness of our MPRA screen of functional regulatory variants and identified over
211 500 SNVs with reproducible and concordant allele-specific effects on gene expression.

212

213 **Association of functional regulatory variants with putative gene targets**

214 We determined if the 210 functional regulatory variants found in accessible chromatin sites in ALL
215 cell lines were directly associated with target gene regulation. While 34 functional regulatory
216 variants were localized close (+/-2.5kb) to nearby promoters (**Fig 4A, Sup File 4**), 176 variants
217 were promoter-distal, and therefore likely to map to CREs with unclear gene targets (**Fig 4A**). We
218 therefore performed promoter ChIP in 8 of 10 ALL cell lines used in MPRA and determined that
219 19 of the 176 functional regulatory variants showed evidence of connectivity to distal promoters
220 in the same cell line where allele-specific MPRA activity and chromatin accessibility were detected
221 (**Fig 4A, Sup File 4**). In prioritizing functional regulatory variants, we were interested in the gene
222 regulatory impacts of variants at TSS-proximal promoter-associated versus TSS-distal promoter-
223 connected CREs. Interestingly, we found that SNVs found in open chromatin at TSS-distal sites
224 with promoter connectivity showed higher allele-specific changes in MPRA activity than those at
225 promoters (**Fig 4B**). Amongst the TSS-distal promoter-connected functional regulatory variants,
226 we found that distal intergenic and intronic SNVs showed significantly higher allele-specific activity
227 than those in UTRs (**Fig 4C**). These data suggest that the most robust allelic effects attributable
228 to these regulatory variants are likely to occur at distal intergenic and intronic sites >2.5kb from
229 the TSS of the target gene.

230 Next, we ranked TSS-proximal promoter-associated and TSS-distal promoter-connected
231 functional regulatory variants by the geometric mean of their significant MPRA data to account for
232 the magnitude of allele-specific activity and the reproducibility of a significant change across ALL
233 cell lines (**Fig 4D-E**). This analysis identified rs1247117 as the most robust functional regulatory
234 variant to pursue for mechanistic understanding (**Fig 4E**).

235

236 **rs1247117 determines PU.1 binding and impacts sensitivity to vincristine**

237 We pursued functional validation of rs1247117 based on its highest ranking by geometric mean
238 of MPRA allelic effect. rs1247117 is in high LD with two GWAS sentinel variants (rs1312895,

239 $r^2=0.99$; rs1247118, $r^2=1$) that are associated with persistence of MRD after induction
240 chemotherapy (3). This functional regulatory variant maps to a distal intergenic region near the
241 *CACUL1* gene, for which it is an eQTL for in EBV-transformed lymphocytes (37). However,
242 rs1247117 also loops to the *EIF3A* promoter in Nalm6 B-ALL cells (**Fig 5A**). We therefore
243 explored how this accessible chromatin site might recruit transcriptional regulators that would
244 depend on the allele present at rs1247117. For this, we first performed ChIP-seq for RNA pol II
245 and H3K27Ac which further confirmed that rs1247117 is associated with an active CRE in Nalm6
246 cells (**Fig 5A**). Through an examination of the underlying DNA sequence spanning rs1247117,
247 we found that the reference guanine (G) risk allele at rs1247117 resides in a PU.1 transcription
248 factor binding motif that is disrupted by the alternative adenine (A) allele. Although the risk G allele
249 is the reference allele, the alternative A allele is more common in human populations. Supporting
250 PU.1 binding at this location, accessible chromatin profiling in primary ALL cells identified an
251 accessible chromatin site and PU.1 footprint spanning rs1247117 in diverse ALL samples (**Sup**
252 **Fig 4A-B**). Significantly greater chromatin accessibility at rs1247117 was also observed in
253 heterozygous (GA) patient samples compared to patient samples homozygous for the alternative
254 A allele (**Sup Fig 4C**), and the G allele at rs1247117 harbored significantly greater ATAC-seq
255 read count compared to the A allele (**Sup Fig 4D**). Importantly, we determined that PU.1 was
256 bound at this site in Nalm6 cells using CUT and RUN (57) (**Fig 5A-B**).

257 Nalm6 cells contain the alternative A allele that disrupts the PU.1 motif at rs1247117, yet
258 our data suggests that this site still binds PU.1 (**Fig 5A-B**). This led us to hypothesize that PU.1
259 binding affinity for the PU.1 motif surrounding rs1247117 would be strengthened by the risk G
260 allele. Therefore, we designed biotinylated DNA probes containing two tandem 25-bp regions
261 centered on reference G or alternative A allele-containing rs1247117 to test this hypothesis (**Fig**
262 **5C**). Using biotinylated probes we performed an *in vitro* DNA-affinity pulldown from Nalm6 nuclear
263 lysate and found that while PU.1 was indeed bound to the alternative A allele, PU.1 was more

264 robustly bound to the reference G allele at rs1247117 (**Fig 5D**). These data suggest that the risk
265 G allele increases the affinity of PU.1 binding at rs1247117 relative to the alternative A allele.

266 We were next interested in how allele-specific PU.1 binding was related to the expression
267 of nearby putative target genes. Because rs1247117 is a known eQTL associated with *CACUL1*
268 expression in EBV-transformed lymphocytes and our promoter ChIP data demonstrated
269 connectivity between rs1247117 and the promoter of *EIF3A*, we asked if the expression of these
270 two genes was altered by deletion of the CRE containing rs1247117 in Nalm6 cells. Using
271 CRISPR/Cas9 genome editing, we made a heterogeneous pool of Nalm6 cells harboring a
272 deletion of the CRE containing rs1247117 (**Fig 5E**). We found that CRE deletion resulted in a
273 significant up-regulation of both *CACUL1* and *EIF3A* (**Figure 5F-G**), suggesting an inverse
274 relationship between PU.1 binding and transcription of associated genes at this locus. Importantly,
275 this observation is concordant with GTEx eQTL data showing that the risk G allele harboring
276 stronger PU.1 binding is associated with down-regulation of *CACUL1* in EBV-transformed
277 lymphocytes (37).

278 Because the risk G allele at rs1247117 was also associated with vincristine resistance in
279 primary ALL cells from patients (3), we additionally sought to determine the impact of the CRE
280 deletion containing rs1247117 on cellular response to vincristine treatment. We hypothesized that
281 because the risk G allele is associated with enhanced PU.1 binding and resistance to vincristine,
282 complete disruption of PU.1 binding in Nalm6 cells harboring the CRE deletion would show
283 increased sensitivity to vincristine relative to parental Nalm6 cells. As predicted, Nalm6 cells with
284 the CRE deletion exhibited significantly increased sensitivity to vincristine across a range of
285 concentrations after 24, 48, and 72 hours of treatment (**Fig 5H**). Collectively, these data suggest
286 that a functional regulatory variant alters the binding affinity of a key transcription factor, PU.1,
287 and disruption of PU.1 binding at this locus impacts vincristine sensitivity in ALL cells.

288

289

290 **DISCUSSION**

291
292 Using MPRA, we systematically interrogated the functional effects of inherited noncoding variation
293 associated with relapse, persistence of MRD after induction chemotherapy and/or *ex vivo*
294 chemotherapeutic drug resistance in childhood ALL. We refined our search to regulatory variants
295 that were found in accessible chromatin sites in 158 ALL cell models, including primary ALL cells
296 from patients, PDXs and ALL cell lines, as those noncoding regions were likely to be participating
297 in transcriptional regulation. Using MPRA we identified 556 functional regulatory variants showing
298 reproducible and concordant changes in an allele-specific gene regulatory activity. To further
299 explore the impact of these variants on gene regulation in ALL cell lines, we selected a subset of
300 functional regulatory variants from MPRA that were within an accessible chromatin site in an ALL
301 cell line. We overcame difficulties in associating promoter-distal functional regulatory variants with
302 gene targets using promoter CHiC, and found 19 variants with robust looping to a distal promoter,
303 as well as 34 promoter-associated functional regulatory variants.

304 We identified rs1247117 as the top functional regulatory variant showing the highest
305 geometric mean of differential transcription activity, which was identified in 9 of 10 ALL cell lines
306 assayed by MPRA. We found that the allele present at rs1247117 was determinant of PU.1
307 transcription factor binding, with the risk G allele leading to greater chromatin accessibility and
308 PU.1 binding affinity. Interestingly, the allele-specific activities as measured by MPRA and
309 traditional dual-luciferase reporter assays suggest that the reference G allele at rs1247117
310 stimulates transcription more than the alternative A allele, and we suspect this is driven by greater
311 PU.1 binding affinity in these episomal assays. However, our endogenous genetic manipulation
312 that disrupted PU.1 binding altogether at this locus in Nalm6 cells suggest that *CACUL1* and
313 *EIF3A* expression are driven inversely to PU.1 binding. Corroborating these endogenous findings,
314 GTEx eQTL data suggests that the risk G allele harboring great PU.1 affinity is associated with
315 reduced *CACUL1* expression. Moreover, complete disruption of PU.1 binding resulted in greater

316 sensitivity to vincristine, which is consistent with the risk G allele contributing to both greater PU.1
317 binding affinity and vincristine resistance in primary ALL cells from patients. This discrepancy may
318 be due to the ability of PU.1 to act in an activating or repressing manner on gene expression
319 dependent on its genomic context and other transcriptional regulators present (58, 59). While not
320 addressed within the scope of this work, our hypothesis is that MPRA and luciferase reporter
321 assays, which are episomal and utilize a non-native minimal promoter, detected transcription
322 activating PU.1 activities rather than the PU.1 repressive activities we detect within the
323 endogenous locus. Collectively, these observations stress the importance of performing
324 subsequent functional follow-up experimentation within an endogenous sequence context.

325 Although the risk G allele at rs1247117 is associated with decreased *CACUL1* expression,
326 increased risk of MRD after induction chemotherapy and vincristine resistance, it remains unclear
327 why *CACUL1* expression might impact vincristine efficacy. *CACUL1* expression has been
328 correlated with cell cycle progression, and others have shown that *CACUL1* knockdown leads to
329 cell cycle arrest at the G1/S checkpoint (60). However, vincristine can act on microtubules to
330 rapidly kill cells during G1 and later the mitotic spindle to arrest cells during metaphase, so
331 increased *CACUL1* expression may facilitate cell cycle progression, thus increasing the rate at
332 which metaphase mitotic spindles are disrupted by vincristine (61).

333 Our studies also identified *EIF3A* as a target gene for rs1247117 through long-distance
334 promoter looping. *EIF3A* expression has been previously linked to chemotherapeutic sensitivity
335 in both melanoma and lung cancer (62, 63). *EIF3A* expression led to decreased phosphorylation
336 of ERK, supporting the effect of vemurafenib-induced MAP kinase signaling blockade, while
337 *EIF3A* loss led to sustained activation of ERK and therapeutic resistance (62). Interestingly, ERK
338 activation is important in G1/S progression, and therefore it follows that *EIF3A*-dependent
339 inhibition of ERK may support the rapid killing of cells in G1 phase shortly after initial vincristine
340 treatment (61, 64). This notion is supported by significantly greater sensitivity to vincristine of ALL

341 cells harboring greater *EIF3A* expression through disruption of a distal CRE after just 24 hours of
342 treatment.

343 The regulatory variants assayed in this study were originally discovered from GWAS in
344 patient samples, and most of our functional regulatory variant hits from MPRA were present in
345 accessible chromatin sites found only in primary ALL cells from patients. Consequently, these
346 data highlight both substantial differences in the chromatin landscape between immortalized cell
347 lines and primary cells and a limitation of our study that relied on the functional exploration of the
348 top regulatory variant in an ALL cell line model. An optimal approach would be to validate top
349 functional regulatory variants in patient samples; however, this is not currently feasible due to the
350 limited duration of patient sample viability in culture for genetic manipulation. Nonetheless, future
351 implementation of promoter CHiC in patient samples can be used to map gene connectivity of
352 promoter-distal functional regulatory variants found only in primary cells, and these gene targets
353 can then be genetically disrupted in ALL cell line models for functional validation.

354 This translational work represents the largest functional investigation of inherited
355 noncoding variation that is associated with diverse pharmacological traits in ALL. Our study
356 identified hundreds of functional regulatory variants with significant, reproducible, and concordant
357 allele-specific effects on gene expression, and further connected gene regulatory disruptions to
358 differences in chemotherapy response through alterations in antileukemic drug sensitivity in ALL
359 cells. Collectively, these data support the importance of noncoding, gene regulatory disruptions
360 in the pharmacogenomics of ALL treatment. The further functional investigation of these
361 regulatory variants and the discovery of additional inherited variants impacting therapeutic
362 outcome can be used by clinicians to tailor therapies based on a patient's unique genetic makeup
363 through precision or personalized medicine.

364

365

366

367 **MATERIALS AND METHODS**

368

369 **Patient samples and consent**

370 All patients or their legal guardians provided written informed consent. The use of these samples
371 was approved by the institutional review board at St. Jude Children's Research Hospital. Patient
372 samples were obtained from: St. Jude Children's Research Hospital (Memphis, Tennessee) Total
373 Therapy XVI protocol (TOTXVI, NCT00549848) and Total Therapy XVII protocol (TOT17,
374 NCT03117751); Eastern Cooperative Oncology Group (ECOG), The Alliance for Clinical Trials in
375 Oncology, MD Anderson Cancer Center (Houston, Texas) or the University of Chicago (Chicago,
376 Illinois).

377

378 ***Ex vivo* drug resistance assays in primary ALL cells**

379 Primary leukemia cells were isolated from the bone marrow or peripheral blood of newly
380 diagnosed ALL patients from St. Jude Total Therapy XVI protocol (TOTXVI, NCT00549848) and
381 tested for antileukemic drug sensitivity by a 96-hour MTT assay using a range of drug
382 concentrations, as previously described (20, 21). Primary ALL cells were treated with
383 prednisolone (n=320), dexamethasone (n=312), bacterially derived L-asparaginase (n=335),
384 vincristine (n=323), 6-mercaptopurine (n=344) and 6-thioguanine (n=325). Following drug
385 treatment, the lethal concentration resulting in 50% viability (LC₅₀) was calculated for each patient
386 sample.

387

388 **Nalm6 vincristine sensitivity assays**

389 Drug viability assays were performed as previously with slight modification (22). Nalm6 parental
390 cells (WT) and rs1247117 CRE deletion Nalm6 cells (Del) were seeded at 20,000 cells per well
391 in a 96-well plate and co-treated with the indicated concentrations of vincristine (Hospira, 61703-
392 0309-16). Following the indicated duration of incubation, cell viability was measured using the

393 CellTiter-Glo® 2.0 Cell Viability Assay (Promega, G9243). The luminescence was measured
394 using a BioTek Cytation1 cell imaging multimode reader (Agilent). The obtained values were
395 normalized and plotted as % of control. All the experiments were performed in 3 biological
396 replicates having 4-6 technical replicates in each group. Data was plotted as Mean +/- SD.

397

398 **Selection of SNVs impacting treatment outcome in patients from published GWAS**

399 We chose 13 relapse SNVs with $p < 1 \times 10^{-5}$ from (4), 19 ancestry-specific SNVs relapse SNPs
400 associated with relapse in both discovery and replication ALL patient cohorts ($p < 0.05$) from (5)
401 and 3 SNVs associated with persistence of MRD with $p < 1 \times 10^{-6}$ from (3) ($n=35$). In addition, we
402 chose all SNVs ($n=126$) from these GWAS with nominal genome-wide association ($p < 0.05$) but
403 that were additionally associated with *ex vivo* drug resistance phenotypes in primary ALL cells
404 from patients ($p < 0.05$) enrolled in St. Jude Total Therapy XIIIIB, XV and/or XVI protocols
405 (TOTXIIIIB, TOTXV and TOTXVI).

406

407 **Genotyping in primary ALL cells**

408 DNA was extracted from the ALL cells of bone marrow or peripheral blood samples from patients
409 using the Blood and Cell Culture DNA kit (Qiagen). Genotyping was performed using the
410 Affymetrix GeneChip Human Mapping 500K set or the SNP 6.0 array (Affymetrix). Genotypes
411 were called BRLMM algorithm in the Affymetrix GTYPE software
412 (<http://www.affymetrix.com/products/software/specific/gtype.affx>) as previously described (65).

413 We excluded SNVs for call rates $< 95\%$ among patients or minor allele frequencies $< 1\%$.

414

415 **Gene expression profiling in primary ALL cells**

416 Total RNA from primary ALL cells was isolated using RNAeasy Mini kit (Qiagen) and mRNA
417 sequencing using an Illumina HiSeq platform was performed by the Hartwell Center for
418 Bioinformatics and Biotechnology at St. Jude Children's Research Hospital.

419

420 **Quantitative real time PCR (qPCR)**

421 Nalm6 parental (WT) and rs1247117 deleted Nalm6 cells were cultured in RPMI 1640 media. 10
422 million cells were collected from each group in triplicates and resuspended in RLT/BME mixture
423 for total RNA extraction. RNA was isolated using RNeasy Mini Kit (Qiagen #74104).
424 Complimentary DNA synthesis was done using the High-Capacity RNA-to-cDNA Kit (Applied
425 Biosystems #4387406). RT-PCR reactions were prepared using TaqMan Fast Advanced Master
426 Mix (Applied Biosystems #4444557) and TaqMan Gene Expression Assays (Thermo) (CACUL1:
427 Hs00403870_m1, EIF3A: Hs01025769_m1, TBP (endogenous control, Hs00427620_m1). The
428 samples were run on a QuantStudio 3 Real-Time PCR Instrument using the recommended
429 TaqMan Fast Advanced Master Mix PCR conditions.

430

431 **Dual-luciferase reporter assays**

432 A 300-bp of sequence centered on reference or the alternative allele of rs1247117, rs10411204,
433 rs4742260, rs12660691, rs2166631, rs11879659, rs41380646, rs16857207 was cloned upstream
434 of the minimal promoter into the pGL4.23-basic vector. Sequences used in luciferase reporter
435 experiments are shown in **Supplemental File 5**. Nalm6, SUPB15, REH, and 697 cells (10 million
436 cells per replicate, 60 µg plasmid DNA and 6 µg pRL-TK control vector) were used for transfection.
437 Using Neon Transfection system (Thermo Fisher Scientific, MPK5000), the constructs were co-
438 transfected with renilla plasmid to enable normalization of the luciferase signal. 24 h post-
439 transfection, firefly luciferase and renilla luciferase activity was measured using Dual Luciferase
440 Reporter Assay System (Promega, E1960) on a BioTek Cytation1 cell imaging multimode reader
441 (Agilent). The ratio of firefly luciferase to renilla luciferase activity readings reflect the luciferase
442 activity of the reference allele relative to the alternative allele. All experiments were performed in
443 10 samples from each replicate and repeated 2-3 times.

444

445 **Chromatin accessibility mapping in ALL cell models**

446 Fast-ATAC in fresh primary ALL cells from patients (n=120), PDXs (n=3) and in a subset of ALL
447 cell lines (BALL1, CEM, Jurkat and P12-Ichikawa) was performed on 10,000 cells as described
448 in (22, 66). Paired-end Illumina next-generation sequencing of Fast-ATAC libraries was performed
449 at the Hartwell Center for Bioinformatics and Biotechnology at St. Jude Children's Research
450 Hospital. Data were analyzed as in (22). For cryo-preserved primary ALL cells from patients and
451 for B-ALL cell lines, Fast-ATAC data was obtained from the Gene Expression Omnibus
452 (GSE161501 and GSE129066).

453

454 **Massively parallel reporter assays**

455

456 **MPRA Oligo design**

457 Oligo libraries were designed by following previous work with modified protocols (27, 31,
458 67, 68). MPRA oligos ordered from Agilent (230 bp) were structured as follows: 5'-Primer1-
459 enh-KpnI-XbaI-barcode-primer2-3' where primer1 and primer2 are universal primer sites,
460 *enh* denotes the 175bp variant containing region to test for enhancer activity, KpnI and
461 XbaI denote recognition sequences for cut sites, and *barcode* denotes 10-bp tag sequence
462 (see **Sup File 6**). Agilent oligos were resuspended in 100 ul nuclease free water. All 10-
463 bp barcodes for each variant allele used in MPRA are provided in **Supplemental File 7**.

464

465 **MPRA Plasmid Cloning-Input (DNA) Library construction**

466 For plasmid cloning, oligo libraries were amplified by 20 cycles of emulsion PCR (Micellula
467 DNA Emulsion & Purification Kit #E3600, EURx Molecular Biology Products) using
468 Herculase II fusion DNA polymerase (#600675, Agilent), forward and reverse primers (see
469 Sup File 3) to introduce SfiI restriction enzyme sites (GGCCNNNNNGGCC) (NEB) and

470 homology arms to the pMPRA1 plasmid (Cat: #49349, Addgene). Purified PCR products
471 were separated on a 2-4% agarose gel to verify the expected amplification size of 281bp.
472 The pMPRA1 backbone vector was SfiI digested overnight and size selected on a 1%
473 agarose gel. Vector backbone gel extraction was done with the Qiagen gel extraction kit
474 and QIAquick PCR purification kit (28706X4 and 28104). Gibson assembly was used to
475 clone oligos into vector using 79 ng of inserts, 100 ng of digested vector, and 20 ul Gibson
476 assembly 2x master mix (# E2611S, NEB). The reaction was purified using MinElute PCR
477 purification column (Qiagen), and drop analysis was performed with Millipore filters (Type
478 VSWP 0.25 um Millipore #VSWP02500). For the transformation step, we aimed to obtain
479 10x CFU bacterial cells than the distinct promoter-tag combinations (unique sequences)
480 in the oligo library. We transformed the Gibson assembly reactions into MegaX DH10B
481 electrocompetent bacteria (#C6400-03, Invitrogen) using GenePulser II electroporator
482 (Bio-Rad). Plasmids were extracted with Qiagen Maxi Prep Kit. For quality control studies,
483 an aliquot of the isolated plasmid library was digested with SfiI and run on 1 % agarose
484 gel to confirm the presence of inserts. To generate linear enhancer-barcode backbone
485 sequences for reporter insertion, 2 ug of plasmid was digested with KpnI/XbaI. A minimal
486 promoter + truncated eGFP was then ligated to the linearized enhancer-barcode
487 backbone and purified using the Qiagen MinElute PCR purification kit. The ligation was
488 then transformed into 1 vial of MegaX electrocompetent bacteria (#C6400-03, Invitrogen)
489 as before. Plasmids were then extracted using a Qiagen Maxiprep kit as before and the
490 elution was verified as a single size band by gel electrophoresis.

491

492 **MPRA library transfection and sequencing**

493 MPRA plasmid library (10 µg) transfections were done in 10 ALL cell lines having at >95%
494 cell viability (45 million cells x 4 replicates x 10 cell lines) using electroporation with the
495 Neon Transfection system (ThermoFisher). Next day, RNA was harvested using the

496 RNeasy plus mini kit (Cat: #74134, Qiagen) using 4 columns per sample. Once the RNA
497 was isolated, we performed additional DNase digestion using RQ1 RNase-free DNase
498 (Cat: # M6101, Promega). All tubes from the same replicates were combined and added
499 1 volume of 70% ethanol to the combined lysate and mixed well by pipetting. The DNase
500 treated RNA was again purified with RNeasy mini kit and eluted in 60 ul RNase free water.
501 Total RNA was quantified using DeNovix Ds-11 FX instrument. We yielded 16µg-112µg
502 of total RNA from each replicate depending on the cell line used. mRNA purification was
503 performed using Dynabeads mRNA purification kit (Cat: #61006, Invitrogen). mRNA
504 concentration was measured using Qubit HS RNA (Cat: #Q32852, Invitrogen). We yielded
505 from 0.75-2 ug of mRNA in average from each replicate. cDNA was synthesized using
506 three primers (2 uM) cDNA P1, cDNA P4, and cDNA 6, with the Superscript III first-strand
507 synthesis system (Cat: #18080051, Invitrogen) (see **Sup File 6**).

508
509 Final multiplexing of 50ng cDNA and input plasmid DNA (4 aliquots of MPRA
510 plasmid pool that were independently prepared for next-generation sequencing) was
511 carried out using Q5 Hot Start 2x Master Mix (NEB #M0494S), index primers, and
512 Multiplexing primer 1 for 15 cycles of PCR. The reactions were size selected using
513 AMPure XP beads (Cat: #A63881, Beckman Coulter, Indianapolis, IN) and eluted in 20 ul
514 nuclease free water. The final library concentration was measured using Qubit DNA HS.
515 20-40 ng of each library was sequenced on the Illumina NovaSeq (200 million x 150bp
516 paired-end reads per sample) at the Hartwell Center for Bioinformatics and Biotechnology
517 at St. Jude Children's Research Hospital.

518

519 **MPRA sequencing analysis**

520 Following next-generation sequencing, the MPRA sequence data was trimmed to
521 contain only barcode sequences without allowing for any mismatches and read counts

522 were determined for all barcodes. To identify significant allele-specific effects mpralm (69)
523 was performed on RNA and DNA barcode counts after merging RNA or DNA counts
524 across all barcodes for each allele.

525

526 **Promoter capture Hi-C**

527 Arima promoter capture HiC (Arima: A510008, A303010, A302010) was performed according to
528 the manufacturers provided instructions using unspecified proprietary buffers, solutions,
529 enzymes, and reagents. Briefly, 10 million ALL cells were harvested, suspended in 5ml RT PBS
530 which was brought to 2% formaldehyde by adding 37% methanol-stabilized paraformaldehyde for
531 a 10-minute fixation. The amount of fixed cell suspension equal to 5 μ g of cell DNA was used for
532 HiC. Cells were lysed with Lysis Buffer and conditioned with Conditioning Solution before their
533 DNA was digested in a cocktail consisting of Buffer A, Enzyme 1, and Enzyme 2. The digested,
534 fixed chromatin was biotinylated using Buffer B and Enzyme B before being ligated using Buffer
535 C and Enzyme C. The fixed, biotinylated, ligated DNA was then subjected to reversal of
536 crosslinking and digestion of proteins before being purified. 100ul containing 1500ug of purified
537 large proximally ligated DNA was fragmented for 24 cycles (30s on/ 30s off) using a Diagenode
538 Bioruptor Plus bath sonicator. The fragmented DNA was then subjected to two-sided size
539 selection targeting fragments between 200-600bp using AMPure XP DNA purification beads. Size
540 selected DNA was then subjected to biotin enrichment using T1 streptavidin beads. Bead bound,
541 enriched HiC DNA was then subjected to Arima library prep. Briefly, the sample underwent end
542 repair followed by adapter ligation, at which point the sample was then subjected to 10 cycles of
543 PCR amplification. The library DNA was then purified using AMPure XP DNA purification beads.
544 The HiC library was then subjected to Arima promoter capture enrichment. The library was
545 precleared of biotinylated DNA using T1 streptavidin beads before being subjected to promoter
546 enrichment with biotinylated RNA probes. After washing, the captured fragments were then
547 amplified an additional 13 PCR cycles. These libraries were submitted for deep sequencing on

548 an Illumina Nova-seq where >200M 150bp paired-end reads were obtained. Analysis of Promoter
549 capture HiC data was performed using the Arima CHiC pipeline (v1.5,
550 <https://github.com/ArimaGenomics/CHiC>). Briefly, this pipeline uses HiCUP v0.8.0 for mapping
551 and quality assessment of promoter capture HiC data and CHiCAGO to identify significant looping
552 interactions in the promoter capture HiC data using 5kb resolution and adj. p < 0.05 (70, 71).

553

554 **PU.1 *in vitro* binding affinity assay**

555 DNA pulldown assay was adopted from previous article and performed using a modified protocol
556 (72). Briefly, biotinylated ssDNA probes were ordered via custom synthesis from IDT with their
557 non-biotinylated reverse complement sequences. The DNA probe sequences used in the
558 experiment are listed in **Supplemental File 8**. The probes containing reference and alternative
559 alleles featuring the nucleotide of interest (rs1247117) and its flanking +/-12 bp nucleotides were
560 arranged side-by-side in tandem for a total of 50 bp each. Probes were annealed by combining
561 biotinylated probes and the non-biotinylated reverse complement at 1.5M excess (50µM:75µM)
562 with an equal volume of 2x annealing buffer (10 mM Tris pH 8.0, 100 mM NaCl, 2 mM EDTA) and
563 incubating at 98C for 10 minutes before cooling at RT overnight. To isolate nuclear lysate, we
564 washed 75 million cells in 5 mL PBS and pelleted at 500xg for 3 min at RT. The cells were
565 resuspended in 2 mL of homogenization buffer (1M KCl, 1M MgCl₂, 1M HEPES, 0.5 M EGTA, 1x
566 Halt protease inhibitors, Thermo Fisher 78429) and passed through 26 G needle 10x. The nuclei
567 were pelleted, and the supernatant was discarded. The nuclei were gently washed with an
568 additional 1ml of homogenization buffer, pelleted, and the supernatant discarded again. Washed
569 nuclei were suspended in 300 µL of SKT buffer (1M HEPES, 1M MgCl₂, Glycerol, 1 M KCl, EDTA,
570 0.1% triton). To extract nuclear proteins 33ul of 3M NaCl was added and samples were vortexed
571 every 2-3 min on ice for the next 20 min. Insoluble nuclear material was pelleted at 13000 rpm for
572 10 min at 4°C. To make a nuclear lysate master mix 1100µg of nuclear lysate was transferred to
573 a fresh tube with 22µg nonspecific DNA (11µg Poly (dl-dC), (Thermo Fisher 20148E) + 11µg

574 Poly(dA:dT), (Cell Signaling Technologies)) and brought to a final volume of 1320 μ L with protein
575 binding buffer (PBB, 150 mM NaCl, 0.25% NP40, 50 mM Tris pH 8.0, 1 mM DTT and EDTA free
576 protease inhibitors (Roche). Reference and Alternative allele annealed biotinylated DNA probes
577 (500pMol) were prebound to 20ul Streptavidin T1 Dynabeads (#65601, Thermo) in 600 μ L DNA
578 binding buffer (DBB, 1 M NaCl, 0.05% NP40, 10 mM TRIS, pH 8.0 and 1 mM ETDA) rotating at
579 4°C for 30 min. Streptavidin T1 beads bound to probes were washed 1x DBB, 2x PBB on ice and
580 600 μ L of nuclear lysate master mix (prepared above) was added to reference or alternative allele
581 bead-bound probe tubes. Lysate and probe-bound beads were rotated for 90 min at 4°C and
582 washed 3x PBB, and 2x PBS by pipetting up and down 5x each wash. Proteins were eluted in 40
583 μ L 1x LDS sample buffer (+10% BME) (Cat: #NP0007, Invitrogen) by heating at 99.9°C in a
584 thermal mixer at 1200 rpm for 10 min (#13687712, Thermo). Samples were subsequently
585 assessed by western blot as previously described (22) using anti-rabbit PU.1 antibody (#2258S,
586 Cell Signaling Technology).

587

588 **ChIP-seq**

589 RNA polymerase II ChIP-seq data were generated by first fixing 20 million Nalm6 cells in 1%
590 formaldehyde (diluted from sigma F87750) at room temp for 10 minutes. Crosslinking was
591 stopped with the addition of 2.5M glycine to a concentration of 0.125M, and the cells were then
592 washed in ice-cold PBS. 5 μ g anti-RNA polymerase II CTD repeat YSPTSPS (phospho S5)
593 antibody [4H8] (ab5408, lot: GR3264797-1) was prebound to 200ul of protein G dynabeads
594 (Invitrogen 10003D) overnight in 0.5% BSA in PBS. Fixed cell pellets (20M cells) were suspended
595 in 1ml Farnham lysis buffer (5mM PIPES pH 8, 85mM KCl, 0.5% NP40, 1x protease inhibitors
596 (Roche 11836170001)) and passed through a 18G needle 10x. Nuclei were pelleted and
597 resuspended in 275ul of RIPA buffer (1x PBS, 1% NP40, 0.5% Sodium Deoxycholate, 0.1% SDS,
598 1x protease inhibitors) and sonicated on high power in 1.5ml tubes for 25 minutes (30s on/ 30s
599 off) using a Diagenode Bioruptor Plus. 5% Input samples were taken from sonicated material and

600 the remaining sonicated material was added to the pre-bound antibody/protein G beads to rotate
601 overnight at 4C. The next day the supernatant was discarded, and the beads were washed 5x
602 with ice cold LiCl buffer (100mM Tris pH 7.5, 500mM LiCl, 1% NP40, 1% sodium deoxycholate)
603 and 1x with ice cold TE buffer (10mM Tris pH 7.5, 1mM EDTA). Samples were eluted from the
604 washed beads using room temperature IP elution buffer (1% SDS, 0.1 M NaHCO₃) at 65C for 1hr,
605 vortexing every 15 minutes. The elution was then incubated at 65C overnight to reverse
606 crosslinks. The next day DNA was purified using the QIAquick PCR purification kit (Qiagen
607 28104). DNA quantification was performed using the PicoGreen assay (Molecular Probes,
608 Eugene, OR, P-7581). Sequencing libraries were generated from ChIP and input DNA by using
609 the KAPA Hyper Prep kit (Roche, Basel, Switzerland, # 7962363001) according to the included
610 manufacturer's specifications, and quality was determined by using the Agilent TapeStation with
611 D1000 screentape. Then, >50M 50-bp paired-end reads per sample were generated on the
612 NovaSeq 6000. Reads were quality checked using fastqc (v0.11.5) and trimmed using trimgalore
613 (v0.4.4) before being mapped to the hg19 reference genome using bowtie2 (v2.2.9). Sam files
614 were converted to bam format using samtools (v1.2), which were sorted using picard (v1.141).
615 Duplicates were removed using picard and mitochondrial reads were removed using samtools.
616 For visualization, bam files from replicates were merged using samtools and converted to bigwig
617 format using deeptools (v3.5.0). For peak calling, we used macs2 (v2.1.1), and only considered
618 peaks called in both samples. H3K27ac ChIP-seq data in Nalm6 cells was obtained from the
619 Gene Expression Omnibus (GSE161501).

620

621 **PU.1 CUT and RUN**

622 CUT and RUN data were generated using the Epicyper Cutana CUT&RUN kit v3.0 (14-1048)
623 according to the manufacturers provided instructions. Briefly 500k NALM6 cells per reaction were
624 bound to 10µl of provided activated ConA beads in 0.2ml PCR tubes. Bead-bound cells were
625 suspended in Antibody Buffer (Wash buffer with 0.1% digitonin, 0.5mM Spermidine, 2mM EDTA,

626 and 1x HALT protease inhibitors) and incubated with 1ul PU.1 antibody (Cell Signaling 2258) or
627 IgG (Epiccypher 13-0042k) overnight on a nutator mixer at 4C. The next day after washing, pAG
628 MNase was bound and targeted digestion was carried out for 2 hours at 4C. Digestion was
629 stopped using 33 μ l Stop buffer + 1 μ l (0.5ng) E.coli spike-in DNA and then cleaved DNA were
630 released for 10 minutes at 37C. DNA were then purified for library preparation using the included
631 purification kit. >30M 75bp paired end reads were generated per sample using the Illumina
632 Novaseq. The Nextflow CUT and RUN pipeline (v2.0) was used in spike-in mode to assess quality
633 and map reads to the HG19 (human) and K12-MG1655 (E. coli) reference genomes (73, 74). The
634 spike-in normalized .bam files from Nextflow were exported to Easeq (v1.111), where peaks were
635 called against the IgG sample using adaptive local thresholding ($p < 1 \times 10^{-5}$, $FDR < 1 \times 10^{-5}$,
636 $\text{Log}_2(\text{Fold Change}) > 1$, merge within = 100bp, window size = 100bp) (75). Data shown are spike-
637 in normalized bigwig files generated in Nextflow.

638

639 **CRISPR/Cas9 deletion**

640 rs1247117 deletions in Nalm6 were generated using CRISPR-Cas9 technology. In brief, one
641 million Nalm6 cells were transiently transfected with precomplexed ribonuclear proteins (RNPs)
642 consisting of 100pmol of each chemically modified sgRNA (Synthego, see **Sup File 9**), 35pmol
643 of Cas9 protein (St. Jude Protein Production Core), and 3 μ g of ssODN (Alt-R modifications, IDT;
644 see Table 1 below) via nucleofection (Lonza, 4D-Nucleofector™ X-unit) using solution P3 and
645 program CV-104 in a large (100ul) cuvette according to the manufacturer's recommended
646 protocol. Three days post-nucleofection, genomic DNA was harvested via crude lysis and used
647 for PCR amplification (see Sup File 3 for primers). The presence of the desired deletion was
648 confirmed via gel electrophoresis and sequencing. To validate disruptions, targeted amplicons
649 were generated using gene specific primers with partial Illumina adapter overhangs and
650 sequenced. Cell pellets of approximately 10,000 cells were lysed and used to generate gene
651 specific amplicons with partial Illumina adapters in PCR#1. Amplicons were indexed in PCR#2

652 and pooled with targeted amplicons from other loci to create sequence diversity. Additionally, 10%
653 PhiX Sequencing Control V3 (Illumina) was added to the pooled amplicon library prior to running
654 the sample on an Miseq Sequencer System (Illumina) to generate paired 2 X 250bp reads.
655 Samples were demultiplexed using the index sequences, fastq files were generated, and NGS
656 analysis was performed using CRIS.py (76).

657

658 **Statistical analysis**

659

660 **GWAS for *ex vivo* drug resistance in primary ALL cells**

661 Multiple linear regression was used with log-transformed LC₅₀ as dependent variable and
662 genotype as independent variable. Genotypes were coded as 0, 1 and 2. Patient genetic
663 ancestry was included in the linear model as covariates. Two tailed p-values were
664 generated using a Wald test. All statistical analysis was performed in R v4.0.2. All SNVs
665 with a p-value < 0.05 were further evaluated to determine if were eQTLs.

666

667 **Gene expression profiling in primary ALL cells**

668 Genotype information was correlated with RNA expression across patient samples to
669 identify expression quantitative trait loci (eQTLs). eQTL mapping was performed using
670 multiple linear regression and log-transformed FPKM gene expression as dependent
671 variable and genotype as independent variable. Genotypes were coded as 0, 1 and
672 2. Patient genetic ancestry was included in the linear model as covariates. Two tailed p-
673 values were generated using a Wald test. All statistical analysis was performed in R v4.0.2.

674

675

676

677

678 **Promoter-capture Hi-C**

679 Statistical modeling of chromatin looping was performed as described in the publication
680 introducing the CHiCAGO tool (71). We used 5kb resolution and an adjusted P value cutoff
681 of 0.05.

682

683 **PU.1 binding affinity assay**

684 A one-tailed student's T-test was used to test the hypothesis that the alternate allele would
685 show less affinity for PU.1 binding.

686

687 **PU.1 CUT and RUN**

688 PU.1 peaks were called against the IgG sample using adaptive local thresholding ($p <$
689 1×10^{-5} , $FDR < 1 \times 10^{-5}$, $\text{Log}_2(\text{Fold Change}) > 1$, merge within = 100bp, window size = 100bp)
690 (75). Data shown are spike-in normalized bigwig files generated in Nextflow.

691

692 **Nalm6 vincristine drug sensitivity assays**

693 Individual Students T-tests were performed at each dose in each time point to test the
694 hypothesis that deletion of the regulatory region containing rs1247117 in Nalm6 cells
695 would alter sensitivity to vincristine.

696

697 **Comparisons between groups of variant MPRA data**

698 When comparing promoter-associated and distal promoter-connected variant MPRA data,
699 the Mann-Whitney test was used. Comparisons within the distal promoter-connected
700 variants from introns, UTRs and distal intergenic regions were carried out using the
701 Kruskal-Wallis test with Dunn's correction for multiple comparisons.

702

703

704 **CACUL1 and EIF3A qPCR and Dual Luciferase reporter assays**

705 Student's T tests were used to determine the significance of differences between samples.

706

707 **Acknowledgments**

708 We would like to thank the Hartwell Center at St. Jude for ATAC-seq, ChIP-seq and MPRA next-
709 generation sequencing. We would also like to thank Jeremy Hunt and Brandon Smart for technical
710 support. This work was performed at St. Jude Children's Research Hospital and was also
711 conducted in part by the ECOG-ACRIN Cancer Research Group. This work is supported by the
712 National Cancer Institute (R01CA234490, P30CA021765, U10CA180820, UG1CA232760
713 and UG1CA189859), the National Institute of General Medical Studies (P50GM115279) and the
714 American Lebanese Syrian Associated Charities (ALSAC). The content is solely the responsibility
715 of the authors and does not necessarily represent the official views of the National Institutes of
716 Health.

717

718

719 **Author contributions:**

720 Conceptualization: DS, RJM

721 Methodology: DS, KR Bhattarai, RJM

722 Investigation: KR Bhattarai, RJM, DCF, JDD, BPB

723 Formal Analysis: KR Bhattarai, RJM, KR Barnett, WY, CC

724 Data Curation: KR Barnett, WY, KRC, CC

725 Sample Acquisition: CSM, EJ, EP, MRL, SMK, WS, HI, SJ, CHP, MVR, JJY,
726 WEE

727 Visualization: RJM, KR Bhattarai

728 Supervision: DS, RJM

729 Writing—original draft: DS, KR Bhattarai, RJM

730 Writing—review & editing: KR Bhattarai, RJM, KR Barnett, DCF, JDD, BPB,

731 WY, KRC, CSM, EJ, EP, MRL, SMK, WS, HI, SJ, CHP, CC, MVR, JJY, WEE, DS

732

733 **Competing interests:** No conflicts of interest to disclose.

734

735 **Data and materials availability:** Cell line ATAC-seq, Promoter capture HiC, RNA pol II,
736 and PU.1 genomic binding data are available on GEO (GSE224204). Previously
737 published H3K27Ac ChIP-seq data, "GSE175482_Nalm6_H3K27ac_0hr_merged.bw",
738 are located in GEO GSE175482. Patient sample and PDX associated genomics data are
739 available upon request via St. Jude cloud (<https://www.stjude.cloud/>).

740

741

742 References

- 743
- 744 1. S. P. Hunger, C. G. Mullighan, Acute Lymphoblastic Leukemia in Children. *N Engl J*
745 *Med* **373**, 1541-1552 (2015).
- 746 2. A. Oriol *et al.*, Outcome after relapse of acute lymphoblastic leukemia in adult patients
747 included in four consecutive risk-adapted trials by the PETHEMA Study Group.
748 *Haematologica* **95**, 589-596 (2010).
- 749 3. J. J. Yang *et al.*, Genome-wide interrogation of germline genetic variation associated with
750 treatment response in childhood acute lymphoblastic leukemia. *JAMA* **301**, 393-403
751 (2009).
- 752 4. J. J. Yang *et al.*, Genome-wide association study identifies germline polymorphisms
753 associated with relapse of childhood acute lymphoblastic leukemia. *Blood* **120**, 4197-
754 4204 (2012).
- 755 5. S. E. Karol *et al.*, Genetics of ancestry-specific risk for relapse in acute lymphoblastic
756 leukemia. *Leukemia* **31**, 1325-1332 (2017).
- 757 6. H. Cave *et al.*, Clinical significance of minimal residual disease in childhood acute
758 lymphoblastic leukemia. European Organization for Research and Treatment of Cancer--
759 Childhood Leukemia Cooperative Group. *N Engl J Med* **339**, 591-598 (1998).
- 760 7. M. J. Borowitz *et al.*, Minimal residual disease detection in childhood precursor-B-cell
761 acute lymphoblastic leukemia: relation to other risk factors. A Children's Oncology
762 Group study. *Leukemia* **17**, 1566-1572 (2003).
- 763 8. E. Coustan-Smith *et al.*, Clinical importance of minimal residual disease in childhood
764 acute lymphoblastic leukemia. *Blood* **96**, 2691-2696 (2000).
- 765 9. J. Zhou *et al.*, Quantitative analysis of minimal residual disease predicts relapse in
766 children with B-lineage acute lymphoblastic leukemia in DFCI ALL Consortium Protocol
767 95-01. *Blood* **110**, 1607-1611 (2007).
- 768 10. S. H. R. Lee *et al.*, Pharmacotypes across the genomic landscape of pediatric acute
769 lymphoblastic leukemia and impact on treatment response. *Nat Med* **29**, 170-179 (2023).
- 770 11. R. Pieters *et al.*, Relation of cellular drug resistance to long-term clinical outcome in
771 childhood acute lymphoblastic leukaemia. *Lancet* **338**, 399-403 (1991).
- 772 12. A. G. Bosanquet, Correlations between therapeutic response of leukaemias and in-vitro
773 drug-sensitivity assay. *Lancet* **337**, 711-714 (1991).
- 774 13. W. S. Hwang, L. M. Chen, S. H. Huang, C. C. Wang, M. T. Tseng, Prediction of
775 chemotherapy response in human leukemia using in vitro chemosensitivity test. *Leuk Res*
776 **17**, 685-688 (1993).
- 777 14. R. Pieters, G. J. Kaspers, E. Klumper, A. J. Veerman, Clinical relevance of in vitro drug
778 resistance testing in childhood acute lymphoblastic leukemia: the state of the art. *Med*
779 *Pediatr Oncol* **22**, 299-308 (1994).
- 780 15. E. Klumper *et al.*, In vitro cellular drug resistance in children with relapsed/refractory
781 acute lymphoblastic leukemia. *Blood* **86**, 3861-3868 (1995).
- 782 16. T. Hongo, S. Yajima, M. Sakurai, Y. Horikoshi, R. Hanada, In vitro drug sensitivity
783 testing can predict induction failure and early relapse of childhood acute lymphoblastic
784 leukemia. *Blood* **89**, 2959-2965 (1997).
- 785 17. G. J. Kaspers *et al.*, In vitro cellular drug resistance and prognosis in newly diagnosed
786 childhood acute lymphoblastic leukemia. *Blood* **90**, 2723-2729 (1997).

- 787 18. M. L. Den Boer *et al.*, Patient stratification based on prednisolone-vincristine-
788 asparaginase resistance profiles in children with acute lymphoblastic leukemia. *J Clin*
789 *Oncol* **21**, 3262-3268 (2003).
- 790 19. B. M. Frost *et al.*, Increased in vitro cellular drug resistance is related to poor outcome in
791 high-risk childhood acute lymphoblastic leukaemia. *Br J Haematol* **122**, 376-385 (2003).
- 792 20. A. Holleman *et al.*, Gene-expression patterns in drug-resistant acute lymphoblastic
793 leukemia cells and response to treatment. *N Engl J Med* **351**, 533-542 (2004).
- 794 21. R. J. Autry *et al.*, Integrative genomic analyses reveal mechanisms of glucocorticoid
795 resistance in acute lymphoblastic leukemia. *Nat Cancer* **1**, 329-344 (2020).
- 796 22. B. P. Bergeron *et al.*, Epigenomic profiling of glucocorticoid responses identifies cis-
797 regulatory disruptions impacting steroid resistance in childhood acute lymphoblastic
798 leukemia. *Leukemia* **36**, 2374-2383 (2022).
- 799 23. R. P. Smith, E. T. Lam, S. Markova, S. W. Yee, N. Ahituv, Pharmacogene regulatory
800 elements: from discovery to applications. *Genome Med* **4**, 45 (2012).
- 801 24. M. R. Luizon, N. Ahituv, Uncovering drug-responsive regulatory elements.
802 *Pharmacogenomics* **16**, 1829-1841 (2015).
- 803 25. N. J. Sakabe, D. Savic, M. A. Nobrega, Transcriptional enhancers in development and
804 disease. *Genome Biol* **13**, 238 (2012).
- 805 26. M. A. White, C. A. Myers, J. C. Corbo, B. A. Cohen, Massively parallel in vivo enhancer
806 assay reveals that highly local features determine the cis-regulatory function of ChIP-seq
807 peaks. *Proc Natl Acad Sci U S A* **110**, 11952-11957 (2013).
- 808 27. P. Kheradpour *et al.*, Systematic dissection of regulatory motifs in 2000 predicted human
809 enhancers using a massively parallel reporter assay. *Genome Res* **23**, 800-811 (2013).
- 810 28. D. Savic *et al.*, Promoter-distal RNA polymerase II binding discriminates active from
811 inactive CCAAT/ enhancer-binding protein beta binding sites. *Genome Res* **25**, 1791-
812 1800 (2015).
- 813 29. J. C. Ulirsch *et al.*, Systematic Functional Dissection of Common Genetic Variation
814 Affecting Red Blood Cell Traits. *Cell* **165**, 1530-1545 (2016).
- 815 30. M. Kircher *et al.*, Saturation mutagenesis of twenty disease-associated regulatory
816 elements at single base-pair resolution. *Nat Commun* **10**, 3583 (2019).
- 817 31. J. Choi *et al.*, Massively parallel reporter assays of melanoma risk variants identify MX2
818 as a gene promoting melanoma. *Nat Commun* **11**, 2718 (2020).
- 819 32. A. C. Joslin *et al.*, A functional genomics pipeline identifies pleiotropy and cross-tissue
820 effects within obesity-associated GWAS loci. *Nat Commun* **12**, 5253 (2021).
- 821 33. R. Ajore *et al.*, Functional dissection of inherited non-coding variation influencing
822 multiple myeloma risk. *Nat Commun* **13**, 151 (2022).
- 823 34. E. Long *et al.*, Massively parallel reporter assays and variant scoring identified functional
824 variants and target genes for melanoma loci and highlighted cell-type specificity. *Am J*
825 *Hum Genet* **109**, 2210-2229 (2022).
- 826 35. J. C. Klein *et al.*, Functional testing of thousands of osteoarthritis-associated variants for
827 regulatory activity. *Nat Commun* **10**, 2434 (2019).
- 828 36. S. Smemo *et al.*, Obesity-associated variants within FTO form long-range functional
829 connections with IRX3. *Nature* **507**, 371-375 (2014).
- 830 37. G. T. Consortium *et al.*, Genetic effects on gene expression across human tissues. *Nature*
831 **550**, 204-213 (2017).

- 832 38. B. van Steensel, J. Dekker, Genomics tools for unraveling chromosome architecture. *Nat*
833 *Biotechnol* **28**, 1089-1095 (2010).
- 834 39. B. Mifsud *et al.*, Mapping long-range promoter contacts in human cells with high-
835 resolution capture Hi-C. *Nat Genet* **47**, 598-606 (2015).
- 836 40. L. E. Montefiori *et al.*, A promoter interaction map for cardiovascular disease genetics.
837 *Elife* **7**, (2018).
- 838 41. B. M. Javierre *et al.*, Lineage-Specific Genome Architecture Links Enhancers and Non-
839 coding Disease Variants to Target Gene Promoters. *Cell* **167**, 1369-1384 e1319 (2016).
- 840 42. E. Thulson *et al.*, 3D chromatin structure in chondrocytes identifies putative osteoarthritis
841 risk genes. *Genetics* **222**, (2022).
- 842 43. C. Shi *et al.*, Chromatin Looping Links Target Genes with Genetic Risk Loci for
843 Dermatological Traits. *J Invest Dermatol* **141**, 1975-1984 (2021).
- 844 44. V. Perez-Andreu *et al.*, Inherited GATA3 variants are associated with Ph-like childhood
845 acute lymphoblastic leukemia and risk of relapse. *Nat Genet* **45**, 1494-1498 (2013).
- 846 45. V. Perez-Andreu *et al.*, A genome-wide association study of susceptibility to acute
847 lymphoblastic leukemia in adolescents and young adults. *Blood* **125**, 680-686 (2015).
- 848 46. H. Xu *et al.*, Inherited coding variants at the CDKN2A locus influence susceptibility to
849 acute lymphoblastic leukaemia in children. *Nat Commun* **6**, 7553 (2015).
- 850 47. E. A. Hungate *et al.*, A variant at 9p21.3 functionally implicates CDKN2B in paediatric
851 B-cell precursor acute lymphoblastic leukaemia aetiology. *Nat Commun* **7**, 10635 (2016).
- 852 48. J. Vijayakrishnan *et al.*, Genome-wide association study identifies susceptibility loci for
853 B-cell childhood acute lymphoblastic leukemia. *Nat Commun* **9**, 1340 (2018).
- 854 49. J. Vijayakrishnan *et al.*, Identification of four novel associations for B-cell acute
855 lymphoblastic leukaemia risk. *Nat Commun* **10**, 5348 (2019).
- 856 50. J. Vijayakrishnan *et al.*, The 9p21.3 risk of childhood acute lymphoblastic leukaemia is
857 explained by a rare high-impact variant in CDKN2A. *Sci Rep* **5**, 15065 (2015).
- 858 51. J. B. Studd *et al.*, Genetic and regulatory mechanism of susceptibility to high-
859 hyperdiploid acute lymphoblastic leukaemia at 10p21.2. *Nat Commun* **8**, 14616 (2017).
- 860 52. H. Yang *et al.*, Noncoding genetic variation in GATA3 increases acute lymphoblastic
861 leukemia risk through local and global changes in chromatin conformation. *Nat Genet* **54**,
862 170-179 (2022).
- 863 53. X. Zhao *et al.*, Molecular Mechanisms of ARID5B-Mediated Genetic Susceptibility to
864 Acute Lymphoblastic Leukemia. *J Natl Cancer Inst* **114**, 1287-1295 (2022).
- 865 54. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of
866 native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-
867 binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218 (2013).
- 868 55. J. D. Diedrich *et al.*, Profiling chromatin accessibility in pediatric acute lymphoblastic
869 leukemia identifies subtype-specific chromatin landscapes and gene regulatory networks.
870 *Leukemia*, (2021).
- 871 56. S. L. Klemm, Z. Shipony, W. J. Greenleaf, Chromatin accessibility and the regulatory
872 epigenome. *Nat Rev Genet* **20**, 207-220 (2019).
- 873 57. P. J. Skene, S. Henikoff, An efficient targeted nuclease strategy for high-resolution
874 mapping of DNA binding sites. *Elife* **6**, (2017).
- 875 58. H. Hosokawa *et al.*, Transcription Factor PU.1 Represses and Activates Gene Expression
876 in Early T Cells by Redirecting Partner Transcription Factor Binding. *Immunity* **49**, 782
877 (2018).

- 878 59. H. A. Carey *et al.*, Enhancer variants reveal a conserved transcription factor network
879 governed by PU.1 during osteoclast differentiation. *Bone Res* **6**, 8 (2018).
- 880 60. N. Chen *et al.*, CAC1 knockdown reverses drug resistance through the downregulation of
881 P-gp and MRP-1 expression in colorectal cancer. *PLOS ONE* **14**, e0222035 (2019).
- 882 61. A. Kothari, W. N. Hittelman, T. C. Chambers, Cell Cycle–Dependent Mechanisms
883 Underlie Vincristine-Induced Death of Primary Acute Lymphoblastic Leukemia Cells.
884 *Cancer Research* **76**, 3553-3561 (2016).
- 885 62. S.-L. Jiang *et al.*, The Downregulation of eIF3a Contributes to Vemurafenib Resistance
886 in Melanoma by Activating ERK via PPP2R1B. *Frontiers in Pharmacology* **12**, (2021).
- 887 63. J.-Y. Yin *et al.*, Effect of eIF3a on Response of Lung Cancer Patients to Platinum-Based
888 Chemotherapy by Regulating DNA Repair. *Clinical Cancer Research* **17**, 4600-4609
889 (2011).
- 890 64. S. Torii, T. Yamamoto, Y. Tsuchiya, E. Nishida, ERK MAP kinase in G1 cell cycle
891 progression and cancer. *Cancer Science* **97**, 697-702 (2006).
- 892 65. D. French *et al.*, Acquired variation outweighs inherited variation in whole genome
893 analysis of methotrexate polyglutamate accumulation in leukemia. *Blood* **113**, 4512-4520
894 (2009).
- 895 66. M. R. Corces *et al.*, Lineage-specific and single-cell chromatin accessibility charts human
896 hematopoiesis and leukemia evolution. *Nature Genetics* **48**, 1193-1203 (2016).
- 897 67. A. Melnikov *et al.*, Systematic dissection and optimization of inducible enhancers in
898 human cells using a massively parallel reporter assay. *Nature biotechnology* **30**, 271-277
899 (2012).
- 900 68. A. Melnikov, X. Zhang, P. Rogov, L. Wang, T. S. Mikkelsen, Massively parallel reporter
901 assays in cultured mammalian cells. *JoVE (Journal of Visualized Experiments)*, e51719
902 (2014).
- 903 69. L. Myint, D. G. Avramopoulos, L. A. Goff, K. D. Hansen, Linear models enable
904 powerful differential activity analysis in massively parallel reporter assays. *BMC*
905 *Genomics* **20**, 209 (2019).
- 906 70. S. W. Wingett *et al.*, HiCUP: pipeline for mapping and processing Hi-C data.
907 *F1000Research* **4**, 1310 (2015).
- 908 71. J. Cairns *et al.*, CHiCAGO: robust detection of DNA looping interactions in Capture Hi-
909 C data. *Genome Biology* **17**, (2016).
- 910 72. M. M. Makowski *et al.*, An interaction proteomics survey of transcription factor binding
911 at recurrent TERT promoter mutations. *Proteomics* **16**, 417-426 (2016).
- 912 73. P. A. Ewels *et al.*, The nf-core framework for community-curated bioinformatics
913 pipelines. *Nature Biotechnology* **38**, 276-278 (2020).
- 914 74. P. P. Ewels, Alexander; Fillinger, Sven; Patel, Harshil; Alneberg, Johannes; Wilm,
915 Andreas; Garcia, Maxime Ulysse; Di Tommaso, Paolo; Nahnsen, Sven, The nf-core
916 framework for community-curated bioinformatics pipelines. *Zenodo*, (2022).
- 917 75. M. Lerdrup, J. V. Johansen, S. Agrawal-Singh, K. Hansen, An interactive environment
918 for agile analysis and visualization of ChIP-sequencing data. *Nature Structural &*
919 *Molecular Biology* **23**, 349-357 (2016).
- 920 76. J. P. Connelly, S. M. Pruett-Miller, CRIS.py: A Versatile and High-throughput Analysis
921 Program for CRISPR-based Genome Editing. *Sci Rep* **9**, 4194 (2019).
- 922
- 923
- 924

925 **Figure Legends**

926

927 **Figure 1: Identification and mapping of regulatory variants impacting the**
928 **pharmacogenomics of ALL treatment. (A)** SNVs of interest from GWAS were pursued based
929 on association with *ex vivo* chemotherapeutic drug resistance in primary ALL cells from patients
930 and/or treatment outcome. Dex= dexamethasone, Pred= prednisolone, VCR= vincristine, 6MP=
931 6-mercaptopurine, 6TG= 6-thioguanine, LASP= L-asparaginase. **(B)** GWAS SNVs were
932 combined with ALL disease susceptibility control GWAS SNVs and SNVs in high LD ($R^2 > 0.8$) and
933 **(C)** mapped to accessible chromatin sites in ALL cell lines, ALL PDXs and primary ALL cells from
934 patients.

935

936 **Figure 2: MPRA identifies regulatory variants with allele-specific effects on gene**
937 **expression. (A)** Diagram describing design of MPRA (also see **Methods**). **(B)** Distribution of
938 significant changes in allele-specific transcriptional activity across all SNVs. **(C)** Number of
939 MPRA SNVs showing significant (Adj. $p < 0.05$) changes in allele-specific transcriptional activity
940 in each ALL cell line. **(D)** Pair-wise linear correlation between changes in allele-specific
941 transcriptional activity for all significant (Adj. $p < 0.05$) changes across all cell lines. R^2 correlation
942 and p-value are provided.

943

944 **Figure 3: Identification of functional regulatory variants with reproducible and**
945 **concordant effects in allele-specific stimulation of transcriptional activity. (A)** 556 of the
946 1696 SNVs assayed are functional regulatory variants with reproducible ($FDR < 0.05$ in > 2 cell
947 lines) and concordant (same directionality in > 2 cell lines) changes in allele-specific activity. **(B)**
948 Plot showing the distribution of \log_2 -adjusted activity between alternative (Alt) and reference
949 (Ref) alleles across 556 functional regulatory variants. **(C)** Pie chart shows how many functional
950 regulatory variants map to open chromatin in diverse ALL cell models. 210 of the 556 functional

951 regulatory variants are found in accessible chromatin sites that was identified in an ALL cell line.
952 **(D)** Top hits from the 210 functional regulatory variants found in accessible chromatin in ALL cell
953 lines were orthogonally validated by luciferase reporter assays. Data show significant correlation
954 between the allele-specific effects detected by MPRA and dual-luciferase reporter assays.

955
956 **Figure 4: Promoter CHiC identifies target genes of functional regulatory variants. (A)** Data
957 show the number of functional regulatory variants mapping to open chromatin in cell lines that
958 associate directly with promoters or that are distally promoter-connected via promoter CHiC. **(B)**
959 MPRA data show promoter-connected functional regulatory variants in accessible chromatin
960 exhibit stronger effects on allele specific activity than promoter-associated functional regulatory
961 variants. **(C)** Amongst distally promoter-connected functional regulatory, variants that map to
962 intronic and distal intergenic sequences showed greater activity than those in UTRs. **(D-E)** Data
963 show the ranked allele-specific activity distribution of MPRA data for **(D)** promoter-associated
964 functional regulatory variants and **(E)** distally promoter-connected functional regulatory variants.

965
966 **Figure 5: Functional exploration of rs1247117 in B-ALL cells. (A)** IGV genome browser
967 image in Nalm6 cells showing the genomic context, chromatin accessibility, and *EIF3A* promoter
968 connectivity using promoter CHiC of the top functional regulatory variant, rs1247117, with the
969 highest allele-specific MPRA activity. Genomic binding profiles are also shown for RNA
970 polymerase II (RNA Pol2), histone H3 lysine 27 acetylation (H3K27Ac) and PU.1. **(B)** rs1247117
971 lies in a PU.1 binding motif. The human genome reference sequence, Nalm6 genome
972 sequence, location of rs1247117 and PU.1 position weight matrix is shown. **(C)** Design of
973 biotinylated DNA-probes for *in vitro* rs1247117 pulldown. **(D)** Biotinylated DNA pulldown shows
974 rs1247117 allele-dependent enrichment of PU.1 binding. Blot shown is representative of two
975 independent experiments. P-value from densitometric quantification of two blots is shown. **(E)**
976 Diagram on the left showing the genomic context of the rs1247117 CRE deletion in Nalm6 cells

977 in relation to chromatin accessibility, PU.1 binding and rs1247117. Black bar represents ATAC-
978 seq peak, green bar represents PU.1 peak, and red bar represents region deleted using
979 CRISPR/Cas9 genome editing. Gel shows validation of deletion using primers flanking deleted
980 region. Arrow points to PCR fragment with deletion in heterogeneous Nalm6 cell pools
981 harboring deletion compared to wild-type parental Nalm6 cells. **(F-G)** *CACUL1* **(F)** and *EIF3A*
982 **(G)** expression is upregulated upon deletion of the CRE containing rs1247117. RT-qPCR data
983 show the mean +/- SEM of three independent experiments. **(H)** Drug sensitivity data comparing
984 survival of wild-type parental Nalm6 cells and Nalm6 cells with rs1247117 CRE deletion after
985 vincristine (VCR) treatment for 24 (n=3), 48 (n=3) and 72 (n=3) hours. Vincristine concentration
986 is provided below. *, p<0.05; **, p<0.01; ***, p<0.001; ****, p<0.0001. Data show the mean +/-
987 SEM relative to untreated cells.
988

Figure 1

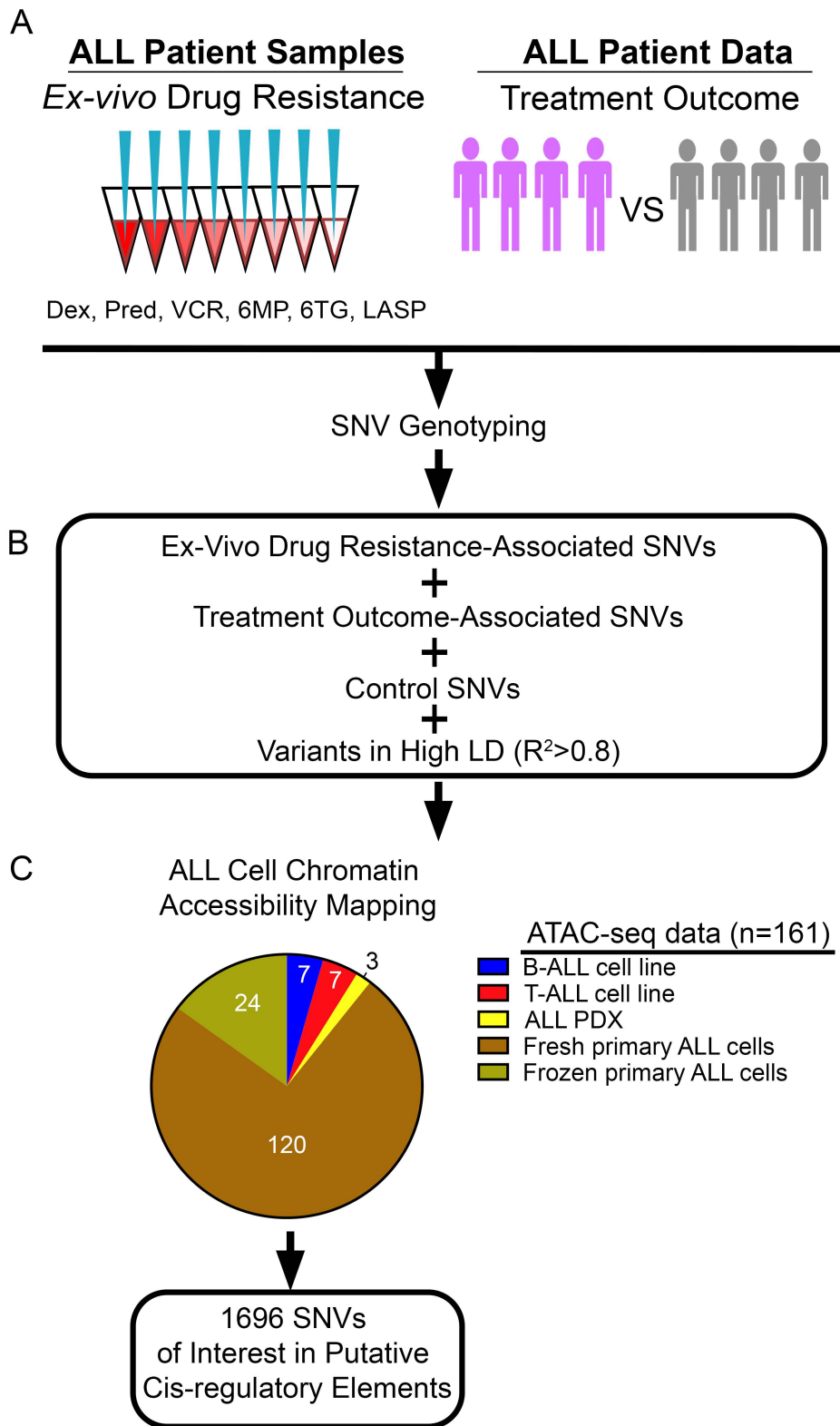


Figure 2

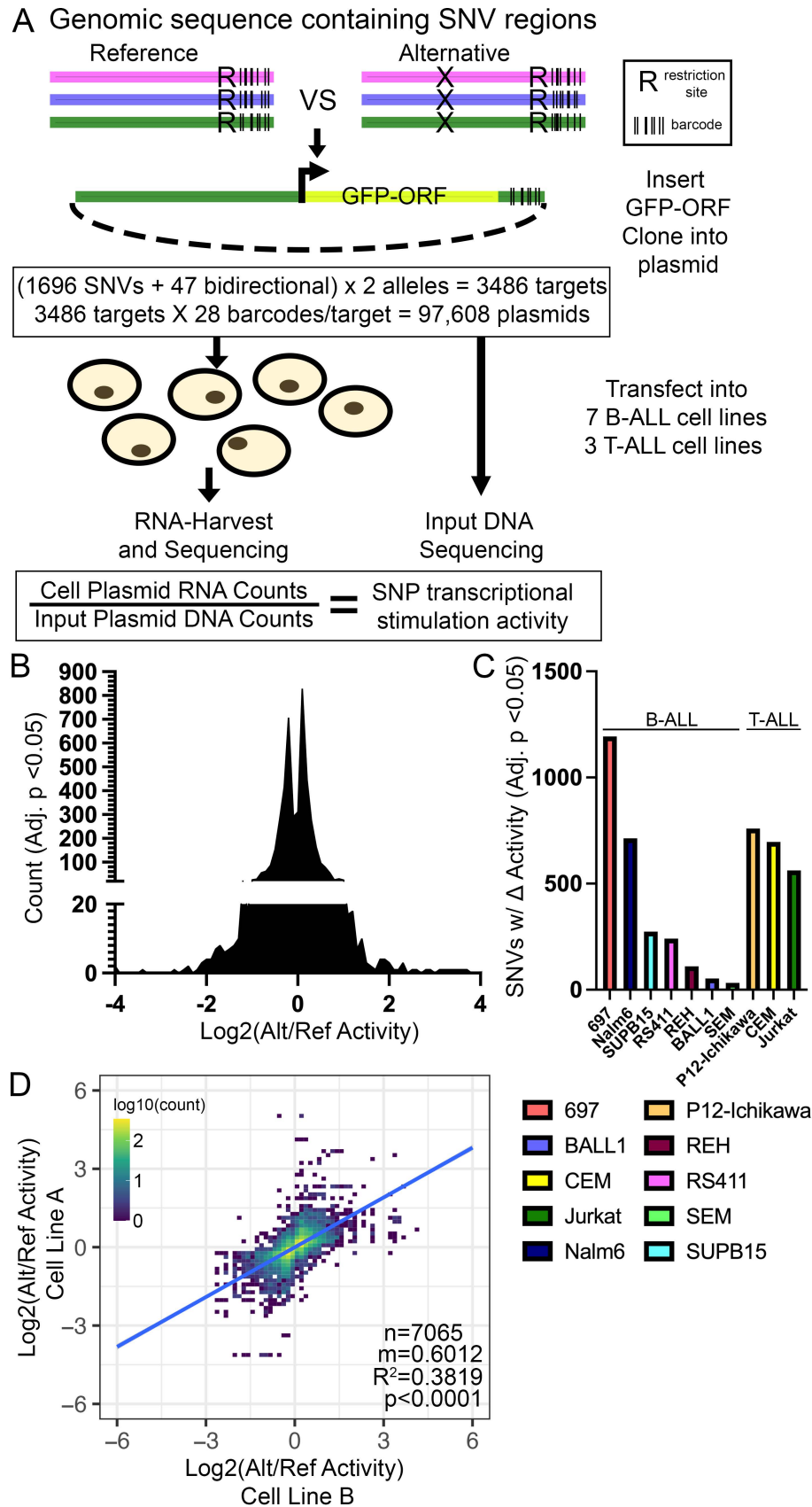


Figure 3

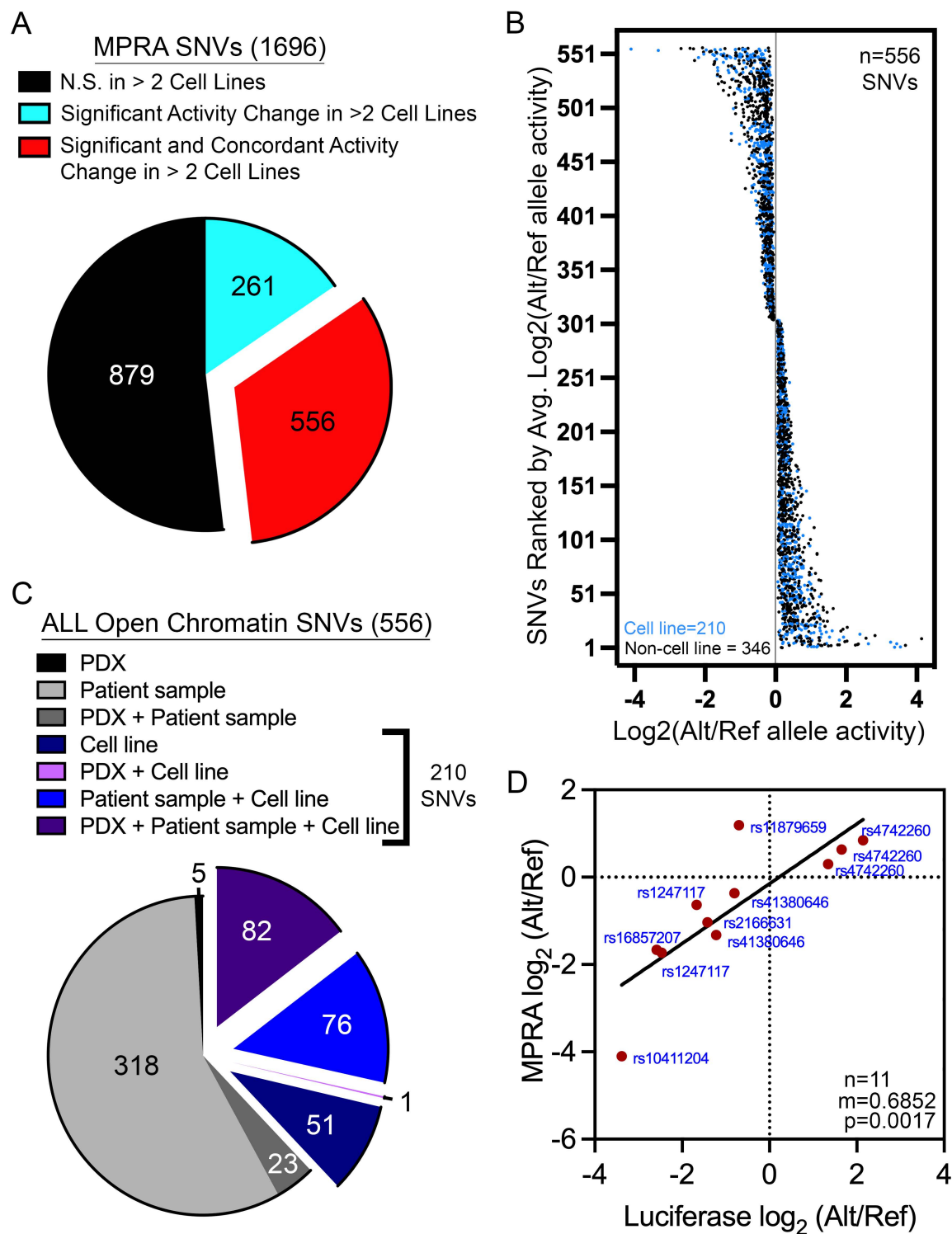


Figure 4

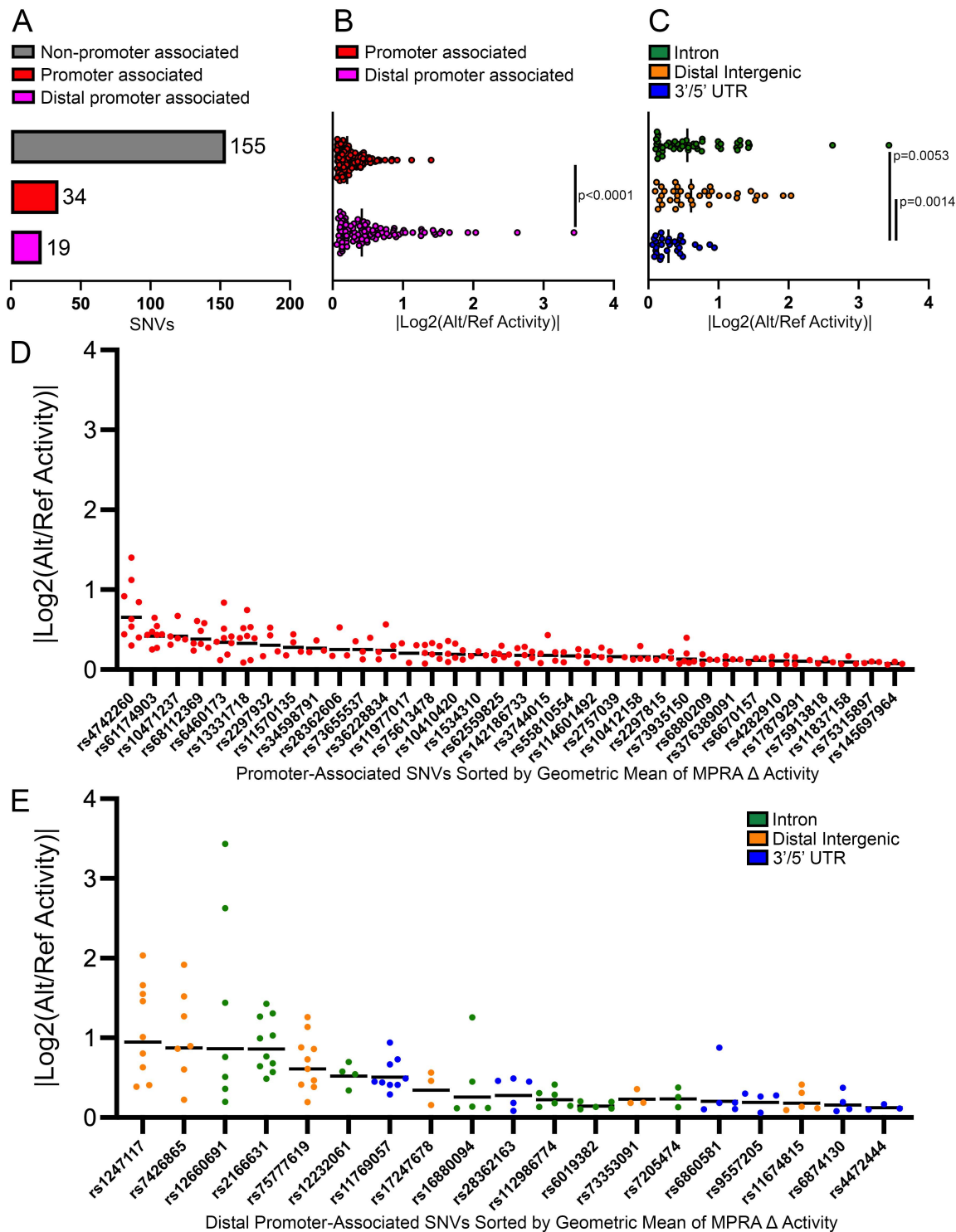
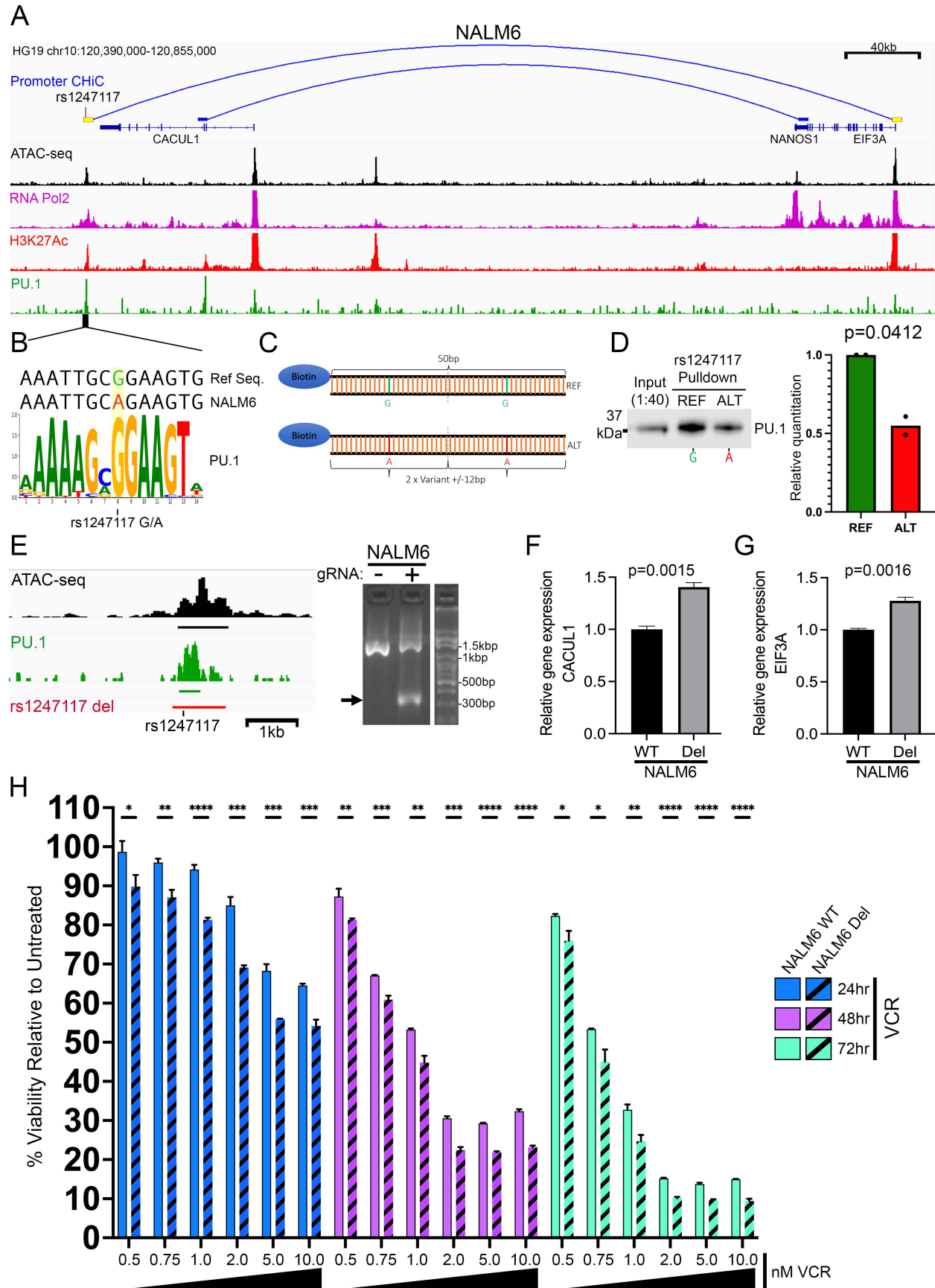


Figure 5



Supplemental Materials

Supplemental Figures

Supplemental File 1 – SNP IDs, Sentinel SNVs, and associated Phenotypes of MPRA SNPs

Supplemental File 2 – MPRA source data

Supplemental File 3 – 556 reproducible and concordant regulatory variants

Supplemental File 4 – 53 reproducible and concordant regulatory variants at promoters or connected to promoters as determined by promoter CHiC.

Supplemental File 5 – Reference and alternative allele sequences used in dual-luciferase reporter assays

Supplemental File 6 – Nucleic acids associated with MPRA method

Supplemental File 7 – MPRA barcode sequences

Supplemental File 8 – PU.1 binding affinity assay probe sequences

Supplemental File 9 – CRISPR/Cas9 rs1247117 deletion pool sequences

SUPPLEMENTAL FIGURES

Functional investigation of inherited noncoding genetic variation impacting the pharmacogenomics of childhood acute lymphoblastic leukemia treatment

Kashi Raj Bhattarai, PhD^{1,2†}, Robert J. Mobley, PhD^{1,2†}, Kelly R. Barnett, PhD^{1,2}, Daniel C. Ferguson, PhD^{1,2}, Baranda S. Hansen, MS^{3,4}, Jonathan D. Diedrich, PhD^{1,2}, Brennan P. Bergeron, PhD^{1,2,5}, Wenjian Yang, PhD^{1,2}, Kristine R. Crews, PharmD^{1,2}, Christopher S. Manning MBA⁶, Elias Jabbour, MD⁷, Elisabeth Paietta, PhD⁸, Mark R. Litzow, MD⁹, Steven M. Kornblau, MD⁷, Wendy Stock, MD¹⁰, Hiroto Inaba, MD, PhD^{1,11}, Sima Jeha, MD^{1,11}, Ching-Hon Pui, MD^{1,11}, Cheng Cheng, PhD¹², Shondra M. Pruett-Miller, PhD^{3,4}, Mary V. Relling, PharmD^{1,2}, Jun J. Yang, PhD^{1,2,5,13}, William E. Evans, PharmD^{1,2} and Daniel Savic, PhD^{1,2,5,13,*}

¹Hematological Malignancies Program, St. Jude Children's Research Hospital, Memphis, TN

²Department of Pharmacy and Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, TN

³Center for Advanced Genome Engineering, St. Jude Children's Research Hospital, Memphis, TN 38105, USA.

⁴Department of Cell and Molecular Biology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA.

⁵Graduate School of Biomedical Sciences, St. Jude Children's Research Hospital, Memphis, TN

⁶Alliance Hematologic Malignancy Biorepository; Clara D. Bloomfield Center for Leukemia Outcomes Research, Columbus, OH 43210, USA

⁷Department of Leukemia, The University of Texas MD Anderson Cancer Center, Houston, TX

⁸Albert Einstein College of Medicine, New York, NY

⁹Division of Hematology, Department of Medicine, Mayo Clinic, Rochester, MN 55905, USA.

¹⁰Comprehensive Cancer Center, University of Chicago Medicine, Chicago, IL

¹¹Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN

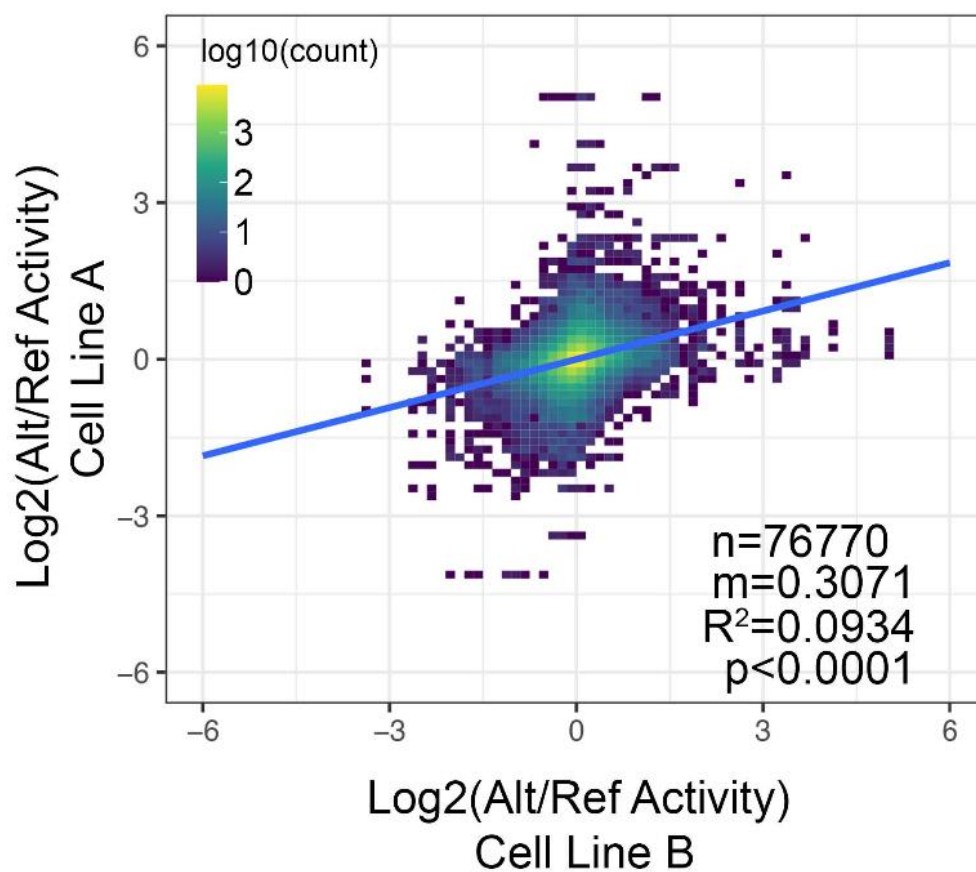
¹²Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN

¹³Integrated Biomedical Sciences Program, University of Tennessee Health Science Center, Memphis, TN

†Authors contributed equally to this work

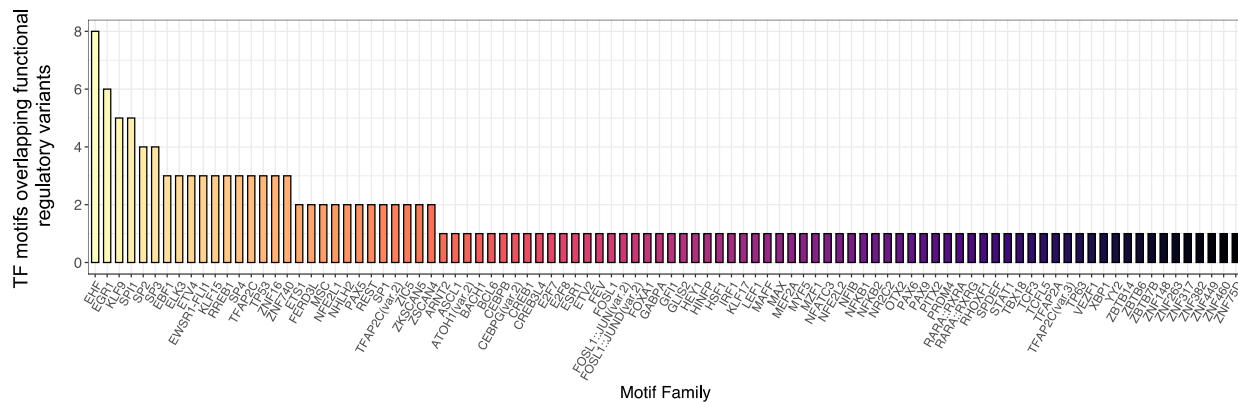
*Corresponding author

Supplemental Figure 1. MPRA activity comparisons among all cell lines. Pair-wise linear correlation between changes in allele-specific transcriptional activity for all measurements and across all cell lines. R^2 correlation and p-value are provided.

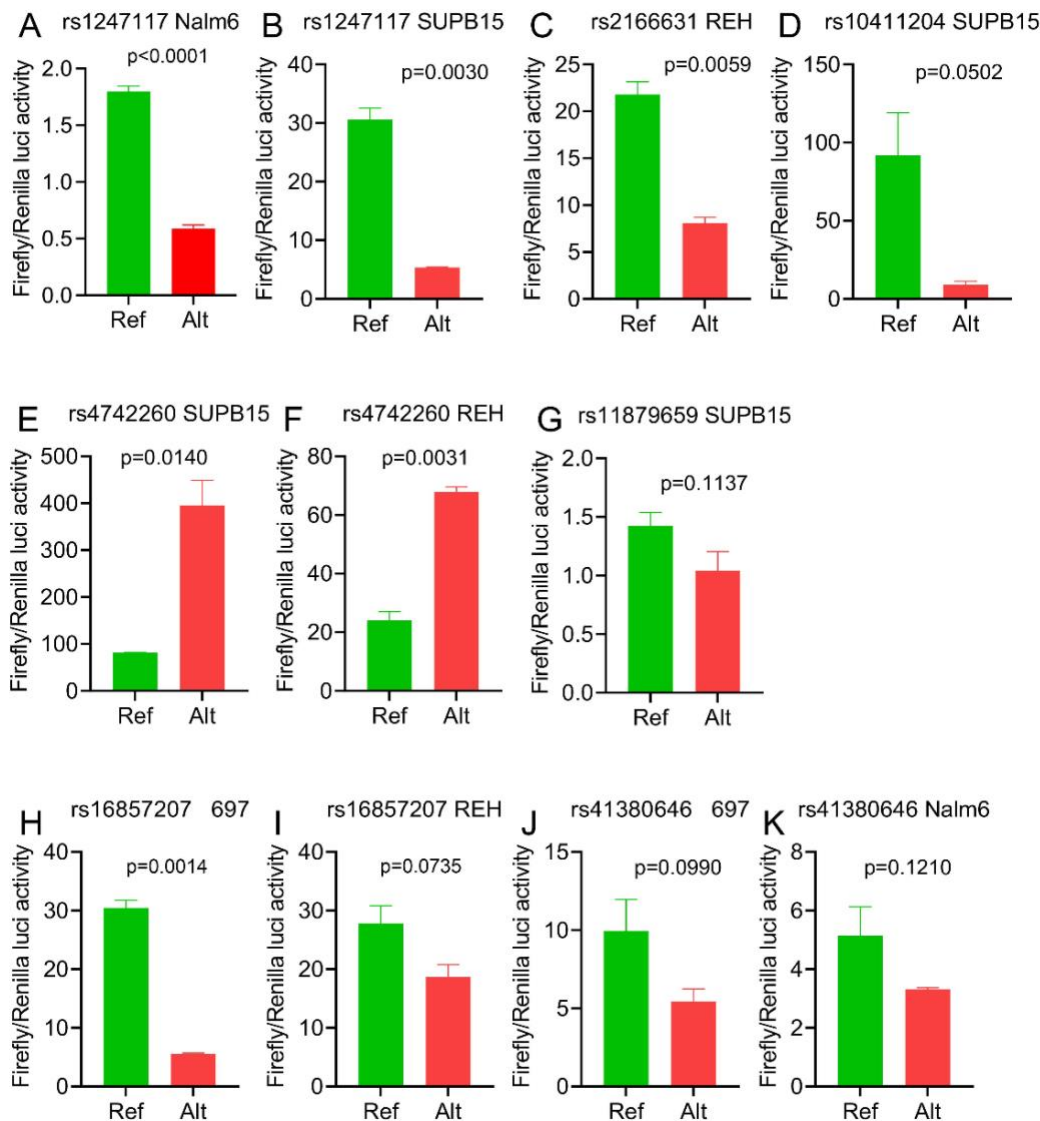


Supplemental Figure 2. Transcription factor footprints at functional regulatory variants.

Transcription factor (TF) footprints identified at 54 of 556 functional regulatory variants are shown and ranked by the total number of motifs identified.



Supplemental Figure 3. Dual-luciferase reporter assay validation of the indicated functional regulatory variant. (A-K) Dual-luciferase reporter assays comparing the reference (Ref, in green) and alternate (Alt, in red) alleles ability to drive luciferase expression is depicted. Variant rs number and the ALL cell line the luciferase reporter assay was tested in is provided. Data show the mean \pm SEM of three (A) or two (B-K) independent experiments. P-value is calculated using a student's t-test.



Supplemental Figure 4. Chromatin accessibility at rs1247117 in primary ALL cells. (A) IGV genome browser image of ATAC-seq chromatin accessibility spanning rs1247117 in diverse molecular subtypes of ALL is provided. **(B)** PU.1 footprint analysis comparing normalized ATAC-seq cut count signal for all bound PU.1 sites (red) compared to unbound (blue) sites across all primary ALL cells from patients. **(C)** Primary ALL cells with SNV genotype information were analyzed (n=69). Normalized ATAC-seq read counts in heterozygous (GA) primary ALL cells (n=12) at rs1247117 compared to homozygous (AA) primary ALL cells (n=57). Mann Whitney U test p-value is provided. **(D)** Normalized ATAC-seq read counts per allele in primary ALL cells for G allele (n=12) and A allele (n=69). Normalized counts for G and A alleles are shown. Mann Whitney U test p-value is provided.

