Epigenomic mapping in B-cell acute lymphoblastic leukemia identifies transcriptional regulators and noncoding variants promoting distinct chromatin architectures

Kelly R. Barnett, PhD[1,2], Robert J. Mobley, PhD[1,2], Jonathan D. Diedrich, PhD[1,2], Brennan P. Bergeron, PhD[1,2,3], Kashi Raj Bhattarai, PhD[1,2], Wenjian Yang, PhD[1,2], Kristine R. Crews, PharmD[1,2], Christopher S. Manring, MBA[4], Elias Jabbour, MD[5], Elisabeth Paietta, PhD[6], Mark R. Litzow, MD[7], Steven M. Kornblau, MD[5], Wendy Stock, MD[8], Hiroto Inaba, MD, PhD[1,9], Sima Jeha, MD[1,9], Ching-Hon Pui, MD[1,9], Charles G. Mullighan, MBBS (Hons), MSc, MD[1,10], Mary V. Relling, PharmD[1,2], Jun J. Yang, PhD[1,2,3,11], William E. Evans, PharmD[1,2] and Daniel Savic, PhD[1,2,3,11,*]

[1] Hematological Malignancies Program, St. Jude Children's Research Hospital, Memphis, TN 38105, USA.
[2] Department of Pharmacy and Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, TN 38105, USA.
[3] Graduate School of Biomedical Sciences, St. Jude Children's Research Hospital, Memphis, TN 38105, USA.
[4] Alliance Hematologic Malignancy Biorepository; Clara D. Bloomfield Center for Leukemia Outcomes Research, Columbus, OH 43210, USA
[5] Department of Leukemia, The University of Texas M. D. Anderson Cancer Center, Houston, TX, USA
[6] Department of Oncology, Montefiore Medical Center, Bronx, NY 10467, USA.
[7] Division of Hematology, Department of Medicine, Mayo Clinic, Rochester, MN 55905, USA.
[8] University of Chicago Comprehensive Cancer Center, Chicago, IL 60637, USA.
[9] Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA.
[10] Department of Pathology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA.
[11] Integrated Biomedical Sciences Program, University of Tennessee Health Science Center, Memphis, TN 38105, USA.

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

Key words: Acute lymphoblastic leukemia, ATAC-seq, chromatin accessibility, gene regulation, gene regulatory network, transcription factor, transcription factor footprints, genetic variation, ATAC-QTLs

SUMMARY

B-cell lineage acute lymphoblastic leukemia (B-ALL) is comprised of diverse molecular subtypes and while transcriptional and DNA methylation profiling of B-ALL subtypes has been extensively examined, the accompanying chromatin landscape is not well characterized for many subtypes. We therefore mapped chromatin accessibility using ATAC-seq for 10 B-ALL molecular subtypes in primary ALL cells from 154 patients. Comparisons with B-cell progenitors identified candidate B-ALL cell-of-origin and AP-1-associated *cis*-regulatory rewiring in B-ALL. *Cis*-regulatory rewiring promoted B-ALL-specific gene regulatory networks impacting oncogenic signaling pathways that perturb normal B-cell development. We also identified that over 20% of B-ALL accessible chromatin sites exhibit strong subtype enrichment, with transcription factor (TF) footprint profiling identifying candidate TFs that maintain subtype-specific chromatin architectures. Over 9000 inherited genetic variants were further uncovered that contribute to variability in chromatin accessibility among individual patient samples. Overall, our data suggest that distinct chromatin architectures are driven by diverse TFs and inherited genetic variants which promote unique gene regulatory networks that contribute to transcriptional differences among B-ALL subtypes.

HIGHLIGHTS

- Pro-B progenitor cells as the most common cell-of-origin for B-ALL
- AP-1 TF-associated *cis*-regulatory rewiring in B-ALL
- Subtype-specific accessible chromatin signatures representing 20% of all B-ALL sites
- Role for distinct TFs in promoting subtype-specific chromatin architectures
- Thousands of inherited genetic variants identified impacting chromatin state

INTRODUCTION

Acute lymphoblastic leukemia (ALL) is derived from B- and T-cell lineage precursor cells and is the most common childhood cancer [1]. A majority of acute lymphoblastic leukemias are derived from B-cell lineages (B-ALL) that are comprised of distinct molecular subtypes characterized by unique chromosomal lesions, including aneuploidy, translocations, gene fusions, point mutations and other chromosomal rearrangements that drive leukemogenesis [2]. Numerous studies have identified extensive heterogeneity in transcriptomes [3,4] and DNA methylomes [5,6] among B-ALL subtypes in large patient cohorts, but there is limited understanding of chromatin landscapes. Here we provide an extensive survey of accessible chromatin state and *cis*-regulatory element activity in primary B-ALL cells from over 150 patients across the United States.

Chromatin accessibility or open chromatin is a hallmark of active *cis*-regulatory elements that control spatial and temporal gene expression [7]. Because ALL typically involves mutations (*PAX5*-altered), complex rearrangements (*DUX4*-rearranged, *PAX5*-altered, *ZNF384*-rearranged, etc.) and/or oncogenic gene fusions (*ETV6::RUNX1*, *TCF3::PBX1*, *KMT2A*-rearranged, etc.) of transcription factor (TF) genes as well as disruptions of *cis*-regulatory elements[8], chromatin accessibility maps can provide valuable information to better understand the leukemogenic process. Accessible chromatin sites can be mapped using transposases by performing assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) [9,10]. Although DNase treatment has also been used [11], one key advantage of ATAC-seq is the low sample input requirements compared to DNase-based assays. This makes ATAC-seq an attractive assay for mapping open chromatin in primary cells from patients wherein sample availability is limited. Additionally, chromatin accessibility allows for identification of bound TFs through an examination of TF footprints which are defined by a depletion in DNA transposition [12] or DNase [13] cleavage events within regions of accessible chromatin signal. As a result, the underlying TF-binding gene regulatory networks that promote chromatin accessibility and differential gene expression can be predicted.

Previous large-scale studies of chromatin accessibility in primary cells have predominantly focused on distinct cell types [10,14] or distinct tumor types and locations [15,16]. Therefore, large-scale analyses aimed to better understand chromatin state in a single heterogeneous malignancy are currently lacking. To address this knowledge gap, we mapped chromatin accessibility in fresh primary ALL cells from 154 patients across 10 molecular subtypes of B-ALL (*BCR::ABL1*, *DUX4*-rearranged, *ETV6::RUNX1*, high hyperdiploid, low hypodiploid, *KMT2A*-rearranged, *BCR::ABL1*-like (Ph-like), *PAX5*-altered, *TCF3::PBX1*, *ZNF384*-rearranged) and B-other patient samples. Notably, these subtypes span the entire spectrum of clinical prognoses, including patients with excellent (*DUX4*-rearranged, *ETV6::RUNX1*, high hyperdiploid), good (*TCF3::PBX1*), intermediate (*ZNF384*-rearranged,

106 *PAX5*-altered) and poor (*BCR::ABL1*, low hypodiploid, *KMT2A*-rearranged and Ph-like) prognosis. We
107 also mapped histone H3 lysine 27 acetylation (H3K27ac) enrichment using ChIP-seq in a subset of
108 these patient samples to additionally infer functional activity.
109 Using ATAC-seq chromatin accessibility and histone profiling in primary ALL cells, we mapped
110 *cis*-regulatory element activity in B-ALL. In complement to chromatin accessibility profiling, we identified
111 thousands of chromatin loops targeting promoters in multiple B-ALL cell lines to better inform linkages
112 of *cis*-regulatory elements to cognate genes. We coupled these maps to transcription factor (TF)
113 footprints at accessible chromatin sites to identify key TFs and gene regulatory networks across B-ALL
114 samples and within distinct B-ALL subtypes. Our results identified extensive chromatin reprogramming
115 between B-cell progenitors and B-ALL, as well as extensive heterogeneity in accessible chromatin
116 landscapes among B-ALL subtypes. Specifically, we uncovered a focused subset of over 42,000 B-ALL
117 open chromatin sites exhibiting extensive subtype-enrichment and subtype-enriched TF binding events.
118 Notably, these sites can predict and classify B-ALL samples with 86% cross-validation accuracy. We
119 additionally explored the impact of inherited genetic variation on chromatin state and delineated over
120 9000 ATAC-seq chromatin accessibility quantitative trait loci (ATAC-QTLs) in B-ALL cells, including a
121 subset that alter neighboring gene expression. Using the largest accessible chromatin accessibility
122 dataset for B-ALL to date, our data collectively support substantial subtype-specificity in chromatin
123 accessibility that is driven in part by distinct TFs, as well as pronounced inter-individual heterogeneity in
124 chromatin state through inherited genetic variants. Our work further supports the role of these distinct
125 chromatin architectures in establishing unique gene regulatory networks that impact gene expression
126 and B-ALL cell biology.
127
128
129

130 # RESULTS

131 **Chromatin accessibility profiles of B-ALL patient samples spanning multiple subtypes**

132 ATAC-seq using the Fast-ATAC [10] method was performed on recently-harvested primary ALL cells from
133 154 patients spanning 10 B-ALL molecular subtypes (*BCR::ABL1*, *DUX4*-rearranged, *ETV6::RUNX1*,
134 high hyperdiploid, low hypodiploid, *KMT2A*-rearranged, Ph-like, *PAX5*-altered, *TCF3::PBX1*, *ZNF384*-
135 rearranged) and B-other samples (**Table S1**) from diverse medical centers, research groups and
136 clinical trials networks across the United States (see **Methods**). To identify high-confidence sites, we
137 identified ATAC-seq peak summits using subtype merged data and selected only loci reproducible
138 among unmerged individual patients. Using this approach we identified 110,468 accessible chromatin

139 fsites, on average, in each B-ALL subtype (range= 71,797–142,498), with 217,240 merged sites
140 identified in total representing the final genomic regions of interest (**Figure 1A, Table S2**).
141 Using H3K27ac ChIP-seq data generated from a subset of 11 B-ALL patient samples, as well as
142 primary B-ALL cell H3K27ac, H3K4me1 and H3K27me3 ChIP-seq data from the Blueprint Epigenome
143 Consortium (https://www.blueprint-epigenome.eu/), we determined that nearly all open chromatin sites
144 mapped to regions containing only active histone marks (H3K27ac and/or H3K4me1, 89.6%;
145 H3K27ac= 3.3%, H3K4me1=34% and H3K4me1+H3K27ac=52.3%) or regions with bivalent marks
146 suggesting a poised chromatin state (H3K27ac and/or H3K4me1 and H3K27me3, 8.9%), compared to
147 only 1.5% of ATAC-seq sites that mapped to regions solely harboring repressive chromatin
148 (H3K27me3; **Figure 1B**). Because these histone modifications are typically found at transcriptional
149 enhancers and promoters[17-20], these findings suggest that these accessible chromatin regions are B-
150 ALL *cis*-regulatory elements implicated in gene regulation.
151 In most cases, these candidate *cis*-regulatory elements map within intergenic or intragenic loci
152 with unclear gene targets. Therefore, to better inform gene connectivity we produced chromatin looping
153 data using promoter capture Hi-C [21] across seven B-ALL cell lines (697, BALL1, Nalm6, REH, RS411,
154 SEM and SUPB15) to complement B-ALL patient chromatin accessibility profiles. Collectively, across
155 the B-ALL cell lines we detected approximately 400,000 chromatin loops, with approximately 50% of
156 the 217,240 chromatin accessible regions of interest intersecting with a promoter loop, including 15,929
157 chromatin accessible sites that looped to a cancer implicated gene set (**Figure 1C**) [22,23] . In many
158 instances, large domains of extensive chromatin looping are present, which with chromatin accessibility
159 and active histone marks emphasize the gene regulatory networks present across B-ALL patient
160 samples (e.g., **Figures 1D** and **1E**).
161
162 **Chromatin accessibility identifies Pro-B cell-of-origin for most B-ALL patient samples**
163 To better understand chromatin remodeling during leukemogenesis we sought a comparison of
164 chromatin accessibility between B-ALL and B-cell progenitors. Moreover, although it is widely accepted
165 that the B-ALL cell-of-origin is a B-cell precursor, exactly which precursor is not always clear,
166 particularly at the chromatin accessibility level [24]. To resolve this uncertainty, we examined publicly
167 available ATAC-seq data from several human B-cell progenitors[10,25] (**Figure 2A**). When comparing
168 chromatin accessibility signal between B-cell progenitor groups, we identified a set of approximately
169 42,344 genomic loci which demonstrate a chromatin accessibility enrichment or depletion trend for a B-
170 cell progenitor (**Figure 2B, Table S3**). We refer to these chromatin loci as B-progenitor identity loci due
171 their distinct patterning across B-progenitor differentiation and are likely representations of stage-
172 specific gene regulatory programs.

173 Next, we examined patient B-ALL cell chromatin accessibility across these B-progenitor identity
174 loci. When plotting chromatin accessibility signal as a heatmap comparing B-cell progenitors and B-ALL
175 patient samples, a high degree of similarity was observed with prePro-B cells and Pro-B cells (**Figure**
176 **2B**). Further, when applying the K-nearest neighbor classification model previously trained on B-
177 progenitor identity loci the majority of B-ALL samples classified as either prePro-B or Pro-B (**Figures**
178 **2C** and **2D**). However, prePro-B cells have been reported to be an extremely rare population beyond
179 embryonic and fetal development [25]. Overall, Pro-B cells demonstrate the most similarity to B-ALL cells
180 at the chromatin accessibility level when focusing specifically on B-cell precursor defining loci,
181 emphasizing this precursor B-cell as a common cell-of-origin for B-ALL.
182
183 **Extensive differences in chromatin accessibility between B-ALL and Pro-B cells**
184 To better understand chromatin remodeling during leukemogenesis we next compared accessible
185 chromatin sites between B-ALL and Pro-B cells (n=3) and uncovered 42,661 differentially accessible
186 chromatin sites (DAS) exhibiting lesser or greater accessibility in B-ALL samples (**Figures 3A** and **3B**;
187 **Figure S1** and **Table S4**). Ontology analysis focusing strictly on DAS with higher chromatin
188 accessibility in B-ALL indicated an enrichment for sites associated with genes involved with toll-like
189 receptor signaling, interleukin production, metabolism (acetyl-CoA production) and cell proliferation
190 (**Figure 3C**). Enriched ontology terms were frequently present at multiple fold change thresholds of
191 input B-ALL DAS (**Table S5)**.
192 In addition to profiling differential chromatin accessibility, global transcription factor (TF) binding
193 was also compared between B-ALL and Pro-B cells. To identify differential TF binding, we performed
194 genome-wide TF footprint profiling [12] using 810 TF motifs comparing B-ALL patient samples and normal
195 Pro-B cell samples across all B-ALL genomic regions of interest (217,240 regions). Differential binding
196 scores indicated the AP-1 family of TFs (e.g., FOS, JUN) as the most prominent TFs with higher
197 binding in B-ALL patient samples compared to normal Pro-B cells (**Figure 3D**). In contrast, prominent
198 TFs with higher binding in Pro-B cells were TFs such as TFAP2A, KLF15, CTCFL, ZBTB14 and EBF1.
199 To further demonstrate AP-1 TF occupancy in B-ALL accessible chromatin sites we performed
200 CUT&RUN for FOSL2, JUN and JUNB in 697 and SUB15 human B-ALL cell lines (**Figure 3E**; **Figure**
201 **S2**). Intersections with B-ALL accessible chromatin sites from primary cells identified that 27% of these
202 sites were occupied by an AP-1 TF in B-ALL cell lines. Strikingly, our results further uncovered that
203 45% of DAS with higher chromatin accessibility in B-ALL (i.e., B-ALL enriched DAS) also exhibit AP-1
204 TF occupancy (**Figure 3F**), thereby supporting AP-1-associated *cis*-regulatory rewiring in B-ALL. We
205 determined that even though most AP-1 occupied B-ALL enriched DAS localized to promoter-distal
206 regions of the human genome (77%), there is a 2.7-fold enrichment for AP-1 occupancy at B-ALL

207 enriched promoters compared to B-ALL enriched DAS devoid of AP-1 occupancy (**Figure 3G**; 16% vs
208 6%). Further integration of AP-1 occupied B-ALL enriched DAS with promoter capture Hi-C in B-ALL
209 cell lines identified target genes that were enriched for cell cycle, autophagy and apoptotic signaling
210 pathways (**Table S6**; example in **Figure 3H**).
211 As an extension of our TF footprinting data we also integrated B-ALL cell line promoter capture
212 Hi-C using the ABC enhancer algorithm to refine identification of TF-target gene relationships across
213 top TFs and a cancer implicated gene set [26]. Specifically, we focused on top TF footprints within B-ALL
214 enriched DAS and the cancer implicated gene targets of these DAS predicted by the ABC enhancer
215 algorithm. Concordant with global TF footprint and AP-1 TF occupancy analyses we identified the AP-1
216 family as top TFs in this network. We also identified other top TFs from TF footprinting such as CEBP
217 family TFs and BACH2 (**Figure 3I**). Other prominent top TFs include NFIC, XBP1, TBX1 and numerous
218 basic helix-loop-helix (bHLH) class TFs (e.g., MYOG, MYF5 and HES5). Top expressed cancer
219 implicated gene targets for each TF converged on notable genes involved in cell signaling (*TGFBR2,*
220 *CXCR4*), histone mark modification (*ARID5B*), transcriptional regulation (*MYC*, *KLF6*, *HIF1A*) and the
221 PI3K-AKT pathway (*PTEN*) (**Figure 3I**). Collectively, these results highlight a rewiring of signaling
222 pathways and TF binding networks that facilitate the proliferative potential of B-ALL samples compared
223 to Pro-B cells.
224
225 **Identification of subtype-enriched chromatin architecture**
226 To better understand chromatin accessibility within B-ALL, inter-subtype analyses were performed to
227 identify DAS exhibiting subtype-enriched signal (i.e., henceforth referred to as subtype-enriched DAS)
228 in 10 B-ALL molecular subtypes harboring known molecular drivers (*BCR::ABL1*, *DUX4*-rearranged,
229 *ETV6::RUNX1*, high hyperdiploid, low hypodiploid, *KMT2A*-rearranged, Ph-like, *PAX5*-altered,
230 *TCF3::PBX1* and *ZNF384*-rearranged; **Figures 4A** and **4C**). For this analysis, we compared a single B-
231 ALL subtype cohort with all other B-ALL cell samples not belonging to that subtype in pairwise fashion
232 covering all subtypes. This approach was utilized to emphasize high degrees of subtype enrichment
233 compared to the full spectrum of chromatin accessibility variability in the remaining sample cohort. We
234 identified between 307 and 10,639 DAS in each B-ALL subtype, with a total of 42,457 subtype-enriched
235 DAS identified across all 10 B-ALL subtypes ($\log_2$ fold change > or < 1, FDR<0.05; **Figure 4B**, **Table**
236 **S7**). We annotated subtype-enriched DAS on a subtype basis and determined that a majority of
237 subtype-enriched DAS in each B-ALL subtype (87%, range=80%-90%) localized to promoter-distal
238 regions of the genome (intronic and distal intergenic; **Figure 4D**), and 43%, on average (range=39%-
239 49%), localized to distal intergenic regions, thereby emphasizing the importance of non-genic loci in
240 defining B-ALL chromatin heterogeneity.

241 To further evaluate subtype-enriched DAS, we determined if they displayed enrichment patterns
242 that were consistent with five established human B-ALL cell lines (697= *TCF3::PBX1*, Nalm6= *DUX4*-
243 rearranged, REH= *ETV6::RUNX1*, SEM= *KMT2A*-rearranged and SUPB15= *BCR::ABL1*). Concordant
244 with DAS in patient samples, subtype-enriched DAS exhibited the strongest (BCR-ABL, *DUX4*-
245 rearranged, *ETV6::RUNX1*, *KMT2A*-rearranged) or second strongest (*TCF3::PBX1*) accessibility in the
246 concordant cell line that was representative of that subtype (**Figure S3**). These data suggest that B-
247 ALL cell lines exhibit chromatin accessibility that is largely consistent with primary B-ALL cell sample
248 from the corresponding subtype.
249 To further determine functional effects on gene expression, we integrated subtype-enriched
250 DAS with DEGs uniquely up-regulated ($\log_2$ fold change >1, FDR<0.05) in each of the 10 B-ALL
251 molecular subtypes to determine if they were enriched near DEGs. We identified a statistically
252 significant enrichment of subtype-enriched DAS near up-regulated DEGs in 9 of 10 subtypes compared
253 to total expressed genes in the corresponding subtype (Kolmogorov-Smirnov test $p< 0.05$; **Figure 4E**,
254 **Figure S4**) and uncovered a strong statistical trend in Ph-like B-ALL (Kolmogorov-Smirnov test $p=$
255 0.06; **Figure S4**). Consequently, these data support the role of subtype-enriched DAS in gene
256 regulation and gene activation and further suggest that differences in chromatin accessibility contribute
257 to transcriptomic differences among B-ALL subtypes [3,4]. Collectively, these results highlight extensive
258 open chromatin heterogeneity among B-ALL molecular subtypes.
259
260 **<u>Mapping transcription factor drivers and gene regulatory networks in B-ALL subtypes</u>**
261 We performed TF footprint profiling using merged ATAC-seq signal from 10 B-ALL subtypes with
262 known molecular drivers to identify subtype-enriched TF drivers. TF footprint profiling [12] identified
263 between 4,303,155 and 5,441,937 bound motifs in each B-ALL subtype, with 49,402,067 TF footprints
264 at 815,992 unique genomic loci identified across all subtypes. Using these data, we next identified key
265 TF footprints that were enriched in each subtype (i.e., subtype-enriched TF footprints) by calculating
266 differential footprint scores between every subtype-subtype pair for each TF motif. The top median
267 differential motif scores for each subtype were selected as subtype-enriched TF footprints. This
268 approach was utilized to emphasize differential TF footprint motifs that were consistent and distinct for
269 each subtype rather than repetitive global trends (**Figure 5A**). Notably, subtype-enriched TF footprints
270 were identified for recognized TF drivers such as DUX4 in *DUX4*-rearranged ALL and ZNF384 in
271 *ZNF384*-rearranged ALL. We also identified HOX family TFs (HOXA9, HOXB9, HOXC9 and HOXD9) in
272 *KMT2A*-rearranged ALL, GATA family TFs (GATA2, GATA3, GATA4, GATA5 and GATA6) in *ZNF384*-
273 rearranged ALL and nuclear receptor family TFs in *PAX5*-altered ALL (ESR1, ESR2, ESRRA, NR2F6,
274 NR2F1, RARA and THRB) that all had strong subtype-enriched TF footprints.

275 Because DNA consensus motifs can be highly redundant within TF families, we integrated
276 subtype-enriched TF footprints with DEGs uniquely up-regulated in each subtype to identify candidate
277 TFs from these TF families that are up-regulated in the corresponding B-ALL subtype. This analysis
278 identified *HOXA9* and *HOXC9*, *RARA* and *GATA3* as up-regulated genes in *KMT2A*-rearranged, *PAX5*-
279 altered and *ZNF384*-rearranged subtypes, respectively (**Figure 5B**, **Figure S5**). In addition, *DUX4*
280 (*DUX4*-rearranged) and *MEIS1* (*KMT2A*-rearranged) were also identified as up-regulated TF genes
281 with subtype-enriched TF footprints (**Figure S5**).
282 To determine if these up-regulated TFs promote unique chromatin accessibility landscapes
283 among B-ALL subtypes, we also performed TF footprinting on subtype-enriched DAS by comparing
284 differential footprint scores at subtype-enriched DAS between each B-ALL subtype and Pro-B cells
285 (**Figure 5C**, **Figure S6**). Notably, these data supported a role of DUX4 in *DUX4*-rearranged ALL,
286 ZNF384 and GATA3 in *ZNF384*-rearranged ALL, and HOXA9 and MEIS1 in *KMT2A*-rearranged ALL in
287 the generation of subtype-specific chromatin landscapes (**Figure 5C**, **Figure S6**).
288
289 **Predictive potential of B-ALL subtype-enriched DAS**
290 We determined how well chromatin accessibility can predict B-ALL subtypes by constructing a stepwise
291 Principal Component Analysis-Linear Discriminant Analysis (PCA-LDA) classification model using the
292 42,457 subtype-enriched DAS ATAC-seq read count matrix as initial input across 10 B-ALL subtypes
293 harboring known molecular drivers (outlined in **Figure 6A**). Notably, the constructed classification
294 model was tested with leave-one-out cross validation at an accuracy of 86%. The most common failure
295 was incorrect classification of *BCR::ABL1* and Ph-like subtypes (**Figure 6B**), as has been observed
296 with other ALL classification algorithms [27]. Taking this into account by grouping *BCR::ABL1* and Ph-Like
297 subtype samples into a common class yielded a re-calculated cross validation accuracy of 91%.
298 Visualization of B-ALL subtype separations using select dimensions output by the LDA model
299 demonstrates distinct groupings of related subtypes emphasizing classification model performance
300 (**Figure 6C**).
301 As a further application of our classification model, we also applied the algorithm to 26 B-ALL
302 patient samples of unknown molecular B-ALL subtype. Although transcriptomic profiling for B-ALL
303 drivers is not available to fully validate these samples, when processed with the constructed PCA-LDA
304 model and projected onto original LDA dimensions they distinctly cluster with known molecular
305 subtypes supporting reasonable predictions (**Figure 6D**). Collectively, these data support the utility of
306 chromatin structure and subtype-enriched DAS in B-ALL subtype classification.
307
308

309 **Mapping inherited DNA sequence variants that impact chromatin accessibility**

310 To determine how germline variation impacts chromatin accessibility, we identified chromatin
311 accessibility quantitative trait loci using ATAC-seq (ATAC-QTLs) in a subset of 69 patient samples with
312 available SNP genotyping information and allele-specific ATAC-seq read counting using RASQUAL [28].
313 In total, 9080 ATAC-QTLs were identified representing both directionalities, with reference or alternative
314 alleles increasing chromatin accessibility (FDR<0.1; **Figure 7A**, **Table S8**). Manual quantification and
315 scaling of allele-specific read counts for select ATAC-QTLs identified with RASQUAL demonstrated a
316 clear concordance and directionality among individual patient samples classified into genotype groups
317 (**Figure 7B**). Visual inspection of merged read counts from patient samples grouped into reference
318 allele homozygote, heterozygote, or alternate allele homozygote for select ATAC-QTLs further supports
319 the high-quality nature of identified ATAC-QTLs (**Figure 7C**). We further determined that 218 ATAC-
320 QTLs where also lead eQTL SNPs when compared to GTEx eQTLs [29] from relevant tissues (blood and
321 lymphoblastoid cells), with 85% also concordant for allele overrepresentation directionality (**Figure 7D**;
322 **Table S9**). ATAC-QTLs were also compared with inherited genome-wide association study (GWAS)
323 variants for ALL disease susceptibility which identified rs3824662 (*GATA3*) [30] and rs17481869 (2p22.3)
324 [31] as ATAC-QTLs that were associated with risk of developing B-ALL. Further supporting the validity of
325 our methodology, rs3824662 was also identified as an ATAC-QTL in ALL PDX samples [32], and we
326 functionally validated differential allele-specific activity for rs17481869 in multiple B-ALL cell lines
327 (**Figure S7**).

328 To infer the impact of TF binding in control of chromatin accessibility at ATAC-QTLs we
329 overlapped ATAC-QTL loci with TF motifs determined as TF-bound by footprint profiling [12]. Nearly one-
330 third (28.8%; 2615/9080 ATAC-QTLs) of these ATAC-QTLs overlapped a TF-bound motif footprint
331 across multiple B-ALL subtypes, suggesting that most ATAC-QTLs do not have a clear TF-binding
332 mechanism in how they impact chromatin accessibility. Analysis of bound TF motif footprint prevalence
333 at ATAC-QTLs identified several ETS family TFs (EHF, ELF3, SPI1/PU.1 and SPIB), zinc finger TFs
334 (ZNF263, ZNF460, ZNF740 and ZNF148) and CTCF as the most altered motifs leading to differences
335 in chromatin accessibility between alleles (**Figure 7E**). Notably, we also identified PAX5 and IKZF1,
336 which have known roles in B-cell development and leukemogenesis [33-36]. Collectively, these data
337 identify inherited DNA sequence variants contributing to chromatin heterogeneity among B-ALL
338 subtypes and indicate specific TFs of interest for further exploration of ATAC-QTLs.
339
340
341
342

343 **DISCUSSION**

344 Our study provides the first, large-scale examination of chromatin accessibility in the B-ALL genome
345 across an expansive set of B-ALL subtypes. We further integrated this data with ChIP-seq histone
346 modification enrichment in primary B-ALL cells and three-dimensional chromatin looping data using
347 promoter capture Hi-C in multiple B-ALL cell lines. Our data demonstrate that most regions of
348 chromatin accessibility harbor activating chromatin marks consistent with *cis*-regulatory elements
349 involved in gene regulation, and we further confirmed direct looping to gene promoters for
350 approximately 50% of accessible chromatin sites. However, this does not rule out more transient
351 chromatin looping interactions difficult to detect by current chromatin conformation capture genomic
352 techniques.

353 Extensive epigenomic reprogramming was uncovered between B-cell progenitors and B-ALL,
354 and cell-of-origin analyses identified Pro-B cells as the most common cell-of-origin. Our comparison of
355 B-ALL and pro-B cell chromatin accessibility suggests epigenomic reprogramming that is, in part,
356 associated with AP-1 TF occupancy. We further identify disruptions to normal B-cell function through
357 the activation of toll-like receptor signaling and interleukin production. Acetyl Co-A synthesis was also
358 identified as an enriched gene ontology term when comparing B-ALL and Pro-B cells. Metabolic
359 alterations in cancer are well known, particularly acetyl-Co-A synthesis alterations which have been
360 previously reported in cancer [37]. In addition to metabolic alterations, *PTEN*, a known tumor suppressor
361 gene is frequently mutated in a large portion of cancers [38]. However, in B-ALL the cancer role of PTEN
362 has been reported to be inverted, functioning instead as an oncogene [39]. Reinforcing this conclusion
363 and further suggesting PTEN as an intriguing target for B-ALL treatment, we also found *PTEN* in our
364 network as a top gene target of B-ALL enriched DAS.

365 We further examined accessible chromatin landscapes among diverse molecular subtypes of B-
366 ALL. Collectively, we identified 42,457 subtype-enriched DAS which strikingly represent 20% of
367 analyzed accessible chromatin sites across a pan-subtype B-ALL genome. Subtype-enriched DAS
368 were enriched near up-regulated DEG in the corresponding subtype, supporting their role in gene
369 activation. Moreover, comparisons between subtype-enriched DAS and chromatin accessibility data
370 from cell lines identified largely consistent patterns. We further identified candidate TFs that exhibited
371 strong subtype-specificity through TF footprinting analyses and validated some of these findings using
372 transcriptomic data from primary B-ALL cells. Collectively, these analyses highlighted the role of
373 HOXA9 and MEIS1 in *KMT2A*-rearranged ALL, GATA3 in *ZNF384*-rearranged ALL and RARA in
374 *PAX5*-altered B-ALL. We further confirmed the previously reported roles of DUX4 and ZNF384 in
375 *DUX4*-rearranged and *ZNF384*-rearranged ALLs, respectively. Concordant with our findings, previous
376 studies have identified the co-upregulation of HOXA9 and MEIS1 in *KMT2A*-rearranged leukemias and

377 further support that these TFs are key drivers of leukemogenesis [40-42]. Our identification of numerous
378 HOX TFs with enriched footprints in *KMT2A*-rearranged ALL is also consistent with observations of
379 HOX gene dysregulation in this subtype [43]. Further supporting our results, ZNF384 fusion proteins in
380 *ZNF384*-rearranged ALL are known to up-regulate *GATA3* expression [44,45]. Although a direct role for
381 RARA in *PAX5*-altered B-ALL has not been established, previous work has identified *PAX5* as a target
382 gene of the PLZF-RARA fusion protein in acute promyelocytic leukemia [46]. Moreover, both *RARA* and
383 *PAX5* genes can form fusions with *PML* in acute promyelocytic leukemia [47] and ALL [48], respectively.
384 While *PAX5*-altered ALL has not been well connected to RARA nuclear receptor signaling, there has
385 been previous work treating IKZF1 mutated BCR-ABL1 ALL with RARA and RXR agonists that
386 suppressed a self-renewal phenotype [49]. Collectively, these data warrant further investigation of RARA
387 and RXR signaling in *PAX5*-altered ALL.
388 Supporting the utility of chromatin accessibility in B-ALL classification, subtype-enriched DAS
389 predicted subtypes with 86% accuracy. As a comparison to chromatin accessibility, transcriptional
390 profiling using ALLSorts correctly assigned B-ALL subtypes with 92% accuracy [27]. However, this RNA-
391 seq dataset included over 1223 transcriptomes from 18 subtypes representing a considerably larger
392 dataset for model development. We therefore suspect that additional chromatin accessibility profiling
393 across more B-ALL subtypes and increased sample sizes will lead to even better subtype prediction
394 that will rival transcriptomic profiling and importantly, incorporate intergenic heterogeneity that can
395 elucidate *cis*-regulatory drivers of B-ALL leukemogenesis.
396 To identify the role of inherited DNA sequence variation on the B-ALL chromatin landscape, we
397 mapped over 9000 ATAC-QTLs (FDR<0.1). A large subset of ATAC-QTLs mapped to TF footprints and
398 were concordant in allelic biases with GTEx lead eQTLs. Further validating our analysis, we functionally
399 validated a variant (rs17481869; 2p22.3) associated with susceptibility to ALL [31]. Collectively, this
400 analysis suggests that chromatin accessibility is additionally modified by inherited DNA sequence
401 variation, thereby further contributing to increased chromatin heterogeneity in B-ALL.
402 Overall, our data support pronounced changes in chromatin accessibility between B-ALL and
403 precursor B-cells, as well as among B-ALL subtypes. Our results further support the role of diverse TFs
404 and inherited genetic variants in modulating and promoting differences in chromatin accessibility among
405 B-ALL subtypes. Ultimately, these diverse chromatin architectures contribute to unique gene regulatory
406 networks and transcriptional programs. Our work therefore provides a valuable resource to the cancer
407 genomics research community and can be further used to better understand biological as well as
408 clinical differences among B-ALL subtypes.
409
410

411 # METHODS

412 **Patient samples**

413 Patient samples were obtained from: St. Jude Children's Research Hospital (Memphis, Tennessee),
414 ECOG-ACRIN Cancer Research Group, The Alliance for Clinical Trials in Oncology, MD Anderson
415 Cancer Center (Houston, Texas), Cook Children's Medical Center (Fort Worth, Texas), Lucile Packard
416 Children's Hospital (Palo Alto, California), The University of Chicago (Chicago, Illinois), Novant Health
417 Hemby Children's Hospital (Charlotte, North Carolina) and Children's Hospital of Michigan (Detroit,
418 Michigan). All patients or their legal guardians provided written informed consent. The use of these
419 samples was approved by the institutional review board at St. Jude Children's Research Hospital.
420

421 **Functional genomic studies**

422 ATAC-seq using the Fast-ATAC[10] protocol was performed on 10,000 fresh primary ALL cells. H3K27ac
423 ChIP-seq was performed as previously described[50] on 20 million fresh primary ALL cells. CUT&RUN for
424 FOSL2/Fra2 (Cell signaling; 19967S), JUN (Epicypher; 13-2019) and JUNB (Cell Signaling; 3753S)
425 was performed using the Epicypher Cutana CUT&RUN kit v3.0 (14-1048) according to the
426 manufacturers provided instructions. Next-generation sequencing of ATAC-seq, CUT&RUN, and ChIP-
427 seq was performed at the Hartwell Center for Bioinformatics and Biotechnology at St. Jude Children's
428 Research Hospital. Transcriptomic and SNP genotyping data from B-ALL patient samples were
429 obtained from St. Jude Children's Research Hospital. Normal B-cell ATAC-seq [10,25] were downloaded
430 from NCBI (GSE122989 and GSE74912). B-ALL cell histone modification ChIP-seq datasets
431 (H3K27ac, H3K4me1 and H3K27me3) were downloaded from the Blueprint Epigenome Consortium
432 (https://www.blueprint-epigenome.eu/). Expression quantitative trait loci (eQTL) data was obtained from
433 previous studies [51]. Arima promoter capture Hi-C (Arima; A510008, A303010, A302010) was performed
434 on 10 million B-ALL cell lines (697, BALL1, Nalm6, RS411, REH, SEM and SUPB15) according to the
435 manufacturers provided instructions using unspecified proprietary buffers, solutions, enzymes, and
436 reagents. See **Supplemental Methods** for additional details.
437

438 **Data analysis**

439 ATAC-seq and ChIP-seq reads were mapped to the hg19 reference genome using bowtie2 [52] and
440 peaks were identified using MACS2 [53]. Regions of interest for ATAC-seq analyses were selected using
441 a reproducible peak summit approach within each subtype cohort with subsequent region merging.
442 DESeq2 [54] was employed to identify B-ALL-enriched or subtype-enriched differentially accessible
443 chromatin sites (DAS). Two B-ALL subtype patient samples (IKZF1 N159Y and iAMP21) were included
444 in B-ALL versus Pro-B cell analyses but were excluded from additional studies due to limited sample

size. Promoter capture Hi-C libraries from B-ALL cell lines were analyzed at 3-kb resolution using the Arima CHiC pipeline (v1.4, https://github.com/ArimaGenomics/CHiC). Genomic regions representing separate loop ends were compiled to facilitate overlap determinations with B-ALL patient chromatin accessible regions of interest using "bedtools intersect". Enhancer and target gene prediction for network construction was analyzed with the ABC enhancer algorithm (https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction). In brief, inputs for the ABC enhancer algorithm included, B-ALL enriched DAS, merged B-ALL patient ATAC-seq, H3K27Ac ChIP-seq, Arima promoter capture Hi-C contact counts with ABC score threshold at 0.04. The Genomic Regions Enrichment of Annotations Tool (GREAT) [55] was used to identify candidate target gene sets and ontologies associated with DAS. TOBIAS [12] was used to identify TF footprints at accessible chromatin sites. The Principal Component Analysis-Linear Discriminant Analysis (PCA-LDA) subtype classification model was constructed stepwise by first PCA transformation of subtype-enriched ATAC-seq counts, then applying LDA on an optimized number of principal components. RASQUAL [28] was used to map chromatin accessibility quantitative trait loci using ATAC-seq (ATAC-QTLs). Significant ATAC-QTLs for each region were identified with a genome-wide computed FDR of 10%. See **Supplemental Methods** for additional details.

DATA AND CODE AVAILABILITY

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Daniel Savic (daniel.savic@stjude.org).

ACKNOWLEDGEMENTS

We would like to thank the Hartwell Center at St. Jude for ATAC-seq, ChIP-seq and promoter capture Hi-C library preparation and next-generation sequencing. We would also like to thank Jeremy Hunt and Brandon Smart for technical support. This work was supported by the National Cancer Institute (R01CA234490, P30CA021765, UG1CA232760, UG1CA189859, and U10CA180820), the National Institute of General Medical Studies (P50GM115279) and the American Lebanese Syrian Associated Charities (ALSAC). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

DECLARATIONS OF INTEREST

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

Conceptualization, K.R. Barnett, D.S.; Methodology, K.R. Barnett, D.S.; Investigation, K.R. Barnett, J.D.D., B.P.B, K.R. Bhattarai; Analysis, K.R. Barnett, D.S.; Data Curation, K.R.B., W.Y.; Patient sample acquisition, K.R.C., C.S.M., E.J., E.P., M.R.L., S.M.K., W.S., H.I., S.J., C.H.P., C.G.M., M.V.R., W.E.E., J.J.Y.; Writing – Original Draft, K.R. Barnett, D.S.; Writing – Review & Editing, K.R. Barnett, J.D.D., B.P.B, K.R. Bhattarai, W.Y., K.R.C., C.S.M., E.J., E.P., M.R.L., S.M.K., W.S., H.I., S.J., C.H.P., C.G.M., M.V.R., W.E.E., J.J.Y, D.S.

REFERENCES

1. Moriyama, T., Relling, M.V., and Yang, J.J. (2015). Inherited genetic variation in childhood acute lymphoblastic leukemia. Blood *125*, 3988-3995. 10.1182/blood-2014-12-580001.
2. Mullighan, C.G. (2012). Molecular genetics of B-precursor acute lymphoblastic leukemia. J Clin Invest *122*, 3407-3415. 10.1172/JCI61203.
3. Lilljebjorn, H., Henningsson, R., Hyrenius-Wittsten, A., Olsson, L., Orsmark-Pietras, C., von Palffy, S., Askmyr, M., Rissler, M., Schrappe, M., Cario, G., et al. (2016). Identification of ETV6-RUNX1-like and DUX4-rearranged subtypes in paediatric B-cell precursor acute lymphoblastic leukaemia. Nat Commun *7*, 11790. 10.1038/ncomms11790.
4. Andersson, A., Olofsson, T., Lindgren, D., Nilsson, B., Ritz, C., Eden, P., Lassen, C., Rade, J., Fontes, M., Morse, H., et al. (2005). Molecular signatures in childhood acute leukemia and their correlations to expression patterns in normal hematopoietic subpopulations. Proc Natl Acad Sci U S A *102*, 19069-19074. 10.1073/pnas.0506637102.
5. Figueroa, M.E., Chen, S.C., Andersson, A.K., Phillips, L.A., Li, Y., Sotzen, J., Kundu, M., Downing, J.R., Melnick, A., and Mullighan, C.G. (2013). Integrated genetic and epigenetic analysis of childhood acute lymphoblastic leukemia. J Clin Invest *123*, 3099-3111. 10.1172/JCI66203.
6. Almamun, M., Levinson, B.T., van Swaay, A.C., Johnson, N.T., McKay, S.D., Arthur, G.L., Davis, J.W., and Taylor, K.H. (2015). Integrated methylome and transcriptome analysis reveals novel regulatory elements in pediatric acute lymphoblastic leukemia. Epigenetics *10*, 882-890. 10.1080/15592294.2015.1078050.
7. Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. Nat Rev Genet *20*, 207-220. 10.1038/s41576-018-0089-8.
8. Bhagwat, A.S., Lu, B., and Vakoc, C.R. (2018). Enhancer dysfunction in leukemia. Blood *131*, 1795-1804. 10.1182/blood-2017-11-737379.
9. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods *10*, 1213-1218. 10.1038/nmeth.2688.
10. Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L., Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., et al. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat Genet *48*, 1193-1203. 10.1038/ng.3646.

11. Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-resolution mapping and characterization of open chromatin across the genome. Cell *132*, 311-322. 10.1016/j.cell.2007.12.014.
12. Bentsen, M., Goymann, P., Schultheis, H., Klee, K., Petrova, A., Wiegandt, R., Fust, A., Preussner, J., Kuenne, C., Braun, T., et al. (2020). ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. Nat Commun *11*, 4267. 10.1038/s41467-020-18035-1.
13. Gusmao, E.G., Allhoff, M., Zenke, M., and Costa, I.G. (2016). Analysis of computational footprinting methods for DNase sequencing experiments. Nat Methods *13*, 303-309. 10.1038/nmeth.3772.
14. Zhang, K., Hocker, J.D., Miller, M., Hou, X., Chiou, J., Poirion, O.B., Qiu, Y., Li, Y.E., Gaulton, K.J., Wang, A., et al. (2021). A single-cell atlas of chromatin accessibility in the human genome. Cell *184*, 5985-6001 e5919. 10.1016/j.cell.2021.10.024.
15. Corces, M.R., Granja, J.M., Shams, S., Louie, B.H., Seoane, J.A., Zhou, W., Silva, T.C., Groeneveld, C., Wong, C.K., Cho, S.W., et al. (2018). The chromatin accessibility landscape of primary human cancers. Science *362*. 10.1126/science.aav1898.
16. Cejas, P., Xie, Y., Font-Tello, A., Lim, K., Syamala, S., Qiu, X., Tewari, A.K., Shah, N., Nguyen, H.M., Patel, R.A., et al. (2021). Subtype heterogeneity and epigenetic convergence in neuroendocrine prostate cancer. Nat Commun *12*, 5775. 10.1038/s41467-021-26042-z.
17. Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317-330. 10.1038/nature14248.
18. Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A *107*, 21931-21936. 10.1073/pnas.1016071107.
19. Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311-318. 10.1038/ng1966.
20. Hoffman, B.G., Robertson, G., Zavaglia, B., Beach, M., Cullum, R., Lee, S., Soukhatcheva, G., Li, L., Wederell, E.D., Thiessen, N., et al. (2010). Locus co-occupancy, nucleosome positioning, and H3K4me1 regulate the functionality of FOXA2-, HNF4A-, and PDX1-bound loci in islets and liver. Genome Res *20*, 1037-1051. 10.1101/gr.104356.109.
21. Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S.W., Andrews, S., Grey, W., Ewels, P.A., et al. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet *47*, 598-606. 10.1038/ng.3286.
22. Sondka, Z., Bamford, S., Cole, C.G., Ward, S.A., Dunham, I., and Forbes, S.A. (2018). The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat Rev Cancer *18*, 696-705. 10.1038/s41568-018-0060-1.
23. Consortium, I.T.P.-C.A.o.W.G. (2020). Pan-cancer analysis of whole genomes. Nature *578*, 82-93. 10.1038/s41586-020-1969-6.
24. Jackson, T.R., Ling, R.E., and Roy, A. (2021). The Origin of B-cells: Human Fetal B Cell Development and Implications for the Pathogenesis of Childhood Acute Lymphoblastic Leukemia. Front Immunol *12*, 637975. 10.3389/fimmu.2021.637975.
25. O'Byrne, S., Elliott, N., Rice, S., Buck, G., Fordham, N., Garnett, C., Godfrey, L., Crump, N.T., Wright, G., Inglott, S., et al. (2019). Discovery of a CD10-negative B-progenitor in human fetal life identifies unique ontogeny-related developmental programs. Blood *134*, 1059-1071. 10.1182/blood.2019001289.
26. Fulco, C.P., Nasser, J., Jones, T.R., Munson, G., Bergman, D.T., Subramanian, V., Grossman, S.R., Anyoha, R., Doughty, B.R., Patwardhan, T.A., et al. (2019). Activity-by-contact model of

572 enhancer-promoter regulation from thousands of CRISPR perturbations. Nat Genet *51*, 1664-
573 1669. 10.1038/s41588-019-0538-0.
574 27. Schmidt, B., Brown, L.M., Ryland, G.L., Lonsdale, A., Kosasih, H.J., Ludlow, L.E., Majewski,
575 I.J., Blombery, P., Ekert, P.G., Davidson, N.M., and Oshlack, A. (2022). ALLSorts: an RNA-Seq
576 subtype classifier for B-cell acute lymphoblastic leukemia. Blood Adv *6*, 4093-4097.
577 10.1182/bloodadvances.2021005894.
578 28. Kumasaka, N., Knights, A.J., and Gaffney, D.J. (2016). Fine-mapping cellular QTLs with
579 RASQUAL and ATAC-seq. Nat Genet *48*, 206-213. 10.1038/ng.3467.
580 29. Consortium, G.T., Laboratory, D.A., Coordinating Center -Analysis Working, G., Statistical
581 Methods groups-Analysis Working, G., Enhancing, G.g., Fund, N.I.H.C., Nih/Nci, Nih/Nhgri,
582 Nih/Nimh, Nih/Nida, et al. (2017). Genetic effects on gene expression across human tissues.
583 Nature *550*, 204-213. 10.1038/nature24277.
584 30. Perez-Andreu, V., Roberts, K.G., Harvey, R.C., Yang, W., Cheng, C., Pei, D., Xu, H., Gastier-
585 Foster, J., E, S., Lim, J.Y., et al. (2013). Inherited GATA3 variants are associated with Ph-like
586 childhood acute lymphoblastic leukemia and risk of relapse. Nat Genet *45*, 1494-1498.
587 10.1038/ng.2803.
588 31. Vijayakrishnan, J., Studd, J., Broderick, P., Kinnersley, B., Holroyd, A., Law, P.J., Kumar, R.,
589 Allan, J.M., Harrison, C.J., Moorman, A.V., et al. (2018). Genome-wide association study
590 identifies susceptibility loci for B-cell childhood acute lymphoblastic leukemia. Nat Commun *9*,
591 1340. 10.1038/s41467-018-03178-z.
592 32. Yang, H., Zhang, H., Luan, Y., Liu, T., Yang, W., Roberts, K.G., Qian, M.X., Zhang, B., Yang,
593 W., Perez-Andreu, V., et al. (2022). Noncoding genetic variation in GATA3 increases acute
594 lymphoblastic leukemia risk through local and global changes in chromatin conformation. Nat
595 Genet *54*, 170-179. 10.1038/s41588-021-00993-x.
596 33. Medvedovic, J., Ebert, A., Tagoh, H., and Busslinger, M. (2011). Pax5: a master regulator of B
597 cell development and leukemogenesis. Adv Immunol *111*, 179-206. 10.1016/B978-0-12-
598 385991-4.00005-2.
599 34. Gu, Z., Churchman, M.L., Roberts, K.G., Moore, I., Zhou, X., Nakitandwe, J., Hagiwara, K.,
600 Pelletier, S., Gingras, S., Berns, H., et al. (2019). PAX5-driven subtypes of B-progenitor acute
601 lymphoblastic leukemia. Nat Genet *51*, 296-307. 10.1038/s41588-018-0315-5.
602 35. Mullighan, C.G., Su, X., Zhang, J., Radtke, I., Phillips, L.A., Miller, C.B., Ma, J., Liu, W., Cheng,
603 C., Schulman, B.A., et al. (2009). Deletion of IKZF1 and prognosis in acute lymphoblastic
604 leukemia. N Engl J Med *360*, 470-480. 10.1056/NEJMoa0808253.
605 36. Churchman, M.L., Qian, M., Te Kronnie, G., Zhang, R., Yang, W., Zhang, H., Lana, T., Tedrick,
606 P., Baskin, R., Verbist, K., et al. (2018). Germline Genetic IKZF1 Variation and Predisposition to
607 Childhood Acute Lymphoblastic Leukemia. Cancer Cell *33*, 937-948 e938.
608 10.1016/j.ccell.2018.03.021.
609 37. Guertin, D.A., and Wellen, K.E. (2023). Acetyl-CoA metabolism in cancer. Nat Rev Cancer.
610 10.1038/s41568-022-00543-5.
611 38. Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and
612 Stratton, M.R. (2004). A census of human cancer genes. Nat Rev Cancer *4*, 177-183.
613 10.1038/nrc1299.
614 39. Shojaee, S., Chan, L.N., Buchner, M., Cazzaniga, V., Cosgun, K.N., Geng, H., Qiu, Y.H., von
615 Minden, M.D., Ernst, T., Hochhaus, A., et al. (2016). PTEN opposes negative selection and
616 enables oncogenic transformation of pre-B cells. Nat Med *22*, 379-387. 10.1038/nm.4062.
617 40. Collins, C., Wang, J., Miao, H., Bronstein, J., Nawer, H., Xu, T., Figueroa, M., Muntean, A.G.,
618 and Hess, J.L. (2014). C/EBPalpha is an essential collaborator in Hoxa9/Meis1-mediated
619 leukemogenesis. Proc Natl Acad Sci U S A *111*, 9899-9904. 10.1073/pnas.1402238111.
620 41. Sun, Y., Zhou, B., Mao, F., Xu, J., Miao, H., Zou, Z., Phuc Khoa, L.T., Jang, Y., Cai, S., Witkin,
621 M., et al. (2018). HOXA9 Reprograms the Enhancer Landscape to Promote Leukemogenesis.
622 Cancer Cell *34*, 643-658 e645. 10.1016/j.ccell.2018.08.018.

42. Miyamoto, R., Kanai, A., Okuda, H., Komata, Y., Takahashi, S., Matsui, H., Inaba, T., and Yokoyama, A. (2021). HOXA9 promotes MYC-mediated leukemogenesis by maintaining gene expression for multiple anti-apoptotic pathways. Elife *10*. 10.7554/eLife.64148.
43. Ferrando, A.A., Armstrong, S.A., Neuberg, D.S., Sallan, S.E., Silverman, L.B., Korsmeyer, S.J., and Look, A.T. (2003). Gene expression signatures in MLL-rearranged T-lineage and B-precursor acute leukemias: dominance of HOX dysregulation. Blood *102*, 262-268. 10.1182/blood-2002-10-3221.
44. Qian, M., Zhang, H., Kham, S.K., Liu, S., Jiang, C., Zhao, X., Lu, Y., Goodings, C., Lin, T.N., Zhang, R., et al. (2017). Whole-transcriptome sequencing identifies a distinct subtype of acute lymphoblastic leukemia with predominant genomic abnormalities of EP300 and CREBBP. Genome Res *27*, 185-195. 10.1101/gr.209163.116.
45. Yaguchi, A., Ishibashi, T., Terada, K., Ueno-Yokohata, H., Saito, Y., Fujimura, J., Shimizu, T., Ohki, K., Manabe, A., and Kiyokawa, N. (2017). EP300-ZNF384 fusion gene product up-regulates GATA3 gene expression and induces hematopoietic stem cell gene expression signature in B-cell precursor acute lymphoblastic leukemia cells. Int J Hematol *106*, 269-281. 10.1007/s12185-017-2220-6.
46. Spicuglia, S., Vincent-Fabert, C., Benoukraf, T., Tiberi, G., Saurin, A.J., Zacarias-Cabeza, J., Grimwade, D., Mills, K., Calmels, B., Bertucci, F., et al. (2011). Characterisation of genome-wide PLZF/RARA target genes. PLoS One *6*, e24176. 10.1371/journal.pone.0024176.
47. Chen, Z., and Chen, S.J. (1992). RARA and PML genes in acute promyelocytic leukemia. Leuk Lymphoma *8*, 253-260. 10.3109/10428199209051004.
48. Kurahashi, S., Hayakawa, F., Miyata, Y., Yasuda, T., Minami, Y., Tsuzuki, S., Abe, A., and Naoe, T. (2011). PAX5-PML acts as a dual dominant-negative form of both PAX5 and PML. Oncogene *30*, 1822-1830. 10.1038/onc.2010.554.
49. Churchman, M.L., Low, J., Qu, C., Paietta, E.M., Kasper, L.H., Chang, Y., Payne-Turner, D., Althoff, M.J., Song, G., Chen, S.C., et al. (2015). Efficacy of Retinoids in IKZF1-Mutated BCR-ABL1 Acute Lymphoblastic Leukemia. Cancer Cell *28*, 343-356. 10.1016/j.ccell.2015.07.016.
50. Ferguson, D.C., McCorkle, J.R., Barnett, K.R., Bonten, E.J., Bergeron, B.P., Bhattarai, K.R., Yang, W., Smith, C., Hansen, B.S., Bajpai, R., et al. (2022). Amino acid stress response genes promote L-asparaginase resistance in pediatric acute lymphoblastic leukemia. Blood Adv *6*, 3386-3397. 10.1182/bloodadvances.2022006965.
51. Autry, R.J., Paugh, S.W., Carter, R., Shi, L., Liu, J., Ferguson, D.C., Lau, C.E., Bonten, E.J., Yang, W., McCorkle, J.R., et al. (2020). Integrative genomic analyses reveal mechanisms of glucocorticoid resistance in acute lymphoblastic leukemia. Nat Cancer *1*, 329-344. 10.1038/s43018-020-0037-3.
52. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359. 10.1038/nmeth.1923.
53. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137. 10.1186/gb-2008-9-9-r137.
54. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550. 10.1186/s13059-014-0550-8.
55. McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol *28*, 495-501. 10.1038/nbt.1630.

671 # FIGURE LEGENDS

672

673 **FIGURE 1: Chromatin accessibility landscapes in B-ALL. (A)** Number and genomic location of
674 accessible chromatin sites for 10 B-ALL subtypes and B-other samples is provided. **(B)** Percentage of
675 B-ALL accessible chromatin sties that maps to H3K4me1 and/or H3K27ac active histone marks (Active;
676 green), H3K27me3 and H3K4me1 and/or H3K27ac bivalent or poised histone marks (Bivalent or
677 Poised; yellow) and H3K27me3 only repressed histone marks (Repressed; red). **(C)** B-ALL cell line
678 chromatin loops detected using promoter capture Hi-C at B-ALL accessible chromatin sites. The total
679 number of B-ALL accessible chromatin sites, number of B-ALL accessible chromatin sites within loops
680 and the total number of accessible chromatin sites with a loop to a gene implicated in cancer is shown.
681 **(D)** UCSC genome browser ATAC-seq signal track of average B-ALL chromatin accessibility and
682 promoter capture Hi-C loops across the *IKZF1* gene locus. **(E)** UCSC genome browser ATAC-seq
683 signal tracks of 10 merged B-ALL subtypes with known molecular drivers across the *IKZF1* gene locus.
684

685 **FIGURE 2: B-ALL cell type-of-origin defined by chromatin accessibility. (A)** Differentiation timeline
686 of B-cell progenitors from least differentiated to most differentiated. HSC= hematopoietic stem cell,
687 MPP= multipotent progenitor cell, LMPP= lymphoid-primed multipotent progenitor cell, CLP= common
688 lymphoid progenitor cell, PreProB= prePro-B cell, ProB= Pro-B cell and CD19+,CD20+= B cell. **(B)**
689 Heatmap of B-cell progenitor or B-ALL patient sample variance stabilized ATAC-seq signal across B-
690 cell progenitor-defining chromatin loci. B-cell progenitor groups most similar to B-ALL patient samples
691 (preProB and ProB) are outlined in yellow. **(C)** Confusion matrix showing number (listed) and
692 percentage (color-coded) of B-cell progenitor truths and predictions for leave-one-out cross validation of
693 a K-nearest neighbor classifier model. **(D)** Distribution of B-cell progenitor classification across B-ALL
694 patient samples using a K-nearest neighbor classifier model trained with B-cell progenitor data.
695

696 **FIGURE 3: Mapping differential accessibility between B-ALL and Pro-B cells. (A)** Heatmap of Pro-
697 B cell or B-ALL patient sample variance stabilized ATAC-seq signal as z-score across Pro-B cell and B-
698 ALL enriched DAS. DAS within heatmap are > 1 or < -1 $\log_2$-adjusted fold change. **(B)** ATAC-seq signal
699 track examples of Pro-B-cell-enriched DAS and B-ALL-enriched DAS on the UCSC genome browser.
700 Flanking genomic regions are included for context. **(C)** Gene ontology analysis of DAS with higher
701 accessibility in B-ALL (B-ALL-enriched) at various $\log_2$-adjusted fold change thresholds. All terms were
702 significant using both binomial and hypergeometric statistical tests. **(D)** Differential transcription factor
703 footprinting between Pro-B cells and B-ALL patient samples across 217,240 B-ALL genomic regions of
704 interest. **(E)** FOSL2 CUT&RUN enrichment heatmaps at all B-ALL accessible chromatin sites and B-

705 ALL enriched DAS (B-ALL enrich) in SUPB15 (left) and 697 (right) cells. **(F)** Number of B-ALL enriched
706 DAS overlapping AP-1 TF occupancy (FOSL2, JUN and/or JUNB) in 697 (left) SUPB15 (middle) and
707 both B-ALL cell lines (right). Number of overlapping sites are shown in purple while non-overlapping
708 sites are shown in yellow. **(G)** Genome annotation of B-ALL enriched DAS with AP-1 TF occupancy
709 (left) or that are devoid of AP-1 TF occupancy (right). **(H)** IGV genome browser image showing a B-ALL
710 enriched DAS that maps to accessible chromatin and sites of AP-1 TF occupancy in SUPB15 cells.
711 Promoter capture Hi-C (PC-HiC) looping between the distal AP-1 occupied sites and the *IGFBP7* gene
712 promoter is shown. B-ALL (red) and pro-B (blue) cell ATAC-seq tracks are overlaid in the top panel.
713 Signal tracks for FOSL2, JUN and JUNB in SUBP15 cells are shown. **(I)** Transcription factor and target
714 gene network of DAS with higher accessibility in B-ALL (B-ALL-enriched). Network is subset for top
715 transcription factor footprints across DAS ranked by the top mean $\log_2$-adjusted fold change
716 transcription factor footprint signal. Target genes are subset for a cancer implicated gene set ranked by
717 the top expressed genes. Network connections are colored as transcription factors (purple blocks) to
718 target gene (green arrow heads) pairs. Select expansive and highly similar transcription factor motif
719 families are grouped (AP-1 and CEBP; AP1-family and CEBP-family).
720
721 **FIGURE 4: Mapping differential accessibility among B-ALL molecular subtypes. (A)** Heatmap of
722 variance stabilized ATAC-seq signal as z-score across subtype-enriched DAS. Enrichment patterns for
723 each subtype DAS set are shown on vertical axis and are grouped by B-ALL subtype patient sample on
724 the horizontal axis. Ph-like and BCR-ABL subtype-enriched DAS are expanded at the right for clarity.
725 **(B)** Pie chart shows the number and percentage of subtype-enriched DAS identified. **(C)** ATAC-seq
726 signal track examples of subtype-enriched DAS on the UCSC genome browser. **(D)** Genomic
727 annotations of subtype-enriched DAS for each B-ALL subtype is provided. The fraction of sites
728 harboring different annotations is plotted. **(E)** Cumulative distribution function for *BCR::ABL1* and
729 *ZNF384*-rearranged ALL comparing the fraction (y-axis) of subtype up-regulated genes (Subtype
730 DEGs; gray or light green) and all expressed subtype genes (Expressed; black) at different distance
731 cutoffs from subtype-enriched DAS and their transcription start sites (x-axis). Kolmogorov-Smirnov (K-
732 S) p-values are provided.
733
734 **FIGURE 5: TF footprinting and gene regulatory networks identify key TF drivers in B-ALL**
735 **subtypes. (A)** Heatmap list of the topmost consistently differential TF footprints between all pairwise
736 subtype-subtype comparisons (y-axis; labeled to the right of the heatmap as TF motif identifiers)
737 enriched in 10 B-ALL subtypes (x-axis; labeled on top of heatmap as z-score of differential TF footprint
738 signal output by TOBIAS). **(B)** RNA-seq transcripts per million (TPM) expression of key TFs with

739 subtype-enriched footprints that are also up-regulated in the corresponding subtype (colored) versus all
740 other subtypes (gray). DESeq2 differentially expressed gene FDR significance values are provided. **(C)**
741 Top TF footprints at *KMT2A*-enriched DAS are shown. Differential footprint score between B-ALL and
742 Pro-B cells is provided on the x-axis and TF footprint significance is provided on the y-axis. Transcripts
743 per million (TPM) transcript abundance of associated TF transcript is shown as both color and size of
744 points.
745
746 **FIGURE 6: Classification model accurately predicts B-ALL subtypes. (A)** Flow chart outlines
747 process for PCA-LDA classification of B-ALL subtypes. **(B)** Confusion matrix showing number (listed)
748 and percentage (color-coded) of B-ALL subtype truths and predictions for leave-one-out cross
749 validation. **(C)** Three-dimensional plots showing clustering of B-ALL subtypes utilizing select
750 dimensions from the LDA model. **(D)** B-ALL subtype identification for unknown B-ALL samples (black
751 points). Clustering for unknown samples identified as *DUX4*-rearranged, *BCR::ABL1* and high
752 hyperdiploid (from left to right) is shown.
753
754 **FIGURE 7: Identification of ATAC-QTLs impacting chromatin accessibility. (A)** ATAC-QTL effect
755 size (x-axis) and significance (y-axis) is plotted for all significant ATAC-QTLs (FDR<0.1). **(B)** Examples
756 of allele-specific effects on ATAC-seq read count at ATAC-QTLs between samples from the three
757 genotype groups. Homozygous reference allele= homoz_REF, heterozygous= heteroz_REFALT and
758 homozygous alternative allele= homoz_ALT. **(C)** UCSC browser ATAC-seq signal tracks of merged
759 BAM files from patients with distinct genotypes at *ARL11* (top panel) and *TTC7B* (bottom panel) gene
760 loci. ATAC-QTLs are marked by an asterisk. Homozygous reference allele= homoz_REF,
761 heterozygous= heteroz_REFALT and homozygous alternative allele= homoz_ALT. ENCODE ChIP-seq
762 TF binding sites are shown below each ATAC-seq signal track. **(D)** Scatterplot of effect size for SNPs
763 significant as both ATAC-QTLs (x-axis) and GTEx lead eQTL (y-axis). **(E)** Abundance of top TF-bound
764 motifs overlapping ATAC-QTLs. Highly similar TF motifs were grouped into motif families via TOBIAS
765 motif clustering as shown on the x-axis.

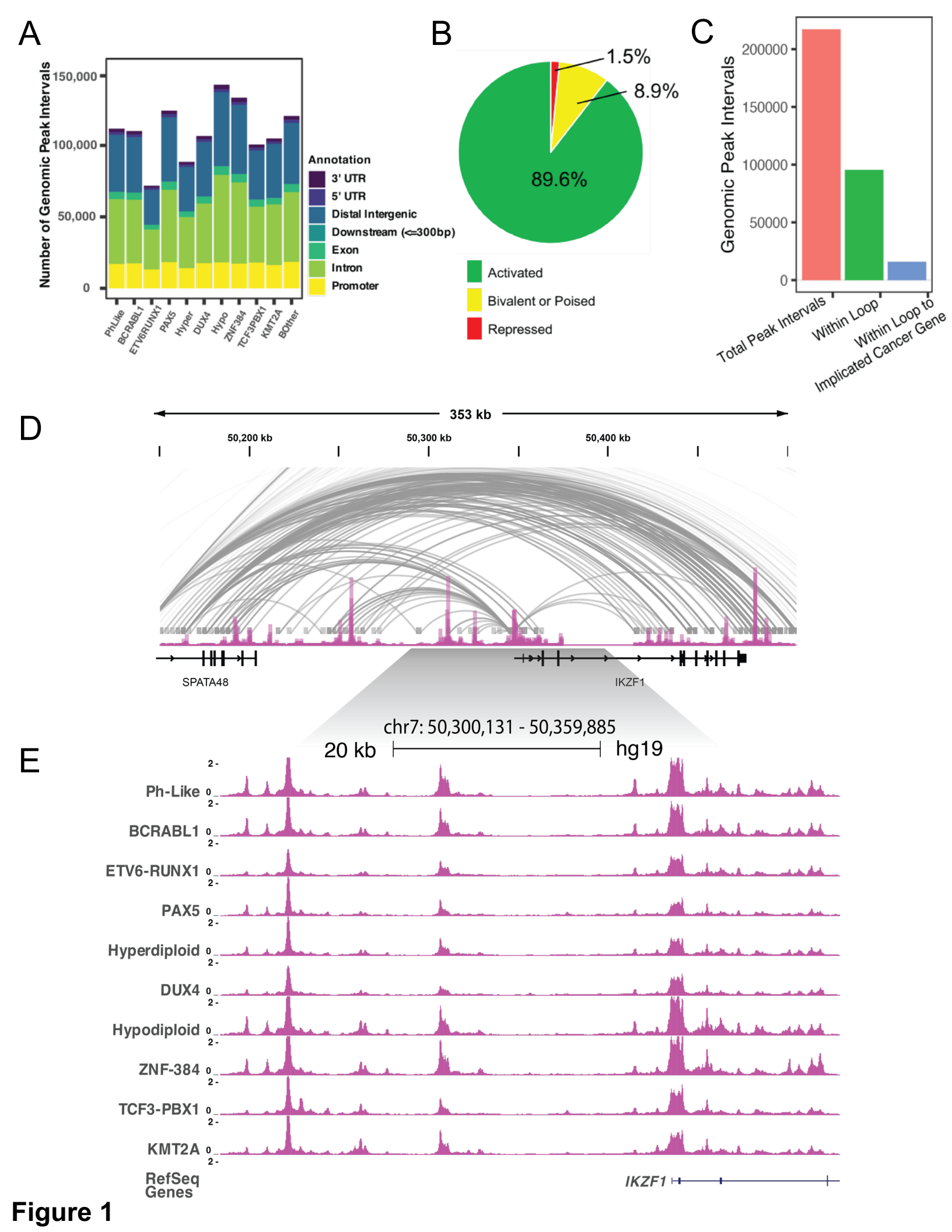

A
Number of Genomic Peak Intervals
150,000
100,000
50,000
0
PhLike
BCRABL1
ETV6RUNX1
PAX5
Hyper
DUX4
Hypo
ZNF384
TCF3PBX1
KMT2A
BOther
Annotation
3' UTR
5' UTR
Distal Intergenic
Downstream (<=300bp)
Exon
Intron
Promoter
B
1.5%
8.9%
89.6%
Activated
Bivalent or Poised
Repressed
C
Genomic Peak Intervals
200000
150000
100000
50000
0
Total Peak Intervals
Within Loop
Within Loop to Implicated Cancer Gene
D
353 kb
50,200 kb
50,300 kb
50,400 kb
SPATA48
IKZF1
chr7: 50,300,131 - 50,359,885
20 kb
hg19
E
Ph-Like
BCRABL1
ETV6-RUNX1
PAX5
Hyperdiploid
DUX4
Hypodiploid
ZNF-384
TCF3-PBX1
KMT2A
RefSeq Genes
IKZF1

Figure 1

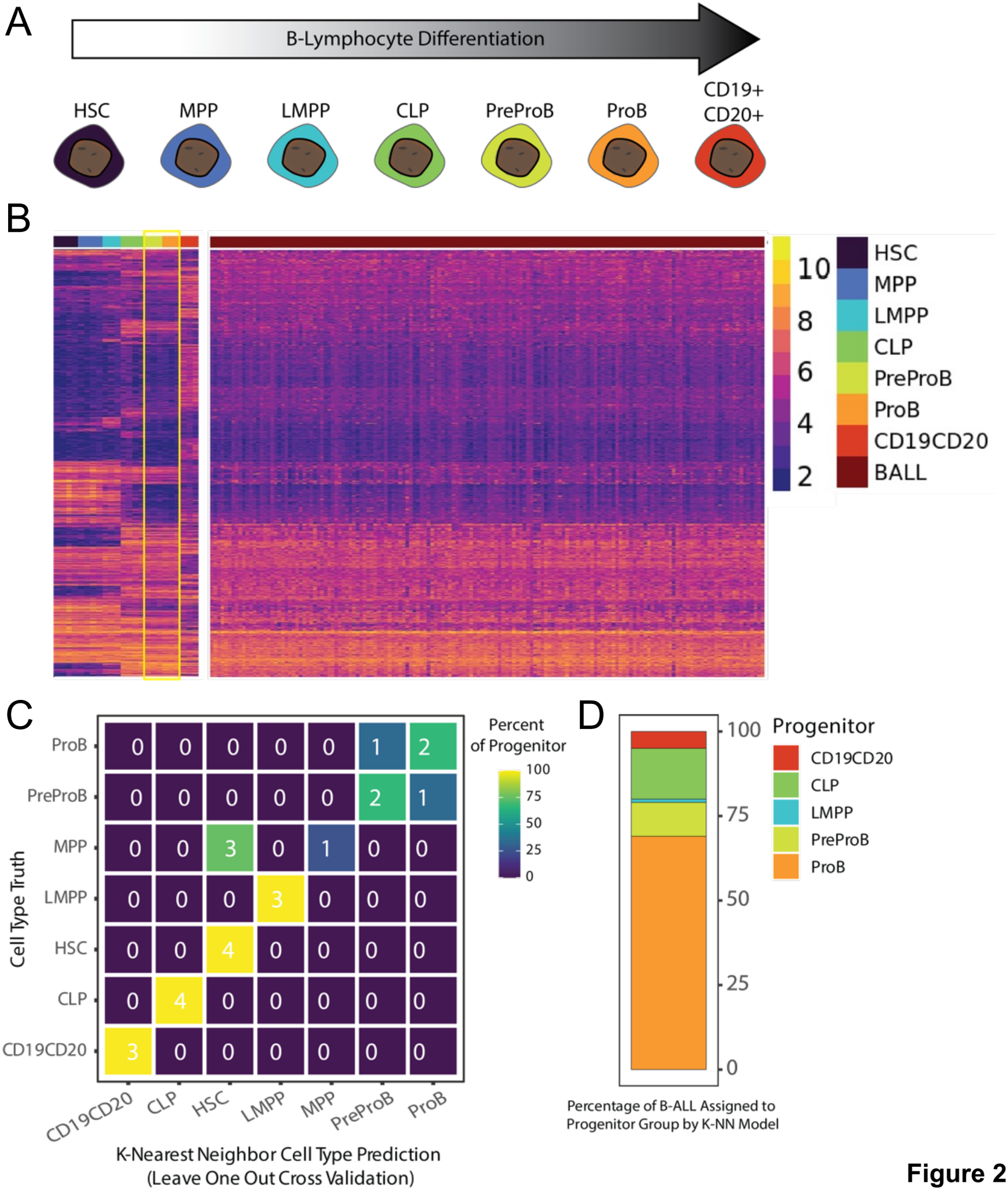

A
B-Lymphocyte Differentiation
HSC
MPP
LMPP
CLP
PreProB
ProB
CD19+
CD20+
B
10
8
6
4
2
HSC
MPP
LMPP
CLP
PreProB
ProB
CD19CD20
BALL
C
Cell Type Truth
ProB
PreProB
MPP
LMPP
HSC
CLP
CD19CD20
CD19CD20
CLP
HSC
LMPP
MPP
PreProB
ProB
K-Nearest Neighbor Cell Type Prediction
(Leave One Out Cross Validation)
Percent
of Progenitor
100
75
50
25
0
D
100
75
50
25
0
Progenitor
CD19CD20
CLP
LMPP
PreProB
ProB
Percentage of B-ALL Assigned to
Progenitor Group by K-NN Model

Figure 2

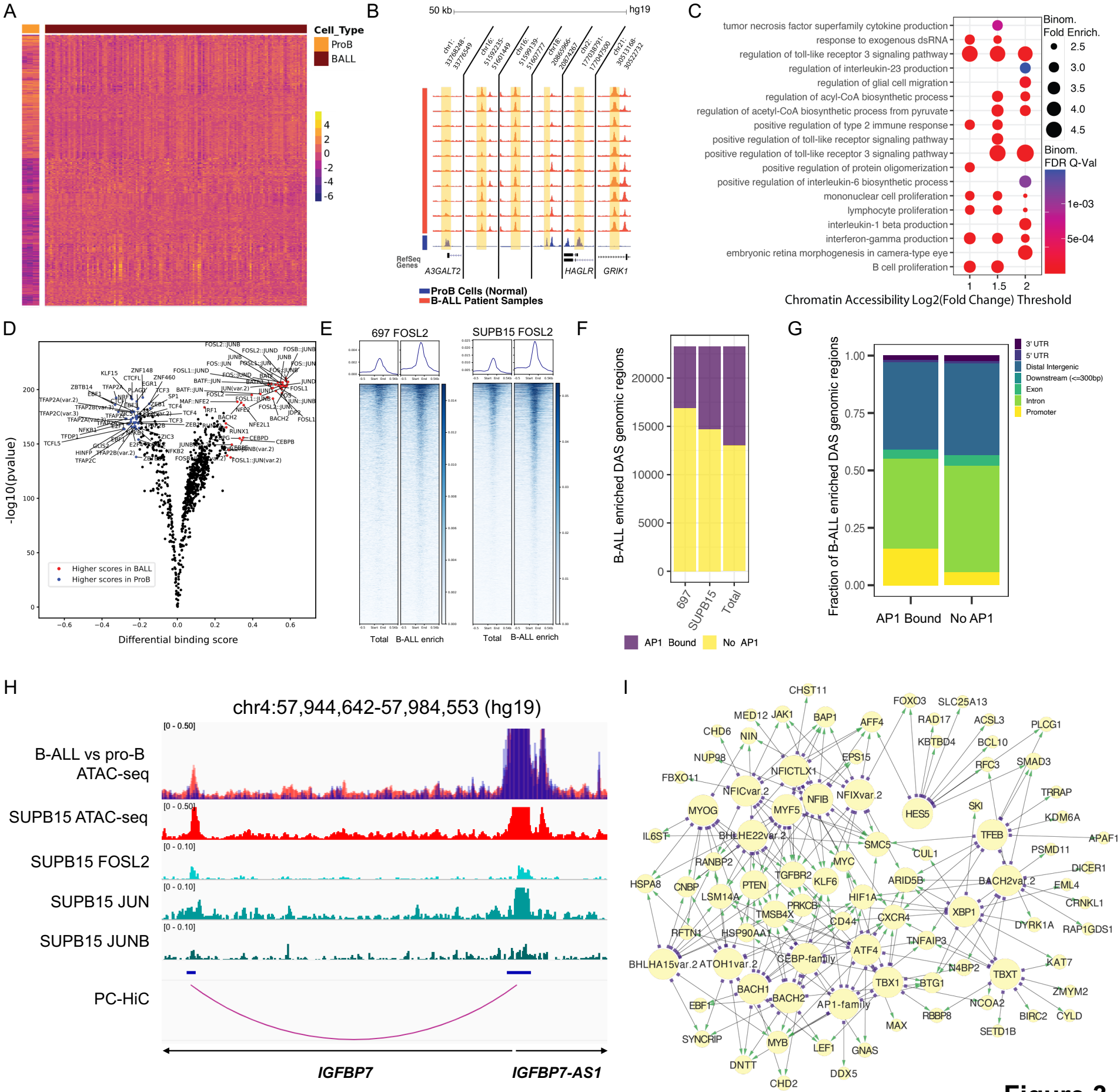

A
Cell_Type
ProB
BALL
B
50 kb
hg19
RefSeq Genes
A3GALT2
HAGLR
GRIK1
ProB Cells (Normal)
B-ALL Patient Samples
C
tumor necrosis factor superfamily cytokine production
response to exogenous dsRNA
regulation of toll-like receptor 3 signaling pathway
regulation of interleukin-23 production
regulation of glial cell migration
regulation of acyl-CoA biosynthetic process
regulation of acetyl-CoA biosynthetic process from pyruvate
positive regulation of type 2 immune response
positive regulation of toll-like receptor signaling pathway
positive regulation of toll-like receptor 3 signaling pathway
positive regulation of protein oligomerization
positive regulation of interleukin-6 biosynthetic process
mononuclear cell proliferation
lymphocyte proliferation
interleukin-1 beta production
interferon-gamma production
embryonic retina morphogenesis in camera-type eye
B cell proliferation
Binom. Fold Enrich.
Binom. FDR Q-Val
Chromatin Accessibility Log2(Fold Change) Threshold
D
-log10(pvalue)
Differential binding score
Higher scores in BALL
Higher scores in ProB
E
697 FOSL2
SUPB15 FOSL2
Total
B-ALL enrich
F
B-ALL enriched DAS genomic regions
697
SUPB15
Total
AP1 Bound
No AP1
G
Fraction of B-ALL enriched DAS genomic regions
AP1 Bound
No AP1
3' UTR
5' UTR
Distal Intergenic
Downstream (<=300bp)
Exon
Intron
Promoter
H
chr4:57,944,642-57,984,553 (hg19)
B-ALL vs pro-B ATAC-seq
SUPB15 ATAC-seq
SUPB15 FOSL2
SUPB15 JUN
SUPB15 JUNB
PC-HiC
IGFBP7
IGFBP7-AS1
I
CHST11
FOXO3
SLC25A13
MED12
JAK1
BAP1
AFF4
RAD17
ACSL3
PLCG1
CHD6
NIN
KBTBD4
BCL10
NUP98
EPS15
RFC3
SMAD3
FBXO11
NFICTLX1
NFICvar.2
NFIB
NFIXvar.2
HES5
TRRAP
MYOG
MYF5
SKI
KDM6A
IL6ST
BHLHE22var.2
SMC5
TFEB
APAF1
CUL1
PSMD11
RANBP2
MYC
DICER1
HSPA8
CNBP
PTEN
TGFBR2
KLF6
ARID5B
BACH2var.2
EML4
LSM14A
HIF1A
CRNKL1
PRKCB
TMSB4X
CD44
CXCR4
XBP1
DYRK1A
RAP1GDS1
RFTN1
HSP90AA1
TNFAIP3
ATF4
BHLHA15var.2
ATOH1var.2
CEBP-family
N4BP2
KAT7
TBXT
BACH1
TBX1
BTG1
ZMYM2
EBF1
BACH2
AP1-family
NCOA2
RBBP8
BIRC2
CYLD
MAX
SETD1B
SYNCRIP
MYB
LEF1
GNAS
DNTT
DDX5
CHD2

Figure 3

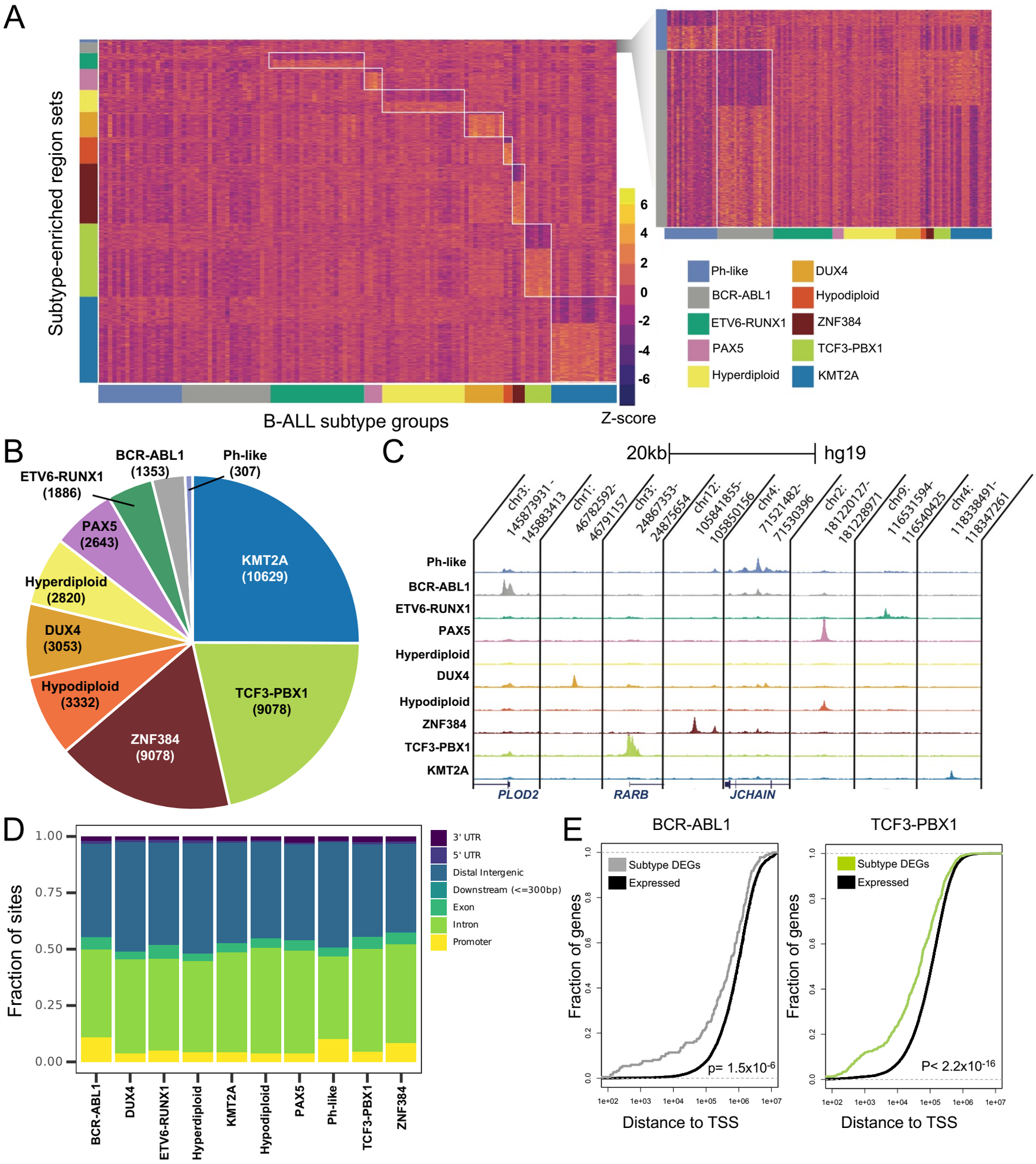

A
Subtype-enriched region sets
B-ALL subtype groups
Z-score
6
4
2
0
-2
-4
-6
Ph-like
BCR-ABL1
ETV6-RUNX1
PAX5
Hyperdiploid
DUX4
Hypodiploid
ZNF384
TCF3-PBX1
KMT2A
B
BCR-ABL1 (1353)
Ph-like (307)
ETV6-RUNX1 (1886)
PAX5 (2643)
Hyperdiploid (2820)
DUX4 (3053)
Hypodiploid (3332)
ZNF384 (9078)
TCF3-PBX1 (9078)
KMT2A (10629)
C
20kb
hg19
chr3: 145873931 - 145883413
chr1: 46782592- 46791157
chr3: 24867353- 24875654
chr12: 105841855- 105850156
chr4: 71521482- 71530396
chr2: 181220127- 181228971
chr9: 116531594- 116540425
chr4: 118338491- 118347261
Ph-like
BCR-ABL1
ETV6-RUNX1
PAX5
Hyperdiploid
DUX4
Hypodiploid
ZNF384
TCF3-PBX1
KMT2A
PLOD2
RARB
JCHAIN
D
Fraction of sites
1.00
0.75
0.50
0.25
0.00
3' UTR
5' UTR
Distal Intergenic
Downstream (<=300bp)
Exon
Intron
Promoter
BCR-ABL1
DUX4
ETV6-RUNX1
Hyperdiploid
KMT2A
Hypodiploid
PAX5
Ph-like
TCF3-PBX1
ZNF384
E
BCR-ABL1
TCF3-PBX1
Subtype DEGs
Expressed
Fraction of genes
p= 1.5x10-6
P< 2.2x10-16
1e+02
1e+03
1e+04
1e+05
1e+06
1e+07
Distance to TSS

Figure 4

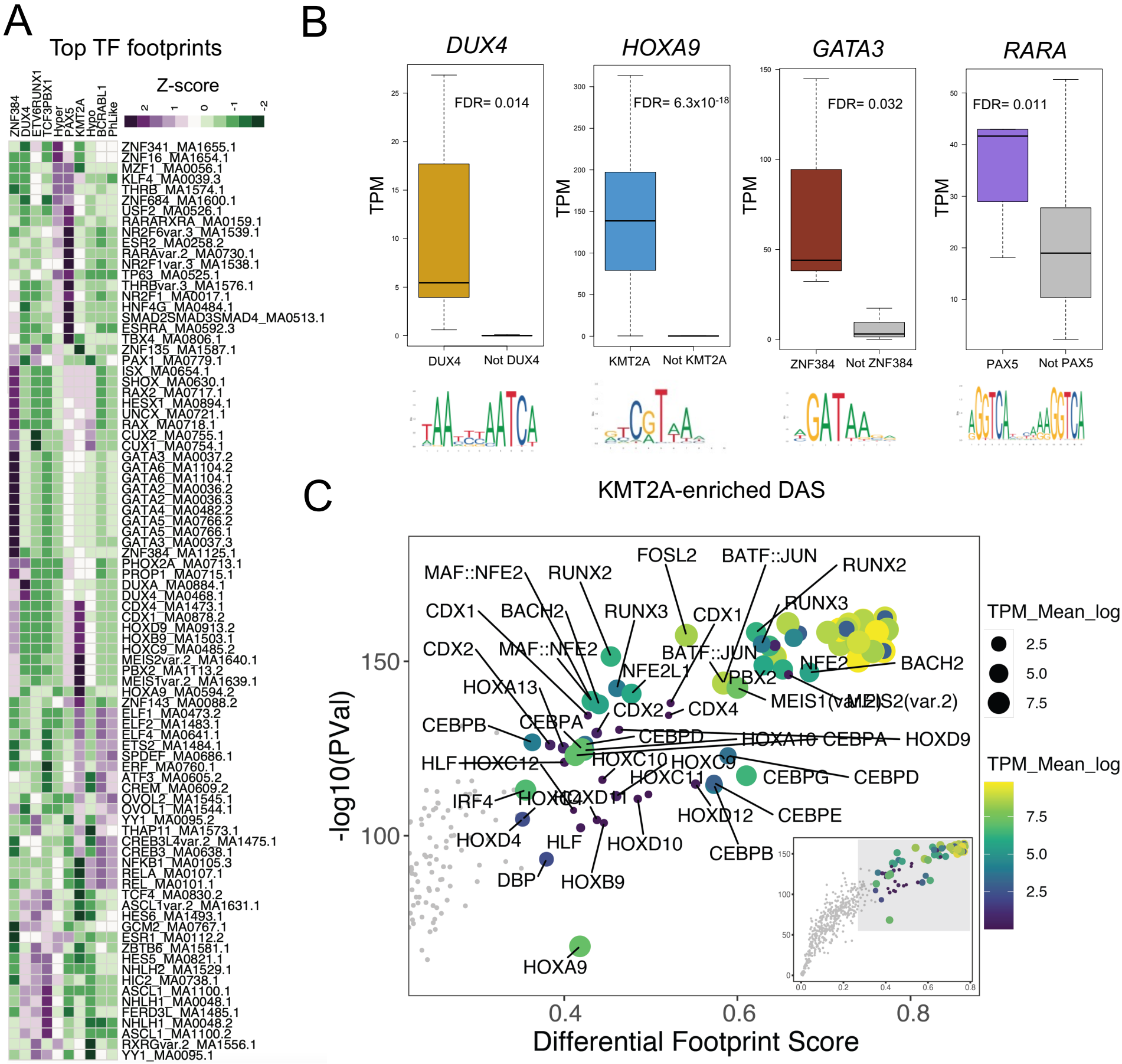
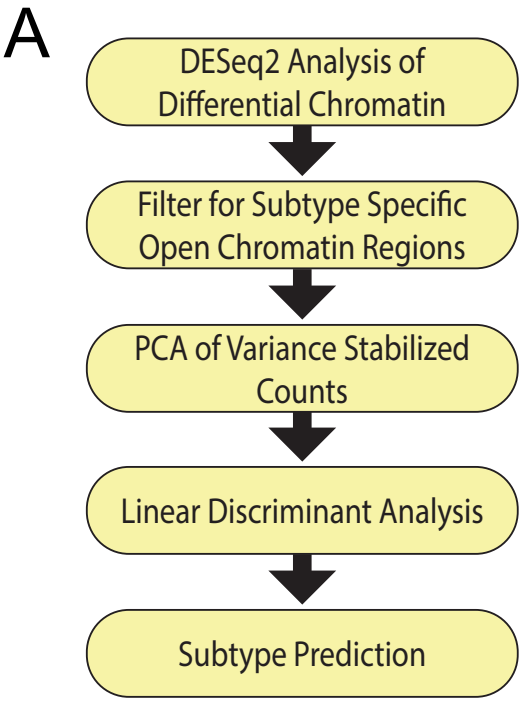
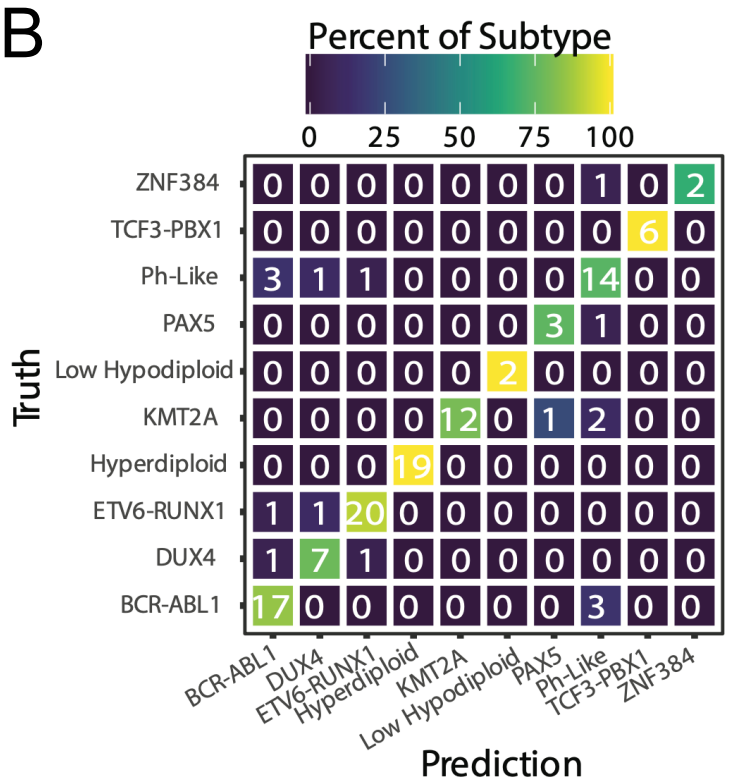
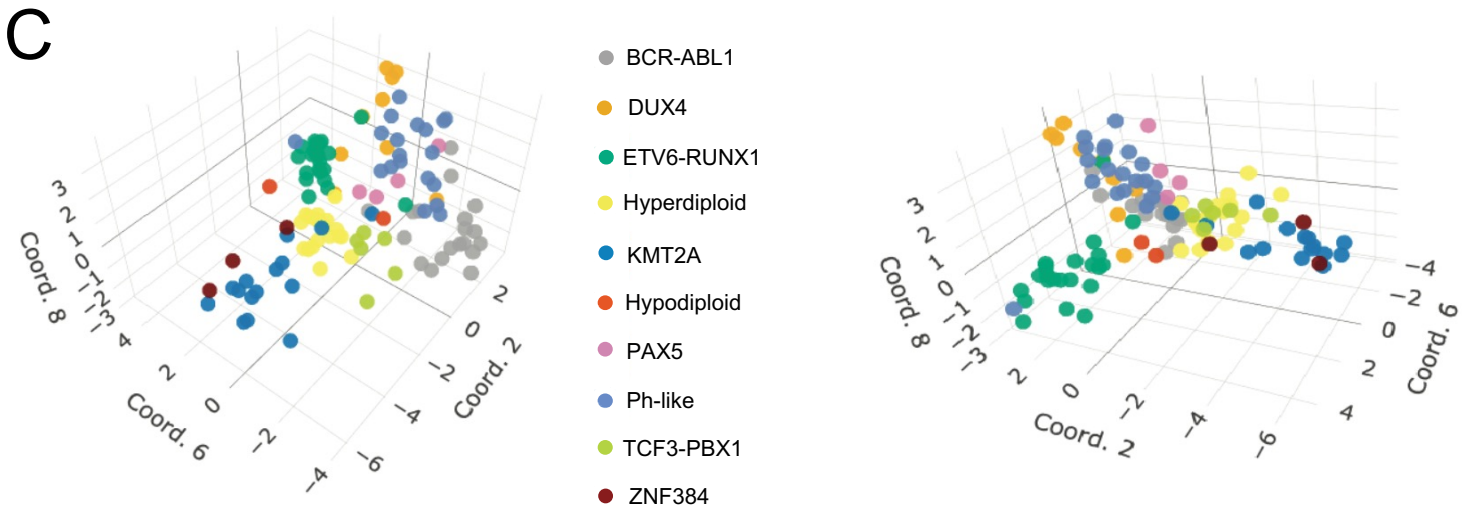

A
Top TF footprints
Z-score
ZNF384
DUX4
ETV6RUNX1
TCF3PBX1
Hyper
PAX5
KMT2A
Hypo
BCRABL1
PhLike
B
DUX4
HOXA9
GATA3
RARA
FDR= 0.014
FDR= 6.3x10-18
FDR= 0.032
FDR= 0.011
TPM
DUX4
Not DUX4
KMT2A
Not KMT2A
ZNF384
Not ZNF384
PAX5
Not PAX5
C
KMT2A-enriched DAS
-log10(PVal)
Differential Footprint Score
TPM_Mean_log
2.5
5.0
7.5
MAF::NFE2
RUNX2
FOSL2
BATF::JUN
CDX1
BACH2
RUNX3
CDX2
NFE2L1
NFE2
HOXA13
CDX4
CEBPB
CEBPA
CEBPD
HOXA10
HOXD9
HLF
HOXC12
HOXC10
HOXC9
HOXC11
CEBPG
IRF4
HOXD11
HOXD12
CEBPE
HOXD4
HOXD10
DBP
HOXB9
HOXA9
MEIS1(var.2)
MEIS2(var.2)
PBX2

Figure 5


A
DESeq2 Analysis of Differential Chromatin
Filter for Subtype Specific Open Chromatin Regions
PCA of Variance Stabilized Counts
Linear Discriminant Analysis
Subtype Prediction


B
Percent of Subtype
0 25 50 75 100
Truth
Prediction
ZNF384 0 0 0 0 0 0 0 1 0 2
TCF3-PBX1 0 0 0 0 0 0 0 0 6 0
Ph-Like 3 1 1 0 0 0 0 14 0 0
PAX5 0 0 0 0 0 0 3 1 0 0
Low Hypodiploid 0 0 0 0 0 2 0 0 0 0
KMT2A 0 0 0 0 12 0 1 2 0 0
Hyperdiploid 0 0 0 19 0 0 0 0 0 0
ETV6-RUNX1 1 1 20 0 0 0 0 0 0 0
DUX4 1 7 1 0 0 0 0 0 0 0
BCR-ABL1 17 0 0 0 0 0 0 3 0 0
BCR-ABL1
DUX4
ETV6-RUNX1
Hyperdiploid
KMT2A
Low Hypodiploid
PAX5
Ph-Like
TCF3-PBX1
ZNF384


C
Coord. 8
Coord. 6
Coord. 2
BCR-ABL1
DUX4
ETV6-RUNX1
Hyperdiploid
KMT2A
Hypodiploid
PAX5
Ph-like
TCF3-PBX1
ZNF384

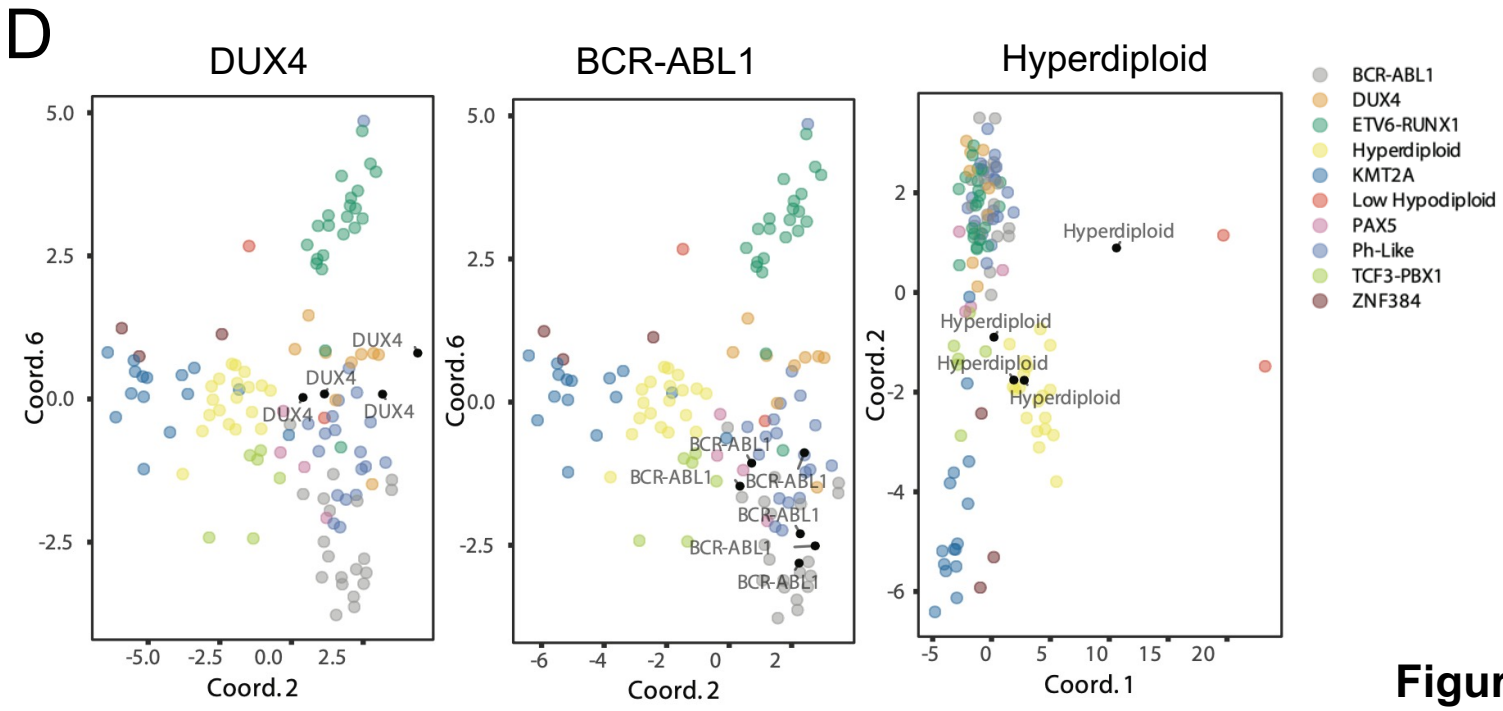

D
DUX4
BCR-ABL1
Hyperdiploid
Coord. 6
Coord. 2
Coord. 1
DUX4
BCR-ABL1
Hyperdiploid
BCR-ABL1
DUX4
ETV6-RUNX1
Hyperdiploid
KMT2A
Low Hypodiploid
PAX5
Ph-Like
TCF3-PBX1
ZNF384

Figure 6

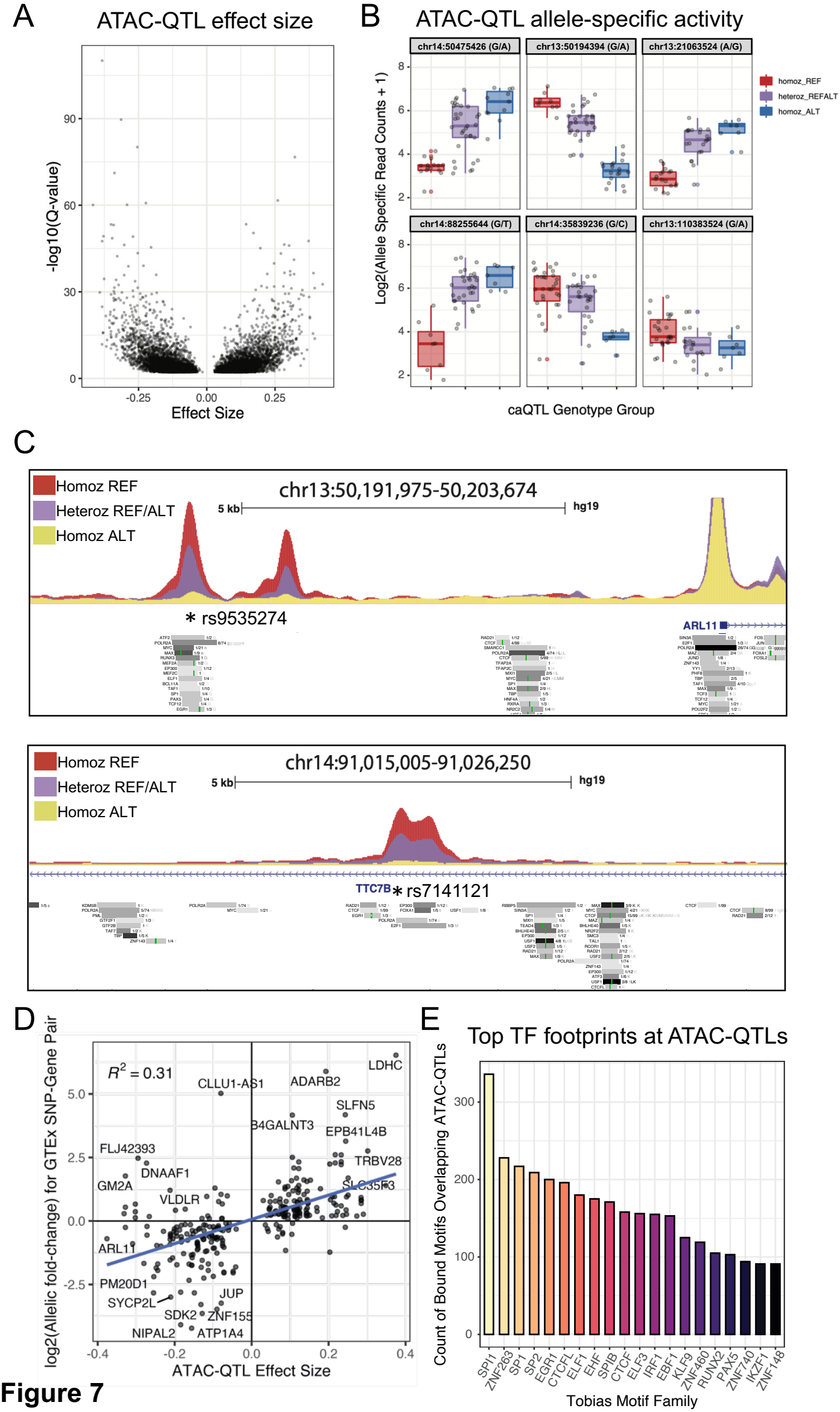

A
ATAC-QTL effect size
-log10(Q-value)
Effect Size
B
ATAC-QTL allele-specific activity
chr14:50475426 (G/A)
chr13:50194394 (G/A)
chr13:21063524 (A/G)
chr14:88255644 (G/T)
chr14:35839236 (G/C)
chr13:110383524 (G/A)
homoz_REF
heteroz_REFALT
homoz_ALT
Log2(Allele Specific Read Counts + 1)
caQTL Genotype Group
C
Homoz REF
Heteroz REF/ALT
Homoz ALT
chr13:50,191,975-50,203,674
5 kb
hg19
* rs9535274
ARL11
chr14:91,015,005-91,026,250
TTC7B
* rs7141121
D
$R^2 = 0.31$
LDHC
CLLU1-AS1
ADARB2
SLFN5
B4GALNT3
EPB41L4B
FLJ42393
TRBV28
DNAAF1
GM2A
SLC35F3
VLDLR
ARL11
PM20D1
JUP
SYCP2L
SDK2
ZNF155
NIPAL2
ATP1A4
log2(Allelic fold-change) for GTEx SNP-Gene Pair
ATAC-QTL Effect Size
E
Top TF footprints at ATAC-QTLs
Count of Bound Motifs Overlapping ATAC-QTLs
SPI1
ZNF263
SP1
SP2
EGR1
CTCFL
ELF1
EHF
SPIB
CTCF
ELF3
IRF1
EBF1
KLF9
ZNF460
RUNX2
PAX5
ZNF740
IKZF1
ZNF148
Tobias Motif Family

Figure 7