**High-throughput single-cell chromatin accessibility CRISPR screens enable unbiased identification of regulatory networks in cancer**

**Authors:** Sarah E. Pierce[1,*], Jeffrey M. Granja[1,2,*], William J. Greenleaf[1,2,†]


**Affiliations:**

[1]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA.

[2]Center for Personal and Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA, USA.

*These authors contributed equally to this work.

†Correspondence to: W.J.G. (wjg@stanford.edu)


**Contact Information:**

William J. Greenleaf, PhD

Stanford University School of Medicine

279 Campus Drive, Beckman Center B257, Stanford, CA 94305-5168

Email: wjg@stanford.edu

Fax: 650-725-1534

Phone: 650-725-3672

1    **Abstract:**

2    Spear-ATAC is a modified droplet-based single-cell ATAC-seq (scATAC-seq) protocol

3    that enables simultaneous read-out of chromatin accessibility profiles and integrated

4    sgRNA spacer sequences from thousands of individual cells at a time. Spear-ATAC

5    profiling of 104,592 cells representing 414 sgRNA knock-down populations revealed the

6    temporal dynamics of epigenetic responses to regulatory perturbations in cancer cells

7    and the associations between transcription factor binding profiles, demonstrating a high-

8    throughput method for perturbing and evaluating dynamic single-cell epigenetic states.

9

10    **Main:**

11    Complex epigenetic regulation is a unique requirement for all multicellular organisms,

12    enabling diverse phenotypes stemming from the same underlying genotype.

13    Understanding how transcription factor binding dynamics drive epigenetic states is

14    essential for the molecular dissection of many core processes, including

15    embryogenesis, differentiation, and cancer[1]. Perturbing the expression levels of

16    epigenetic regulators and observing the subsequent effects on chromatin accessibility

17    provides a powerful means to study transcription factor function. To this end,

18    CRISPR/Cas9 technologies enable precise tuning of gene expression levels using

19    targeted mutation and epigenetic modulation strategies[2,3]. Combined with technologies

20    such as scRNA-seq, the global transcriptional effects of these perturbations can be

21    assayed across thousands of cells at once[4–6]. In contrast, however, current methods to

22    profile the effects of CRISPR/Cas9-based perturbations on single-cell epigenomes are

23    limited to analyzing 96 cells per run on an integrated fluidic circuit[7]. Here, we developed

24    the droplet-based Spear-ATAC protocol (Single-cell perturbations with an accessibility

25    read-out using scATAC-seq) to quantify and map the effects of perturbing transcription

26    factor levels on chromatin accessibility in high throughput. In contrast to previous

27    methods, Spear-ATAC relies on reading out sgRNA spacer sequences directly from

28    genomic DNA rather than off of RNA transcripts.

29         Similar to bulk accessibility profiling using ATAC-seq[8,9], the droplet-based

30    scATAC-seq protocol begins with nuclei isolation and transposition of the sample of

31    interest using a hyperactive transposase (Tn5) that integrates into areas of open

32    chromatin[10,11] (**Fig. 1a** and **Extended Data Fig. 1a**). To first ensure that capture of a

33    single-copy of an integrated sgRNA does not depend on the local accessibility context

34    surrounding the sgRNA sequence, we flanked the lentiviral sgRNA spacer with pre-

35    integrated Nextera Read1 and Read2 adapters (**Extended Data Fig. 2a-b**), enabling

36    the sgRNA sequence to be amplified with the same primers used to amplify ATAC-seq

37    fragments in the library. When testing the detection of sgRNA fragments with bulk

38    ATAC-seq, this design increased our ability to detect sgRNA fragments by ~4-fold

39    without altering sgRNA efficacy (**Extended Data Fig. 2c-d**). Following transposition, the

40    nuclei are loaded into the 10x Controller for the capture of individual nuclei into

41    nanoliter-scale gel-beads in emulsion (GEMs). These GEMs contain barcoded Forward

42    oligos complementary to the Nextera Read1 adapter to amplify all ATAC fragments,

43    thereby tagging each ATAC fragment from the same nucleus with the same 10x

44    barcode. Since this protocol uses a single barcoded primer to tag ATAC fragments, we

45    reasoned that we could preferentially amplify each sgRNA fragment by also spiking in a

46    Reverse oligo specific to the sgRNA backbone (**Fig 1a** and **Extended Data Fig. 3a-b**).

47    This modification allows for exponential amplification of the sgRNA fragment at the

48    same time the rest of the library is amplified linearly, while still ensuring the sgRNA can

49    be used as a substrate for the second round of PCR. We also extended the number of

50    cycles of in-GEM linear amplification of scATAC-seq fragments from 12 to 15, which

51    subsequently adds three rounds of exponential sgRNA amplification without altering

52    scATAC-seq quality (**Extended Data Fig. 3c-d**). Finally, we included a biotin-tagged

53    primer during targeted sgRNA amplification which allows for the specific enrichment of

54    these fragments while minimizing aberrant background signal from scATAC-seq reads

55    (**Extended Data Fig. 3e**). Overall, these changes increase our ability to detect sgRNA

56    fragments by ~40-fold compared to lentiviral integration alone followed by traditional

57    droplet-based scATAC-seq.

58      We first piloted the Spear-ATAC method with a pool of nine CRISPRi sgRNAs

59    targeting two transcription factors (GATA1 and GATA2) and three inert sgRNA controls

60    (Non-targeting or NT) (**Extended Data Fig. 4a**). We introduced this library into K562

61    leukemia cells engineered to express a CRISPRi dCas9-KRAB cassette to knockdown

62    genes of interest, expanded the cells for six days, and then FACS-isolated sgRNA$^+$ cells

63    to process for Spear-ATAC. We captured 6,390 nuclei in the pilot run, of which we were

64    able to directly associate 48% of single-cell epigenetic profiles (n=3,045 nuclei) to their

65    appropriate sgRNA target with >80% specificity (**Fig 1b** and **Extended Data Fig. 4b-e**).

66    Capturing the same number of cell-sgRNA assignments with existing methods would

67    have required ~30 Perturb-ATAC runs costing $9.80/cell compared to one Spear-ATAC

68    run costing $0.46/cell. Perturb-ATAC also requires 4-hour run times on a Fluidigm C1 to

69    process each set of 96 cells, necessitating the handling of multiple batches of frozen

70    cells over several days. Apart from the use of standard PCR machines and bead

71    purification steps, Spear-ATAC only requires an 8-minute run time on a 10x Controller

72    to process up to 80,000 nuclei at once, greatly increasing the potential throughput of

73    these methods.

74         From the 3,045 nuclei assigned to sgRNAs in the pilot Spear-ATAC run, Uniform

75    Manifold Approximation and Projection (UMAP) clearly distinguished cells harboring

76    sgGATA1 from both sgGATA2 and sgNT cells, indicating the high specificity of sgRNA

77    assignments (**Fig. 1c** and **Extended Data Fig. 4c**). GATA1 and GATA2 are both

78    involved in hematopoietic differentiation and development; however, the erythroid

79    transcription factor GATA1 is specifically an essential gene in K562 cells, whereas

80    GATA2 is dispensable for growth and survival in this cell line[12]. We next developed a

81    framework for unbiased identification of changes in TF motif accessibility across

82    multiple populations of cells harboring different sgRNAs. We first computed TF motif

83    accessibility scores (e.g. chromVAR deviation scores[13] across all single cells for a given

84    sgRNA genotype (sgT), then subtracted the average TF motif accessibility scores of the

85    non-targeting (sgNT) cells. We then ranked all of these sgRNA-to-TF motif accessibility

86    difference scores (sgRNA:TF scores) to identify hits. As would be expected, knockdown

87    of GATA1 decreased the accessibility of peaks containing the GATA motif, as well as

88    the accessibility of peaks overlapping with known GATA1 ChIP-seq peaks.

89    Furthermore, GATA1 knockdown resulted in a muted GATA footprint compared to

90    K562;dCas9-KRAB cells expressing non-targeting sgRNAs (**Fig. 1d-f**). Local

91    accessibility at the GATA1 locus also decreased following knockdown, further validating

92    that cells assigned to sgGATA1 are down-regulating expression at this locus (**Fig. 1g**

93 and **Extended Data Fig. 5a**). Interestingly, knocking down GATA1 led to a modest

94 increase in accessibility of GATA3 ChIP-seq peaks as well as an increase in local

95 accessibility near the GATA3 locus (**Extended Data Fig. 5b-c**). GATA3 is typically

96 active in the lymphoid lineage[14] and is not expressed in K562 cells at baseline,

97 suggesting that GATA3 is specifically activated in response to GATA1 knock down.

98 By performing differential accessibility analysis between sgGATA1-containing

99 cells and sgNT-expressing control cells, we observed 14,262 peaks (14.76%)

100 increasing in accessibility and 14,026 peaks (14.52%) decreasing in accessibility (**Fig.**

101 **1h**). Each of the three sgRNAs targeting GATA1 resulted in nearly indistinguishable

102 chromatin accessibility profiles, with peaks decreasing in accessibility following GATA1

103 knockdown enriched for the GATA motif and peaks increasing in accessibility following

104 GATA1 knockdown enriched for SPI/RUNX motifs (**Extended Data Fig. 5d**).

105 Furthermore, individual cells with the lowest aggregate accessibility of genomic regions

106 containing GATA1 motifs had the highest aggregate accessibility of genomic regions

107 containing SPI/RUNX motifs and vice versa, further underscoring these regulatory

108 relationships (**Extended Data Fig. 5e-f**). Supporting these observations, SPI (also

109 known as PU.1) and GATA1 have been previously shown to physically interact and

110 negatively regulate each other[15], exemplifying the type of direct phenotypes that can be

111 assayed and validated using the Spear-ATAC method.

112 Beyond motif enrichment, GREAT[16] enrichment of genomic regions which

113 decreased in accessibility following GATA1 knockdown were enriched for being near

114 erythroid-specific genes (**Extended Data Fig. 5g**), and genomic regions which

115 increased in accessibility following GATA1 knockdown were enriched for being near

116     megakaryocyte-specific genes (**Extended Data Fig. 5h**). This result is particularly

117     interesting given that K562 cells are often used as a model system for erythro-

118     megakaryocytic progenitor cells. Therefore, knocking down GATA1 in K562 leukemia

119     cells appears to prematurely push cells down a more SP1/RUNX1+, megakaryocyte

120     lineage, despite the fact that GATA1 activity is typically required for this differentiation

121     process[14]. Consistent with this idea, genetic disorders that impair GATA1 function often

122     result in both the dysregulation of erythropoiesis as well as an increased incidence of

123     transient myeloproliferative disorder and/or acute megakaryoblastic leukemia in a

124     subset of patients[17].

125         We next took advantage of the throughput of Spear-ATAC to map the dynamic

126     effects of knocking down transcription factors over time. Traditional proliferation based

127     CRISPR screens evaluate the representation of sgRNAs after up to three weeks in

128     culture; therefore, we evaluated knockdown profiles 3, 6, 9, and 21 days post-

129     knockdown. We introduced a library of 18 sgRNAs targeting 6 transcription factors (n=3

130     sgRNAs each) as well as 3 inert sgRNA controls into K562;dCas9-KRAB cells and

131     performed scATAC-seq across the four time-points (**Fig. 2a** and **Extended Data Fig.**

132     **6a-h**). Similar to a proliferation-based CRISPR screen, representation of sgRNAs can

133     be monitored over time using Spear-ATAC; for example, we observed a significant

134     reduction in the representation of sgGATA1-containing cells at days 9 and 21 compared

135     to days 3 and 6, whereas representation of cells with guides targeting KLF1 remained

136     constant across days 3, 6, and 9 before decreasing at day 21 (**Fig. 2b** and **Extended**

137     **Data Fig. 6f**). To identify hits from the screen – i.e. guides with significant effects on the

138     chromatin landscape – we again used chromVAR to rank TF motif accessibility changes

139    following sgRNA perturbations (**Fig. 2c**). Motif regulatory changes following GATA1 and

140    KLF1 knockdown were the most significant across the genotypes, although the

141    responses to sgGATA1 diminished over time corresponding to a decrease in

142    representation of sgGATA1 cells in the population (**Fig. 2c**). The peaks changing in

143    accessibility also changed over time (**Fig. 2d** and **Extended Data Fig. 6j**); for example,

144    peaks enriched for STAT5 motifs increased in accessibility soon after GATA1

145    knockdown at day 3 but returned to near baseline levels of accessibility at days 6 and 9

146    (**Fig. 2d**). STAT5 is known to be involved in the maintenance of erythroid differentiation

147    in a GATA1-dependent process[18]; therefore, decreased accessibility of peaks

148    containing STAT5 motifs followed by the increased accessibility of peaks containing SPI

149    motifs at days 6 and 9 might further suggest a transition to a more megakaryocyte

150    lineage. Local accessibility near erythroid and megakaryocytic genes also changed as a

151    function of time following knockdown, further emphasizing the importance of timing

152    when evaluating the effects of perturbations on chromatin accessibility (**Fig. 2e** and

153    **Extended Data Fig. 6i**).

154        Finally, to test the ability of Spear-ATAC to screen the chromatin accessibility

155    effect of transcription factors in high-throughput, we evaluated the effects of knocking

156    down 38 transcription factors expressed in K562;dCas9-KRAB leukemia cells with 2-3

157    sgRNAs each, in addition to 15 control non-targeting sgRNAs and 16 sgRNAs targeting

158    essential genes (**Fig. 3a** and **Extended Data Fig. 7a-d**). We chose a variety of

159    transcription factors with growth effects when knocked down in K562 cells (Growth TFs)

160    as well as ones with no proliferation phenotype following knockdown (Non-Growth

161    TFs)[19] (**Fig. 3a**). Overall, we captured 32,832 nuclei representing 128 sgRNA genotypes

162    across six Spear-ATAC samples, with on average 372 single cells being assigned to

163    each sgRNA target with high specificity (**Extended Data Fig. 7a**). We next used

164    chromVAR to rank motif accessibility changes following sgRNA perturbations and

165    identified the top sgRNA:TF motif associations. We consistently identified the

166    sgGATA1:GATA and sgKLF:KLF pairs as well as additional pairs such as

167    sgNFE2:NFE2 and sgFOSL1:FOSL (**Fig. 3b**). Similarly, for sgRNA genotypes that

168    resulted in strong motif accessibility differences compared to sgNT-containing cells, the

169    motifs identified were often consistent with the targeted transcription factor, as shown

170    for GATA1, NFE2, KLF1, FOSL1, and NRF1 (**Fig. 3c**).

171         To establish regulatory relationships between TFs, we measured the effects of

172    TF perturbations on co-varying regulatory networks[7] (**Fig. 3d**). To identify these

173    perturbed co-varying networks, we subtracted the TF-TF motif accessibility correlations

174    within the non-targeting cells by the targeting cells. We first analyzed these relationships

175    for the strongest target perturbation, sgGATA1. We identified 5 modules of TF motifs

176    that are differentially perturbed by GATA1 knockdown (**Fig. 3e**). From this analysis, we

177    more unbiasedly found that depletion of GATA1 led to increased coordinated activity of

178    Module 1 consisting of crucial hematopoietic TFs such as SPI1, IRF1, RUNX and

179    others. To further test the specificity and performance of Spear-ATAC in additional cell

180    lines, we performed this same K562-optimized TF screen in GM12878;dCas9-KRAB

181    lymphoblastic cells and MCF7;dCas9-KRAB breast cancer cells. Overall, we captured

182    an additional 12,175 cells with sgRNA associations between the two cell lines. As

183    expected, the sgRNA:TF motif perturbations were strongest and specific to K562 cells,

184    highlighting cell-type specificity for TF regulation. However, shared patterns of co-

9

185    varying regulatory networks were also uniquely observed in MCF7;dCas9-KRAB cells

186    following knockdown of HINFP, CUX1, and NRF1 (**Extended Data Fig. 9a-c**). While

187    HINFP, CUX1, and NRF1 have not previously been shown to directly interact with each

188    other, HINFP and CUX1 are both involved with histone H4 gene regulation and their

189    overlapping regulatory networks suggest a common pathway[20].

190        In conclusion, Spear-ATAC can be used to evaluate the effects of perturbing

191    transcription factor expression on gene regulatory networks, increasing the throughput

192    of previous methods by between 35- and 100-fold (depending on target nuclei capture

193    rate) and decreasing cost by 20-fold. An exciting application for Spear-ATAC will involve

194    the creation of transcription factor interaction maps following multiplexed perturbations,

195    which will enable a higher-level understanding of how proteins interact to regulate the

196    non-coding genome. We additionally envision the application of this method following

197    the perturbation of individual regulatory elements through high-fidelity editing methods

198    such as prime editing[21], allowing a quantitative understanding of how disease-related

199    mutations alter transcription factor occupancy (as inferred by ATAC-seq) and

200    accessibility at these sites. Spear-ATAC also enables facile monitoring of pooled

201    epigenetic perturbations across time, providing insight into the timescales involved in

202    epigenetic reprogramming. Given the time-dependent differences we observe in our

203    chromatin accessibility profiles following CRISPRi perturbations, we believe that this

204    temporal dimension of monitoring is crucial for identifying the appropriate timepoint for a

205    given study to exclude or include downstream effects. Finally, the Spear-ATAC workflow

206    is not inherently limited to reading out CRISPR/Cas9 sgRNAs, but could be adapted to

207    identify sample barcodes for higher throughput multiplexing or to read-out dynamic

208    lineage tracing marks to understand the relationship between cells during differentiation

209    or cancer evolution.

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230    **Materials availability:** Plasmids generated in this study are available from the Lead

231    Contact without restriction.

232

233    **Data availability:** We have made available all matrices (peak matrix and chromVAR)

234    are available through AWS (See Supplementary Table 6). We also made the 10x cell

235    ranger atac output files and all scATAC-seq matrices used in this study available

236    through AWS. All sequencing data have been deposited in the Gene Expression

237    Omnibus (GEO) and are awaiting accessioning.

238

239    **Code availability:** All custom code used in this work is available upon request. We

240    additionally will host a Github website that includes the main analysis code used in this

241    study.

242

243

244

245

246

247

248

249

250

251

252

253 **<u>Methods</u>**

254

255 **<u>Experimental Methods</u>**

256

257 ***Cell lines***

258 Human cell lines (K562, GM12878, and MCF7) were a gift from Michael Bassik and

259 Howard Chang's laboratories, who previously purchased them from ATCC. The dCas9-

260 KRAB derivatives used have been validated and published previously[7,22,23].

261 K562;dCas9-KRAB and GM12878;dCas9-KRAB cells were cultured in RPMI media

262 supplemented with 10% FBS, 1% penicillin-streptomycin-glutamate, and 0.1%

263 amphotericin. MCF7;dCas9-KRAB cells were cultured in DMEM media supplemented

264 with 10% FBS, 1% penicillin-streptomycin-glutamate, and 0.1% amphotericin. All cell

265 lines tested negative for mycoplasma using the MycoAlert Mycoplasma Detection Kit

266 (Lonza).

267

268 ***Lentivirus production***

269 All lentiviruses were produced by co-transfecting lentiviral backbones with packaging

270 vectors (delta8.2 and VSV-G) into 293T cells using PEI (Polysciences). The viral-

271 containing supernatant was collected at 48- and 72-hours post-transfection, filtered

272 through a 0.45uM filter, and combined with fresh media to transduce cells. K562 and

273 GM12878 derivatives were transduced by spinfection at 1000g at 37 degrees C for 2

274 hours. MCF7 derivatives were transduced by incubating with viral-containing

275 supernatant for up to two days prior to the first fresh media change. Cell lines were

13

276    incubated with 8ug/mL polybrene (Sigma) to enhance transduction efficiency. Cells

277    were transduced with varying amounts of virus and Spear-ATAC was only performed on

278    cells with an MOI < 0.05 to reduce the likelihood of multiple transduction events per cell.

279

280    ***Spear-ATAC: Cloning and modifications to the sgRNA plasmid backbone***

281    Nextera Read1 and Read2 sequencing adapters flanking the bovine U6 promoter were

282    inserted into the pMJ114 backbone using site-directed mutagenesis and whole-vector

283    amplification as previously described, thereby generating pSP618. sgRNA spacer

284    sequences of interest were inserted into the pSP618 backbone individually using site-

285    directed mutagenesis. These sequences were originally picked from the Dolcetto

286    CRISPRi genome-wide library available on Addgene and a full list is available in

287    Supplementary Table 7. In addition, to allow for sgRNA read-out directly from

288    sequencing the scATAC-seq library, we also cloned in unique, 10-bp sgRNA barcode

289    sequences immediately adjacent to the Nextera Read1 adapter by site-directed

290    mutagenesis (sequences also available in Supplementary Table 7). However, we found

291    that targeted sgRNA amplification followed by targeted sgRNA sequencing gives the

292    highest quality sgRNA:nuclei associations, and so we would recommend cloning in the

293    sgRNA spacer sequences only and using the custom sequencing primer oMCB1672 (5'-

294    GCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAAACTTGCTATGCTGTTTCC

295    AGCTTAGCTCTTAAAC-3') for Read 1 to directly sequence the sgRNA spacer

296    sequence. sgRNA plasmids were mixed at equimolar ratios before making virus and

297    transducing the cells of interest.

298

14

299    *Spear-ATAC: Modifications to the 10x scATAC-seq protocol*

300    sgRNA+ nuclei were prepared as previously described for the 10x Genomics scATAC-

301    seq protocol[10]. During GEM generation, 1.2uL of 50uM oSP1735 (5'-

302    gctacattttacatgataggcttgg-3') was spiked into the Master Mix. Additionally, PCR1

303    following GEM generation was extended from 12 cycles to 15 cycles.

304

305    *Spear-ATAC: Amplification of sgRNA fragments out of the scATAC-seq libraries*

306    Full protocol is described in the Supplementary Note. In brief, after the final scATAC-

307    seq libraries were prepared (~150nM final concentration in 20uL ddH$_2$O), 2.5uL of the

308    libraries were used as input for a targeted sgRNA linear amplification PCR reaction

309    using a 5' biotinylated, sgRNA-specific primer (oSP2053: 5'-

310    GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTaagtatcccttggagaaccaccttg-3') for

311    25 cycles. PCR product was pooled and purified using a Qiagen Minelute kit (Qiagen).

312    Biotinylated sgRNA fragments were then enriched using Streptavidin MyOne C1 beads

313    (ThermoFisher) and re-suspended in 40uL ddH$_2$O, which was used as input for an

314    exponential PCR amplification reaction for 15 cycles using primers corresponding to P5

315    (oSP1594: 5'- AATGATACGGCGACCACCGAGA-3') and an indexed P7-containing

316    primer (5'-

317    CAAGCAGAAGACGGCATACGAGAT**NNNNNNNN**GTGACTGGAGTTCAGACGTGTG-

318    3', where **NNNNNNNN** is replaced with the index of choice).

319

320    **Analytical Methods**

321

15

322     ***Genome and Transcriptome Annotations***

323     All analyses were performed with the hg38 genome. We used the hg38 genome

324     transcripts for gene annotations from "TxDb.Hsapiens.UCSC.hg38.knownGene".

325

326     ***SpearATAC – Aligning sgRNA data***

327     To identify the sgRNA for each single cell we first aligned each sgRNA (conventionally

328     Read1, i.e. for 10x scATAC "R1_001.fastq.gz") to cell barcode (conventionally Index1,

329     i.e. for 10x scATAC "R3_001.fastq.gz") combination to the sgRNA library and cell

330     barcode library respectively. We first compiled the cell barcode library of all cell

331     barcodes with up to 1 mismatch. We additionally created a dictionary of the sgRNA

332     barcodes using "PDict" in R. With these 2 libraries, we read in the 2 fastq reads (Read1

333     and Index1) in 500,000 read chunks using the package "ShortRead" in R. We next

334     matched the Index reads to the cell barcode library using "fmatch" in R. Then, we

335     matched the Read1 reads to the sgRNA library using "chunkDictMatch" in R. We

336     compiled the match results into a data frame and iterated through the full fastq reads.

337     Finally, we created a cell by sgRNA matrix that encompassed the aligned sgRNA for

338     each cell and identified cells that had a high-fidelity sgRNA assignment as having at

339     least 20 sgRNA counts and a specificity of 0.8 to the top target.

340

341     ***SpearATAC – Aligning scATAC data***

342     Raw sequencing data was converted to fastq format using cellranger atac mkfastq (10x

343     Genomics, version 1.2.0). Single-cell ATAC-seq reads were aligned to the hg38

344     reference genome and quantified using cellranger count (10x Genomics, version 1.2.0).

16

345   The current version of Cell Ranger can be accessed here:

346   https://support.10xgenomics.com/single-cell-atac/software/downloads/latest.

347

348   ***SpearATAC – Pre-processing scATAC data***

349   We used ArchR[24] (version 0.9.5) for all downstream scATAC-seq analysis

350   (https://greenleaflab.github.io/ArchR_Website/). We used the fragments files for each

351   sample with their corresponding csv file with cell information. We then created Arrow

352   files using "createArrowFiles" with using the barcodes from the sample 10x CSV file with

353   "getValidBarcodes". This step adds the accessible fragments a genome-wide 500-bp tile

354   matrix and a gene-score matrix. We did not filter doublets because for these screens

355   the cells will not form many discrete clusters and thus not many heterotypic doublets

356   can be identified. We created an ArchRProjec and then filtered cells that had a TSS

357   enrichment below 4 and less than 1,000 fragments. For QC plots, we used

358   "plotGroups", "plotTSSEnrichment" and "plotFragmentSizes". We added the sgRNA

359   assignments for each individual sgRNA and the sgRNA targets. We reduced

360   dimensionality with "addIterativeLSI" (default parameters), added clusters with

361   "addClusters" (default parameters), and added a UMAP with "addUMAP" (default

362   parameters).

363        To improve the fidelity of our SpearATAC sgRNA assignments, we identified the

364   highest quality assignments for each target (similar to Replogle et al. 2020[25]). To

365   perform this analysis, we first created an individual sgRNA by tile matrix and an sgRNA

366   Target by tile matrix. For each target, we identified the top 5,000 increasing and 5,000

367   decreasing peaks between the target and non-targeting that were reproducibly

17

368    regulated when comparing the individual sgRNA to the non-targeting cells. We used

369    these 10,000 differential tile regions to perform an LSI dimensionality reduction and

370    subsequent UMAP (n_neighbors=40, min_dist=0.4, metric="cosine"). We next

371    computed the "PurityRatio" for each sgRNA target cell based on the proportion of

372    nearest neighbors being targeting cells (n = 20). Cells that had a "PurityRatio" greater

373    than 0.9 kept their assignment (greater than 95% of assigned cells met this criterion) for

374    downstream analysis.

375        Following these assignments, we created a reproducible non-overlapping peak

376    matrix with "addGroupCoverages" and "addReproduciblePeakSet" using the sgRNA

377    targets as groups i.e. sgGATA1, sgGATA2, sgNT, and etc. We quantified the number of

378    Tn5 insertions per peak per cell using "addPeakMatrix". We subsequently added motif

379    annotations using "addMotifAnnotations" with the motifs curated and clustered from

380    Vierstra et al 2020[26] (https://www.vierstra.org/resources/motif_clustering). We computed

381    chromVAR deviations for each single cell with "addDeviationsMatrix". To identify the top

382    sgRNA:TF perturbations, we computed the average TF motif deviations for each target

383    and subtracted the average TF motif deviations for the non-targeting cells. By ranking

384    the top sgRNA:TF perturbations by the absolute differences we could distinguish the top

385    hits in each SpearATAC screen. For TF footprinting of GATA we used "plotFootprints"

386    with normalization method "subtract" which substracts the Tn5 bias from the ATAC

387    footprint. When performing motif based analyses, we first ranked all motifs based on

388    variability (relevant to the analysis) and the kept the highest motif for each motif

389    cluster/family identified from Vierstra et al 2020[26]

390    (https://www.vierstra.org/resources/motif_clustering). This filtration step removed

391    redundant motifs which can confound downstream analysis.

392

393    ***SpearATAC – Analyzing K562 Pilot Screen (9-sgRNA)***

394    Following preprocessing of the SpearATAC data, we identified differential peaks for

395    each target vs non-target cells using "getMarkerFeatures" (testMethod="binomial"). We

396    identified differential peaks as those with a |log2FC| greater than 0.5 and FDR less than

397    0.1. Differential peaks for sgGATA1 (up-regulated and down-regulated independently)

398    were used as input to GREAT[16] (Association="Two nearest genes") to identify inferred

399    regulated biological processes (i.e. GO terms). We next computed the average

400    accessibility per peak for each individual sgRNA using "getGroupSE" (scaleTo=10^6).

401    To create a heatmap of differential peaks for each sgRNA of a target with sgNT (see

402    **Extended Data Figure 5d**), we subset by all differential peaks that were |log2FC|

403    greater than 1 and then plotted a k-means (k=4) z-score (log2-transformed) heatmap

404    using "ArchR:::.ArchRHeatmap". To identify motifs enriched in each k-means cluster of

405    peaks we used "ArchR:::.computeEnrichment" with the motifmatches and all peaks as a

406    background set. Lastly, we computed a chromVAR deviations matrix using the

407    ENCODE ChIP seq data set within ArchR with "addArchRAnnotations" ("EncodeTFBS")

408    and "addDeviationsMatrix".

409

410    ***SpearATAC – Analyzing K562 Time-Course Screen (21-sgRNA)***

411    Following preprocessing of the SpearATAC data, we identified differential peaks for

412    each target vs non-target cells using "getMarkerFeatures" (testMethod="binomial") for

19

413   each time point (day3, day6, day9 and day21). For each time point, we identified

414   differential peaks as those with a |log2FC| greater than 0.5 and FDR less than 0.1. We

415   next computed the average accessibility per peak for each time point and individual

416   sgRNA using "getGroupSE" (scaleTo=10^6). To create a heatmap of differential peaks

417   for each sgRNA of a target with sgNT (see **Figure 2d and Extended Data Figure 6j**),

418   we first subset by the union of all differential peaks that were |log2FC| greater than 1 for

419   each time point. Next, we computed the average log2 fold changes for sgRNA target vs

420   the sgNT at that time point (using the pseudobulk matrix above). We further filtered the

421   differential peaks by those peaks that have a |log2FC| greater than 0.25 in at least 1

422   time point. We plotted a k-means (k=6) z-score (log2-transformed) heatmap using

423   "ArchR:::.ArchRHeatmap". To identify motifs enriched in each k-means cluster of peaks

424   we used "ArchR:::.computeEnrichment" with the motifmatches and all peaks as a

425   background set.

426

427   ***SpearATAC – Analyzing Large Screens for K562, GM12878 and MCF7 (128-***

428   ***sgRNA)***

429   Following preprocessing of the SpearATAC data, we identified differential motifs for

430   each target vs non-target cells using "getMarkerFeatures" (testMethod="wilcoxon",

431   bufferRatio=0.95, maxCells=250, useSeqnames="z"). We filtered sgNT-5,6,8,11,12 cells

432   prior to this differential comparison after identifying these sgRNA as outliers while

433   performing pseudobulk PCA analysis. To identify perturbed co-varying regulatory

434   networks (see **Figure 3e and Extended Data Figure 9a-c**), we first got the motif

435   deviations matrix for each screen (K562, GM12878 and MCF7) and filtered cells

436    corresponding to sgNT-5,6,8,11,12. Next, we computed the average motif deviation

437    scores for each sgRNA target. We subtracted the average motif deviation scores from

438    the non-targeting cells. We rank ordered the motifs by the maximum observed average

439    motif deviation score difference (absolute difference) across all targets in each screen.

440    Finally, we rank ordered the motifs by the average of these maxima across all three

441    screens (K562, GM12878 and MCF7). We de-duplicated the motifs in each cluster (see

442    Vierstra et al 2020[26]; https://www.vierstra.org/resources/motif_clustering) to remove

443    redundant motifs. For each sgRNA we computed all TF-TF deviation score correlations

444    for the non-redundant motifs. Each targeting sgRNA was then subtracted by the non-

445    targeting sgRNA TF-TF correlations. This differential correlation matrix was

446    subsequently hierarchal clustered with "hclust" and split into 5 modules with "cutree". A

447    heatmap of the differential correlations for the sgRNA targeting cells was then

448    constructed across all modules.

449

450

451

452

453

454

455

456

457

458

473

474    **Competing interests:** W.J.G. is a consultant for 10x Genomics who has licensed IP

475    associated with ATAC-seq. W.J.G. has additional affiliations with Guardant Health

476    (consultant) and Protillion Biosciences (co-founder and consultant).

477

478

479

480

481

**References**

1. Lambert, S. A. *et al.* The human transcription factors. *Cell* **172,** 650–665 (2018).

2. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38,** 824–844 (2020).

3. Pickar-Oliver, A. & Gersbach, C. A. The next generation of CRISPR-Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.* **20,** 490–507 (2019).

4. Jaitin, D. A. *et al.* Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* **167,** 1883–1896.e15 (2016).

5. Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14,** 297–301 (2017).

6. Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167,** 1853–1866.e17 (2016).

7. Rubin, A. J. *et al.* Coupled Single-Cell CRISPR Screening and Epigenomic Profiling Reveals Causal Gene Regulatory Networks. *Cell* **176,** 361–376.e17 (2019).

8. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10,** 1213–1218 (2013).

9. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14,** 959–962 (2017).

10. Satpathy, A. T. *et al.* Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **37,**

505    925–936 (2019).

506    11. Granja, J. M. *et al.* Single-cell multiomic analysis identifies regulatory programs in

507    mixed-phenotype acute leukemia. *Nat. Biotechnol.* **37,** 1458–1465 (2019).

508    12. Wang, T. *et al.* Identification and characterization of essential genes in the human

509    genome. *Science* **350,** 1096–1101 (2015).

510    13. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring

511    transcription-factor-associated accessibility from single-cell epigenomic data. *Nat.*

512    *Methods* **14,** 975–978 (2017).

513    14. Romano, O. & Miccio, A. GATA factor transcriptional activity: Insights from genome-

514    wide binding profiles. *IUBMB Life* **72,** 10–26 (2020).

515    15. Burda, P., Laslo, P. & Stopka, T. The role of PU.1 and GATA-1 transcription factors

516    during normal and leukemogenic hematopoiesis. *Leukemia* **24,** 1249–1257 (2010).

517    16. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory

518    regions. *Nat. Biotechnol.* **28,** 495–501 (2010).

519    17. Shimizu, R., Engel, J. D. & Yamamoto, M. GATA1-related leukaemias. *Nat. Rev.*

520    *Cancer* **8,** 279–287 (2008).

521    18. Wierenga, A. T. J., Vellenga, E. & Schuringa, J. J. Down-regulation of GATA1

522    uncouples STAT5-induced erythroid differentiation from stem/progenitor cell

523    proliferation. *Blood* **115,** 4367–4376 (2010).

524    19. Gilbert, L. A. *et al.* Genome-scale CRISPR-mediated control of gene repression and

525    activation. *Cell* **159,** 647–661 (2014).

526    20. Mitra, P. *et al.* Identification of HiNF-P, a key activator of cell cycle-controlled

527    histone H4 genes at the onset of S phase. *Mol. Cell. Biol.* **23,** 8110–8123 (2003).

528   21. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand

529        breaks or donor DNA. *Nature* **576,** 149–157 (2019).

530   22. Corces, M. R. *et al.* The chromatin accessibility landscape of primary human

531        cancers. *Science* **362,** (2018).

532   23. Tycko, J. *et al.* Mitigation of off-target toxicity in CRISPR-Cas9 screens for essential

533        non-coding elements. *Nat. Commun.* **10,** 4063 (2019).

534   24. Granja, J. M. *et al.* ArchR: An integrative and scalable software package for single-

535        cell chromatin accessibility analysis. *BioRxiv* (2020).

536        doi:10.1101/2020.04.28.066498

537   25. Replogle, J. M. *et al.* Combinatorial single-cell CRISPR screens by direct guide

538        RNA capture and targeted sequencing. *Nat. Biotechnol.* **38,** 954–961 (2020).

539   26. Vierstra, J. *et al.* Global reference mapping of human transcription factor footprints.

540        *Nature* **583,** 729–736 (2020).

541

542

543

544

545

546

547

548

549

550

551 **Figure Legends**

552 **Figure 1. Spear-ATAC enables high-throughput CRISPR screening with a**

553 **chromatin accessibility read-out.**

554 **a.** Schematic of the Spear-ATAC method. Modifications to traditional CRISPR screening

555 methods and scATAC-seq approaches are outlined on the right.

556 **b.** Heatmap of the percent of sgRNA reads assigned to 3,045 individual cells with

557 corresponding chromatin accessibility information via scATAC-seq.

558 **c.** UMAP of Spear-ATAC chromatin accessibility profiles for the pilot K562 screen

559 colored by sgRNA assignments. Cells unassigned are in light grey.

560 **d.** Rank ordered plot of sgRNA:TF perturbations to identify top hits in the pilot K562

561 screen.

562 **e.** UMAP of Spear-ATAC chromatin accessibility profiles for the pilot K562 screen

563 colored by chromVAR deviations for GATA1 ENCODE ChIP-seq.

564 **f.** (*Top*) Bias-Normalized footprint of the local accessibility for each scATAC-seq cluster

565 for genomic regions containing GATA motifs. (*Bottom*) Modeled hexamer insertion bias

566 of Tn5 around sites containing each motif.

567 **g.** Pseudo-bulk ATAC-seq track at the GATA1 locus for sgGATA1 and sgNT cells. Light

568 grey box indicates the region targeted by sgGATA1-1, sgGATA1-2, and sgGATA1-2

569 CRISPRi sgRNAs.

570 **h.** Differential accessibility between sgGATA1 and sgNT cells. The x-axis represents the

571 log2 mean accessibility per peak and the y-axis represents the log2 fold change in

572 sgGATA1 cells compared to sgNT cells. Colored points are significant ($|\log_2$ fold

26

573    change|>0.5, FDR <0.05).

574

575    **Figure 2. Assessing trans regulatory perturbations over time using Spear-ATAC.**

576    **a.** Schematic of Spear-ATAC time-course experiment with a 21-sgRNA pool analyzed at

577    4 time-points (3 days, 6 days, 9 days, and 21 days post-transduction).

578    **b.** Change in sgGATA1 representation over time, represented by the number of cells

579    analyzed per time-point (red) and the % of cells in the total pool (black).

580    **c.** Rank ordered plot of sgRNA:TF perturbations to identify top hits in the K562 time-

581    course screen at the indicated time points.

582    **d.** (*Left*) Heatmap of chromatin peak accessibility for each scATAC-seq sub-population

583    using the top differential scATAC-seq peaks for sgGATA1. Each row represents a z-

584    score of $\log_2$ normalized accessibility within each group using scATAC-seq. (*Right*)

585    Transcription factor hypergeometric motif enrichment with FDR indicated in

586    parentheses.

587    **e.** Pseudo-bulk ATAC-seq track at the IRF1 locus for sgGATA1 (day3, day6, day9, and

588    day21) and sgNT cells (day3). Light grey box indicates peak regions that increased in

589    accessibility in the sgGATA1 vs sgNT cells.

590

591    **Figure 3. High throughput chromatin accessibility screening of CRISPR**

592    **perturbations using Spear-ATAC.**

593    **a.** Schematic of Spear-ATAC large screens with a 128-sgRNA pool analyzed for 3

594    different cell lines (K562, MCF7 and GM12878).

595    **b.** Rank ordered plot of sgRNA:TF perturbations to identify top hits in the K562 large

596    screen.

597    **c.** Motif accessibility differences across 6 perturbed transcription factors in the K562

598    large screen. Color indicates whether the motif accessibility difference corresponds to

599    up-regulated (red) or down-regulated (blue) motifs.

600    **d.** Schematic of inferring regulatory relations with Spear-ATAC. Briefly, the correlation

601    for each TF-TF motif is determined in both the targeting (sgT) and non-targeting (sgNT)

602    cells. Next the correlation in the non-targeting cells is subtracted from targeting cells.

603    These TF-TF motif pairs are then assessed for different regulatory relationships.

604    **e.** (*Left*) Heatmap of the differences between sgGATA1 targeting and non-targeting cells

605    for all TF-TF motif accessibility correlations grouped into 5 different modules in the K562

606    large screen. (*Right*) Zoomed-in heatmap of modules 1 and 2 highlighting transcription

607    factors with largely perturbed TF-TF motif accessibility relationships.

608

609

610

611

612

613

614

615

616

617

618 **Extended Data Figure Legends**

619

620 **Extended Data Figure 1. Outline of Spear-ATAC protocol.**

621 **a.** Cells are transduced with a lentiviral sgRNA flanked by Read1/Read2 Nextera

622 adapters, FACS sorted to exclude sgRNA-negative cells, and processed for scATAC-

623 seq using a modified 10x Genomics droplet-based protocol. In brief, nuclei are isolated,

624 transposed, and gel bead emulsions (GEMs) are made with individual nuclei re-

625 suspended in Nuclei Resuspension Buffer combined with an enzymatic master mix

626 containing a sgRNA-specific primer. GEMs are immediately subjected to in-GEM linear

627 amplification of scATAC-seq fragments, while sgRNA fragments are subjected to

628 exponential amplification using the sgRNA-specific primer as a Forward primer. The

629 number of cycles of in-GEM amplification has been extended from 12 cycles (original

630 10x protocol) to 15 cycles (Spear-ATAC protocol). Following a series of clean-ups and

631 exponential amplification to attach flow-cell adapters (P7) and sample indices to the

632 ATAC-seq fragments, sgRNA fragments are specifically enriched using a biotin-

633 conjugated oligo specific to the sgRNA fragments, and then amplified and sequenced

634 separately.

635

636 **Extended Data Figure 2. Flanking a lentiviral sgRNA spacer with Read1/Read2**

637 **adapters increases sgRNA fragment capture efficiency following ATAC-seq.**

638 **a.** Schematic of traditional lentiviral sgRNA.

639 **b.** Schematic of modified Spear-ATAC lentiviral sgRNA. sgRNA spacer sequence is

640 flanked by Read1/Read2 Nextera adapters, obviating the need for tn5 transposase to

29

641    randomly insert sequenc- ing adapters nearby the sgRNA sequence during an ATAC-

642    seq transposition reaction. The Read2 adapter is inserted in a flexible region of the U6

643    promoter to decrease the total length of the sgRNA fragment.

644    **c.** Log2 fold change in sgRNA fragments amplified following bulk ATAC-seq reactions

645    performed with the indicated numbers of input cells transposed with lentiviral sgRNAs

646    with and without Nextera adapters flanking the sgRNA region.

647    **d.** Percentage of GFP negative cells following lentiviral transduction of sgGFP

648    sequences with and without Nextera adapters flanking the sgRNA region.

649

650    **Extended Data Figure 3. Spear-ATAC modifications to the 10x scATAC-seq**

651    **protocol do not hurt scATAC quality.**

652    **a-b.** scATAC-seq quality control metrics with and without a sgRNA-specific primer

653    spiked into the enzymatic master mix during GEM formation on the 10x Controller.

654    Insertion profiles (left a), TSS enrichment (right a), and average fragment distribution (b)

655    in each sample is shown.

656    **c-d.** scATAC-seq quality control metrics after 12, 13, 14, and 15 cycles of in-GEM linear

657    amplification. Insertion profiles (left c), TSS enrichment (right c), and average fragment

658    distribution (d) in each sample is shown.

659    **e.** Biotin tagging and selection increases the specificity of targeted sgRNA amplification,

660    as seen by the fragment distribution of products without (left) and with (right) the first

661    biotin tagging step.

662

663    **Extended Data Figure 4. Quality control metrics and sgRNA assignment methods**

664    **for 9-sgRNA GATA pool.**

665    **a.** Schematic of sgRNAs in the first proof-of-principle 9-sgRNA pool.

666    **b.** scATAC-seq quality control metrics for the proof-of-principle 9-sgRNA GATA pool.

667    Average TSS enrichment (left), individual TSS enrichment x log10 number of unique

668    fragments per cell (middle), and average insertion profiles (right inset) for the 6,390

669    nuclei captured.

670    **c.** UMAP of Spear-ATAC chromatin accessibility profiles (N = 6,390) for the 9-sgRNA

671    K562 screen colored by Seurat graph clustering in ArchR.

672    **d.** Specificity to a single sgRNA spacer sequence and number of sgRNA reads (log10)

673    following targeted amplification for each cell analyzed.

674    **e.** UMAP of Spear-ATAC chromatin accessibility profiles for the 9-sgRNA K562 screen

675    colored by the individual sgRNA. Cells unassigned are in light grey.

676

677    **Extended Data Figure 5. GATA1 maintains accessibility at genomic regions**

678    **important for erythropoiesis and knocking down GATA1 results in altered**

679    **regulation of GATA3.**

680    **a.** Pseudo-bulk ATAC-seq track at the GATA1 locus for individual sgRNAs of sgGATA1

681    and sgNT. Light grey box indicates the region targeted by sgGATA1-1, sgGATA1-2, and

682    sgGATA1-2 CRISPRi sgRNAs.

683    **b.** UMAP of Spear-ATAC chromatin accessibility profiles for the pilot K562 screen

684    colored by chromVAR deviations for GATA3 ENCODE ChIP-seq.

685     **c.** Pseudo-bulk ATAC-seq track at the GATA3 locus for sgGATA1 and sgNT cells. Light

686     grey boxes indicate regions that are significantly changing in accessibility following

687     sgGATA1 knockdown.

688     **d.** Heatmap of chromatin peak accessibility for each individual sgRNA for both

689     sgGATA1 and sgNT using the top differential scATAC-seq peaks identified from ArchR.

690     Each row represents a z-score of log2 normalized accessibility within each group using

691     scATAC-seq. (Right) Transcription factor hypergeometric motif enrichment with FDR

692     indicated in parentheses.

693     **e.** Ridge plot for chromVAR deviations (normalized to the median sgNT) of individual

694     cells for (Left) GATA, (Middle) RUNX and (Right) SPI.

695     **f.** ChromVAR motif correlations for the indicated transcription factor families (RUNX,

696     SPI, or GATA). Pearson correlation (R) is indicated in the top left corner. Each individual

697     point represents a single cell. Cells are colored based on their sgRNA genotype, as

698     indicated in Figure 5b.

699     **g.** GREAT enrichment of peaks that decrease in accessibility with sgGATA1.

700     Enrichments are taken from the Mouse Phenotype with Single KO (Knock-out) dataset.

701     **h.** GREAT enrichment of peaks that increase in accessibility with sgGATA1.

702     Enrichments are taken from the GO Process dataset.

703

704     **Extended Data Figure 6. Spear-ATAC increases the throughput of time-course**

705     **screens to understand the timing of effects following transcription factor**

706     **perturbations.**

707     **a.** Schematic of 21 sgRNAs used in the Spear-ATAC K562 time-course screen.

708 **b.** Specificity to a single sgRNA spacer sequence and number of sgRNA reads (log10)

709 following targeted amplification for each cell analyzed.

710 **c.** TSS enrichment scores per single cell across all Spear-ATAC K562 time-course time

711 points.

712 **d.** Aggregate TSS enrichment scores across all Spear-ATAC K562 time-course time

713 points.

714 **e.** Aggregate ATAC-seq fragment size distributions across all Spear-ATAC K562 time-

715 course time points.

716 **f.** Aggregate ATAC-seq fragment size distributions across all Spear-ATAC K562 time-

717 course time points.

718 **g.** Change in sgKLF1 representation over time, represented by the number of cells

719 analyzed per time-point (orange) and the % of cells in the total pool (black). Cells

720 unassigned are in light grey.

721 **h.** Ridge plot for chromVAR deviations (normalized to the median sgNT) of individual

722 cells across each time point and target sgRNA for (Left) GATA, (Middle) SPI and (Right)

723 KLF.

724 **i.** Pseudo-bulk ATAC-seq track at the (Top Left) RUNX1, (Top Right) PRKAR2B,

725 (Bottom Left) PPBP and (Bottom Right) MPL locus for sgGATA1 (day3, day6, day9, and

726 day21) and sgNT cells (day3). Light grey box indicates peak regions that changed in

727 accessibility in the sgGATA1 vs sgNT cells.

728 **j.** (*Left*) Heatmap of chromatin peak accessibility for each scATAC-seq sub-population

729 using the top differential scATAC-seq peaks for sgKLF1. Each row represents a z-score

730 of $\log_2$ normalized accessibility within each group using scATAC-seq. (*Right*)

731    Transcription factor hypergeometric motif enrichment with FDR indicated in

732    parentheses.

733

734    **Extended Data Figure 7. Quality control metrics for Spear-ATAC transcription**

735    **factor screens in K562;dCas9-KRAB, GM12878;dCas9-KRAB, and MCF7;dCas9-**

736    **KRAB cell lines.**

737    **a.** Specificity to a single sgRNA spacer sequence and number of sgRNA reads (log10)

738    following targeted amplification for each cell analyzed for the (Left) K562, (Middle)

739    GM12878 and (Right) MCF7 large screens.

740    **b.** TSS enrichment scores per single cell across the (Left) K562, (Middle) GM12878 and

741    (Right) MCF7 large screen replicates.

742    **c.** Aggregate ATAC-seq fragment size distributions across the (Left) K562, (Middle)

743    GM12878 and (Right) MCF7 large screen replicates.

744    **d.** UMAP of Spear-ATAC chromatin accessibility profiles for the (Left) K562, (Middle)

745    GM12878 and (Right) MCF7 large screens. This UMAP is colored by (Top)

746    experimental replicate and (Bottom) assigned sgRNA. Cells unassigned are in light

747    grey.

748

749    **Extended Data Figure 8. Transcription factor motif accessibility changes**

750    **following perturbations in K562;dCas9-KRAB, GM12878;dCas9-KRAB, and**

751    **MCF7;dCas9-KRAB cell lines.**

752    **a.** Rank ordered plot of sgRNA:TF perturbations to identify top hits for the (Left) K562,

753    (Middle) GM12878 and (Right) MCF7 large screens.

754    **b.** Ridge plot for chromVAR deviations (normalized to the median sgNT) of individual

755    cells for (Left) GATA, (Middle Left) NFE2, (Middle Right) KLF and (Right) FOSL across

756    the K562 large screen.

757

758    **Extended Data Figure 9. Patterns of differential transcription factor motif**

759    **correlations in MCF7;dCas9-KRAB cells following knockdown of HINFP, CUX1,**

760    **and NRF1.**

761    **a.** (Left) Heatmap of the differences between sgHINFP targeting and non-targeting cells

762    for all TF-TF motif accessibility correlations grouped into 5 different modules in the

763    MCF7 large screen. (Right) Zoomed in heatmap of modules 1 and 2 highlighting

764    transcription factors with largely perturbed TF-TF motif accessibility relationships.

765    **b.** (Left) Heatmap of the differences between sgCUX1 targeting and non-targeting cells

766    for all TF-TF motif accessibility correlations grouped into 5 different modules in the

767    MCF7 large screen. (Right) Zoomed in heatmap of modules 1 and 2 highlighting

768    transcription factors with largely perturbed TF-TF motif accessibility relationships.

769    **c.** (Left) Heatmap of the differences between sgNRF1 targeting and non-targeting cells

770    for all TF-TF motif accessibility correlations grouped into 5 different modules in the

771    MCF7 large screen. (Right) Zoomed in heatmap of modules 1 and 2 highlighting

772    transcription factors with largely perturbed TF-TF motif accessibility relationships.

773

774

775

776

777 **Supplementary Note**

778

779 **Supplementary Note 1. Protocol for targeted amplification of sgRNAs out of final**

780 **scATAC-seq libraries.**

781 This supplementary note contains a full protocol for the targeted amplification of sgRNA

782 fragments out of the final 10x scATAC-seq libraries made with Spear-ATAC.

783

784 **Supplementary Tables**

785

786 **Supplementary Table 1. Spear-ATAC experiment statistics.**

787 This table contains information about each Spear-ATAC dataset generated in this study

788 including QC statistics and number of sgRNA to cell associations captured.

789

790 **Supplementary Table 2. Motif accessibility results from K562 pilot Spear-ATAC**

791 **screen.**

792 This table contains information for the sgRNA:TF motif accessibility changes between

793 targeting and non-targeting cells for the K562 pilot screen.

794

795 **Supplementary Table 3. Motif accessibility results from K562 time-course Spear-**

796 **ATAC screen.**

797 This table contains information for the sgRNA:TF motif accessibility changes between

798 targeting and non-targeting cells for the K562 time-course Spear-ATAC screen.

799

800   **Supplementary Table 4. Motif accessibility results from K562, GM12878, and**

801   **MCF7 transcription factor Spear-ATAC screens.**

802   This table contains information for the sgRNA:TF motif accessibility changes between

803   targeting and non-targeting cells for the K562, GM12878, and MCF7 transcription factor

804   Spear-ATAC screens.

805

806   **Supplementary Table 5. List of Vierstra motif clustering annotations used for all**

807   **motif analyses.**

808   This table contains the motif clustering annotations used for motif de-duplication and

809   motif based analyses.

810

811   **Supplementary Table 6. Spear-ATAC data URLs.**

812   This data contains information for how to access the Spear-ATAC data for each

813   experiment in this study.

814

815   **Supplementary Table 7. List of sgRNA spacer sequences used in this study.**

816   This table contains information about the sgRNA spacer sequences used in this study
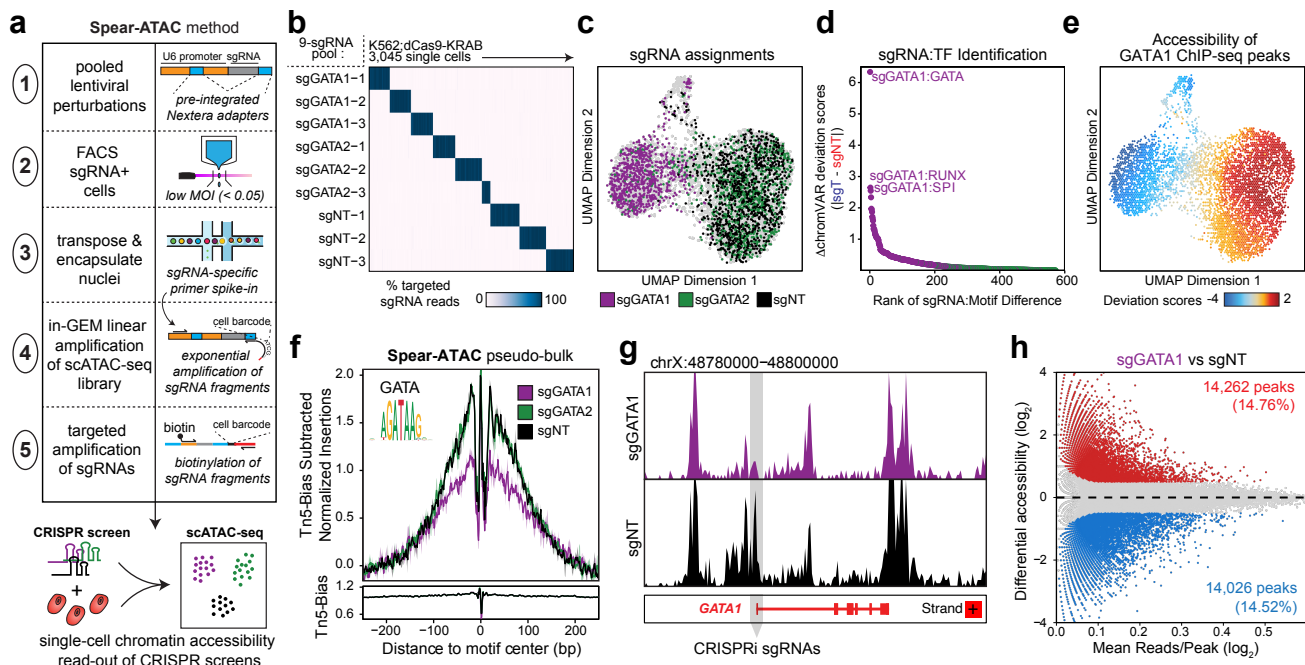
817   and their targets.

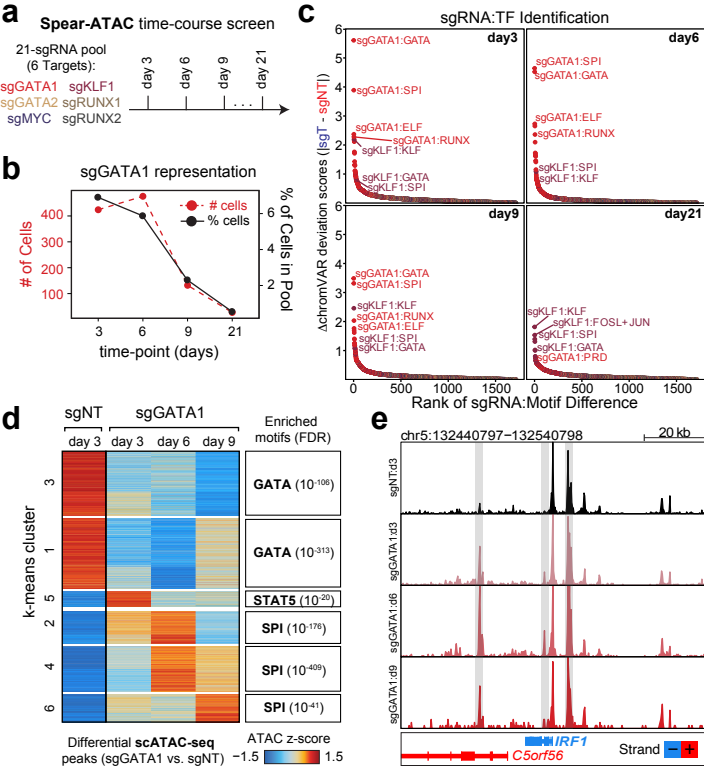818

# Figure 1

# Figure 2



**a** Spear-ATAC time-course screen

21-sgRNA pool
(6 Targets):
sgGATA1 sgKLF1
sgGATA2 sgRUNX1
sgMYC sgRUNX2

day 3 day 6 day 9 day 21

**b** sgGATA1 representation

**c** sgRNA:TF Identification

day3
sgGATA1:GATA
sgGATA1:SPI
sgGATA1:ELF
sgGATA1:RUNX
sgKLF1:KLF
sgKLF1:GATA
sgKLF1:SPI

day6
sgGATA1:SPI
sgGATA1:GATA
sgGATA1:ELF
sgGATA1:RUNX
sgKLF1:SPI
sgKLF1:KLF

day9
sgGATA1:GATA
sgGATA1:SPI
sgKLF1:KLF
sgGATA1:RUNX
sgGATA1:ELF
sgKLF1:SPI
sgKLF1:GATA

day21
sgKLF1:KLF
sgKLF1:FOSL+JUN
sgKLF1:SPI
sgKLF1:GATA
sgGATA1:PRD

ΔchromVAR deviation scores (|sgT⁻ − sgNT|)

Rank of sgRNA:Motif Difference

**d**

sgNT | sgGATA1 | Enriched motifs (FDR)

k-means cluster

3 — GATA (10⁻¹⁰⁶)
1 — GATA (10⁻³¹³)
5 — STAT5 (10⁻²⁰)
2 — SPI (10⁻¹⁷⁶)
4 — SPI (10⁻⁴⁰⁹)
6 — SPI (10⁻⁴¹)

Differential scATAC-seq peaks (sgGATA1 vs. sgNT)

ATAC z-score −1.5 ... 1.5

**e** chr5:132440797−132540798    20 kb

sgNT:d3
sgGATA1:d3
sgGATA1:d6
sgGATA1:d9

IRF1
C5orf56    Strand − +

# Figure 3