

1 **Landscape and dynamics of the transcriptional regulatory network** 2 **during natural killer cell differentiation**

3

4 Kun Li^{1,3}, Yang Wu^{2,3}, Young Li¹, Qiaoni Yu¹, Zhigang Tian², Haiming Wei^{2,*}, Kun Qu^{1,*}

5

6 Affiliation

7

8 ¹Division of Molecular Medicine, Hefei National Laboratory for Physical Sciences at Microscale,
9 the CAS Key Laboratory of Innate Immunity and Chronic Disease, CAS Center for Excellence in
10 Molecular Cell Science, School of Life Sciences, University of Science and Technology of China.

11 ²Division of Molecular Medicine, Hefei National Laboratory for Physical Sciences at Microscale,
12 the CAS Key Laboratory of Innate Immunity and Chronic Disease, School of Life Sciences,
13 University of Science and Technology of China.

14 ³Co-first authors

15

16 *Correspondence: Kun Qu (qukun@ustc.edu.cn) and Haiming Wei (ustcwhm@ustc.edu.cn)

17

18 **Contact Information**

19

20 Kun Qu Ph.D.

21 Division of Molecular Medicine, Hefei National Laboratory for Physical Sciences at Microscale,
22 the CAS Key Laboratory of Innate Immunity and Chronic Disease, CAS Center for Excellence in
23 Molecular Cell Science, School of Life Sciences, University of Science and Technology of China.

24 Hefei, Anhui, China, 230027

25 Email: qukun@ustc.edu.cn

26 Phone: +86-551-63606257

27

28

29 **Abstract**

30

31 Natural killer (NK) cells are essential in controlling cancer and infection. However, little is known
32 about the dynamics of the transcriptional regulatory machinery during NK cell differentiation. In
33 this study, we applied assay of transposase accessible chromatin with sequencing (ATAC-seq)
34 technique in a self-developed *in vitro* NK cell differentiation system. Analysis of ATAC-seq data
35 illustrated two distinct transcription factor (TF) clusters that dynamically regulate NK cell
36 differentiation. Moreover, two TFs from the second cluster, FOSL2 and EGR2, were identified as
37 novel essential TFs that control NK cell maturation and function. Knocking down either of these
38 two TFs significantly impacted NK cell transformation. Finally, we constructed a genome-wide
39 transcriptional regulatory network that provides an understanding of the regulatory dynamics
40 during NK cell differentiation.

41

42 **Keywords**

43

44 Natural killer; ATAC-seq; programmed differentiation; FOSL2 and EGR2; dynamic regulatory
45 networks

46

47 **Introduction**

48

49 Natural killer (NK) cells are innate lymphocytes, and as the name suggests, NK cells
50 secreting inflammatory cytokines and directly killing infected cells represent the early defense line
51 during pathogen invasion, which surveys the environment and to protect the host from infection
52 or cancer cells. [1]. Additionally, NK cell-based immunotherapy has become an emerging force in
53 cancer treatment and will play an essential role in the future treatment [2-6]. For instance, adoptive
54 NK cell immunotherapy has become increasingly popular because it induces graft-versus-
55 leukemia effects without causing graft-versus-host disease in patients [7]. Therefore, major efforts
56 are currently underway to fully utilize the anti-tumor properties of NK cells in the clinic. In
57 addition, a variety of methods include the development of NK cell expansion protocols to establish
58 a microenvironment that favors NK cell activity, redirect the activity of NK cells on tumor cells
59 and release the inhibitory signals that limit NK cell function [8]. On the other hand, many

60 researchers used umbilical cord blood (UCB) CD34⁺ cells to produce abundant NK cells [9-18]
61 that were used for clinical application without feeding cells by adding various cytokines [10]. In
62 our previous work, we have developed a method to obtain sufficient functional NK cells by simply
63 adding a mixture of cytokines and providing a mechanism by which NK cells can be used to treat
64 leukemia [19]. The use of NK cells for immunotherapy relies on a great number of NK cells with
65 optimal cytotoxic activity [4]; therefore, a comprehensive understanding of the regulatory circuits
66 during NK cell differentiation is particularly important for boosting the efficacy of clinical
67 treatments. However, the mechanisms underlying the differentiation of the NK cells are not well
68 understood. Current studies have shown that transcriptional factors (TFs) play an essential role in
69 driving NK cell maturation, and many TFs have been well studied in this process [1]. Additionally,
70 it is known that different TFs play various roles at distinct stages of differentiation [1]. For example,
71 *PU.1* is a TF that is known to drive hematopoietic stem cell (HSC) differentiation into the earliest
72 myeloid and lymphoid progenitors [20], whereas *T-bet* is an essential TF in the control of NK cell
73 maturation and IFN γ secretion regulation [21]. However, how TFs work in concert to enforce the
74 NK cell phenotype is not clear. In our work, we investigated the landscape and dynamics of the
75 transcriptional regulatory network during NK cells differentiation from cord blood CD34⁺ HSCs
76 based on a newly developed epigenomic profiling technique named ATAC-seq [22].

77 ATAC-seq has been widely used to profile the epigenetic landscapes of cells at specific
78 stages of interest and thereby delineate the underlying regulatory mechanisms of gene expression.
79 For example, a previous report used ATAC-seq to help identify an exhaustion-specific enhancer
80 that regulates PD-1 expression and thereby elucidated the regulatory mechanism of gene
81 expression in exhausted CD8⁺ T cells [23, 24]. Alternately, ATAC-seq analysis of pure cancer cell
82 populations of human small cell lung cancer (SCLC) identified a novel TF *Nfib* which is necessary
83 and sufficient to drive the metastatic ability of SCLC cells [25]. ATAC-seq has also enabled
84 researchers to track the epigenomic states changes in patient-derived immune cells [26], and to
85 survey how the personal regulomes of the cutaneous T cell lymphoma patients determines their
86 responses towards histone deacetylase inhibitors (HDACi) anti-cancer drugs [27]. More recently,
87 ATAC-seq was applied to better understand the controlling of NK cells in innate immune memory
88 during infection [28], illustrating the importance of the topic as well as the power of the technique
89 used in this study. Here, we have developed a systematic method to characterize the chromatin
90 accessibility and construct regulatory network dynamics during NK cell differentiation based on

91 ATAC-seq. Motif and enrichment analysis from HOMER [29] and Genomica [26] show that many
92 TFs play important roles in NK cell differentiation, and by integrating gene expression profiles
93 from our previous study [19], we identified two novel transcription factors, *FOSL2* and *EGR2*, that
94 are essential for driving NK cell maturation.

95

96 **Results**

97

98 **Landscape of DNA accessibility during NK cell differentiation**

99

100 To elucidate the regulatory networks during NK cell differentiation, we developed a culture
101 procedure to obtain differentiated NK cells from UCB CD34⁺ cells using a cocktail of cytokines
102 [19]. The differentiation process requires 35 days. Interestingly, after culturing of cord blood stem
103 cells for 3 weeks, the proportion of NK cells rapidly increased from 5% on day21 to approximately
104 60% on day 28, and peaking at close to 100% on day 35 (**Figure 1A** and **Figure S1A**) [19]. To
105 elucidate the molecular mechanism and regulatory network underlying NK cell differentiation, we
106 interrogated the landscape of chromatin accessibility using ATAC-seq at 8 different time points,
107 with 2 replicates each along the process (**Figure 1B**). Multiple bioinformatics analysis methods
108 (see **Materials and Methods**) were then applied to obtain the differential accessible chromatin
109 sites, enriched transcription factors (TFs) and genome-wide regulatory elements. At least 50,000
110 cells were obtained from each sample, and on average, resulted more than 78 million reads with a
111 total of 1260 million measurements (**Table S1**). From this dataset, 143,570 DNA accessible sites
112 were identified ($P < 10^{-7}$, $FDR < 10^{-7}$). Quantitative analysis indicated that the dataset was of high
113 quality with a strong signal to background noise ratio and expected fragment length distribution
114 (**Figure S1B-E**). Moreover, the dynamics of the chromatin accessibilities around the known
115 surface marker genes were consistent with changes of their gene expressions during NK cell
116 differentiation (**Figure S1D**) [30, 31]. Pearson correlation analysis on the biological replicates at
117 each time point also showed that our ATAC-seq profiles were highly reproducible (**Figure S2A-**
118 **B**). Within all the accessible sites, only a very small portion (6.48%) were conserved across all
119 stages (**Figure S3A**), while the majority of peaks changed over time, suggesting significant
120 chromatin dynamics during the NK cell differentiation. Peaks that were detected at all stages were
121 enriched in gene ontology (GO) terms such as RNA transcription and other functions to maintain

122 basic physiological activities (**Figure S3B**). In contrast, stage-specific peaks, especially peaks that
123 emerged at later stage of the process, were highly enriched in immune relevant GO terms,
124 suggesting the activation of critical genes that govern NK cell function.

125 For instance, several TFs are known to regulate NK cell differentiation, such as SPI1 and
126 TBX21. The former is a NK cell repressor and the latter an activator. Gene expression profiles
127 from our previous study [19] indicated down-regulation of the SPI1 gene and up-regulation of the
128 TBX21 gene during NK cell differentiation (**Figure 1C**). The protein levels of these two genes
129 obtained from flow cytometric experiments suggested the same trends (**Figure 1D**). ATAC-seq
130 revealed two regulatory elements at the promoters and intronic enhancers around each of these two
131 genes (**Figure 1E**). Chromatin-accessible sites around the *SPI1* gene clearly became unavailable
132 and those around the *TBX21* gene gradually became available during the process, supporting the
133 notion that the epigenetic dynamics of key regulators are consistent with their corresponding gene
134 and protein expressions. The consistency between the epigenetic and gene expression profiles of
135 transcription factors, such as GATA3 and EOMES, which are known to drive NK cell
136 differentiation, are shown in (**Figure S1E**), indicating the feasibility of predicting transcriptional
137 regulatory networks from ATAC-seq profiles.

138

139 **Epigenomic signatures of different stages during NK cell differentiation**

140

141 To determine the differences in regulatory DNA activity among different stages during NK
142 cell differentiation, we applied pairwise comparisons of the ATAC-seq signals between the
143 corresponding samples, together with intrinsic analysis [32], a method that highlighted elements
144 that varied in accessibility across individuals but not between repeat samples from the same
145 individual. We discovered 6401 peaks of differential DNA accessible sites across the genome, and
146 identified three distinct clusters of chromatin accessible sites via unsupervised hierarchical
147 clustering (**Figure 2A, Figure S4A, B**). Principal component analysis (PCA) of all the samples
148 also illustrated three distinct clusters, which were consistent with the time process of NK cell
149 differentiation representing the early, intermediate and late stages of the entire process (**Figure**
150 **S4C**). Gene ontology analysis of these peaks was performed in GREAT [33]. Cluster I comprises
151 of 1584 elements that were more accessible at the early stage (days7~21) of differentiation. Peaks
152 in cluster I were enriched in the GO terms of metabolic processes, with a certain level of

153 significance (**Figure 2B**, $10^{-8} < P < 10^{-4}$), keeping cells alive, proliferation and prepare to differentiate.
154 Cluster II comprises 4817 elements, which were highly accessible at the late stage (days 24~35)
155 of NK cell differentiation. GREAT analysis revealed that peaks in cluster II were strongly enriched
156 ($P < 10^{-50}$) for immune-relevant GO terms, such as immune system process, immune response and
157 immune system development among others (**Figure 2B**), indicating that the epigenetic states of
158 functional immune cell-specific genes were activated throughout the process and then drove
159 progenitor cells differentiation toward NK cells. Cluster III consisted of only 356 peaks, which
160 showed no enrichment for GO terms and were therefore neglected in downstream analysis.

161 We next examined whether the DNA accessibility chromatin signatures at different stages
162 correlated with those of the corresponding gene expressions. By comparing the ATAC-seq profiles
163 with the genome-wide microarray data during NK cell differentiation, we found that gene loci that
164 gained chromatin accessibility were significantly increased in the gene expression level ($P < 10^{-6}$),
165 while gene loci that lost chromatin accessibility had decreased expression ($P < 10^{-9}$), indicating a
166 high correlation between epigenetic and RNA profiles (**Figure 2C**).

167 Chromatin structure and epigenetic modifications regulate gene expression [34], however,
168 the chronological order of the dynamics of the chromatin accessibility, gene expression, and cell
169 phenotype have not yet been well studied. Here, we integrated the ATAC-seq signals, microarray
170 profile and percentage of NK cell counts to examine the temporal dynamics of these three features.
171 Interestingly, we noticed that the accessibility of mature NK cell-specific peaks (cluster II in
172 **Figure 2A**) started to open up on day 14; and the expression of NK cell specific genes was turned
173 on approximately two days later, while the percentage of NK cell started to grow after day 21
174 (**Figure 2D**). These results suggested a clear chronological order of the changes in chromatin
175 structure, gene expression and cell phenotype during NK cell differentiation.

176

177 **Transcription factor occupancy network during NK cell differentiation**

178

179 Transcription factors bind to cognate DNA sequences with patterns termed motifs, and
180 therefore, we can predict TFs occupancy on chromatin by DNA accessibility data from ATAC-
181 seq, and use it to create regulatory networks [26]. To identify potential drivers of NK cell
182 differentiation, we applied HOMER and used both the default (**Figure 3A**) and setting background
183 (**Figure S5A**) search modes (see **Materials and Methods**) to search for TFs that were enriched at

184 accessible elements in cluster I and cluster II suggesting that NK cell differentiation and maturation
185 require a variety of transcription factors. TFs enriched at cluster I peaks are potential regulators at
186 the early stage of NK differentiation, while those enriched at cluster II peaks could be critical
187 regulators at the late stage. We found several TF families that were significantly enriched ($P < 10^{-10}$)
188 ¹⁰), and many of which are important known TFs of NK cells differentiation, such as the RUNX
189 and ETS1/PU.1 families [20, 35] and CEBP [36], supporting the reliability of this method to detect
190 critical regulators. For instance, PU.1 is widely expressed and controls multiple stages of bone
191 marrow and lymphocyte differentiation in a variety of hematopoietic-derived lineages [37-41], and
192 a reduction in the number of NKPs and iNKs has been detected in chimeric mice [20], indicating
193 that PU.1 may play an important role during the early stages of NK cell differentiation. Several
194 known TFs are also enriched at the late stage, such as RUNX [35], E2A [42], T-bet [21], STAT5
195 [43], and EOMES [21]. The most enriched motif in cluster II was the RUNX family: RUNX1,
196 RUNX and RUNX2 (**Figure S5A**), which are key regulators of lymphocyte lineage-specific gene
197 expression [44].

198 Motif searches of differential ATAC-seq peaks could potentially provide information about
199 the transcriptional regulatory network, but one caveat of this method is unable to distinguish
200 between similar motifs of binding TF family members. Thus, we sought to apply the Genomica's
201 module map algorithm and "TF footprint" analysis to better predict TF occupancy on accessible
202 sites by integrating the ATAC-seq and gene expression microarray profiles (**Figure 3**).

203 From the Jasper database we obtained 242 vertebrate TF motifs [45]. We then used
204 Genomica [27] producing a time points-specific TF occupancy network (**Figure 3B**). This analysis
205 revealed distinct patterns of TF access to DNA at different time points. Many known TFs were
206 enriched, for instance, STAT5, which is an IL-15 downstream signaling molecule and is
207 indispensable throughout the lifetime of NK cells. Deficiency of STAT5a/b has been reported to
208 result in complete elimination of NK cells, which demonstrates the important and non-redundant
209 effects of STAT5 [43]. The JAK-STAT pathway has also been shown to be an important signaling
210 pathway used by various cytokines and growth factors [46]. The interferon regulatory factor (IRF)
211 family regulates a variety of processes, including hematopoietic development, tumorigenesis, host
212 immunization and pathogen defense [47, 48]. IRF2 is required to maintain the normal
213 differentiation of NK cells in a cell-intrinsic manner [49, 50]. And IRF2-deficient NK cells showed
214 reduced levels of mature markers and IFN- γ production during stimulation [49, 50]. T-bet and

215 EOMES are members of the T-box family and are known to control different aspects of NK cell
216 differentiation and maturation. [21, 51, 52].

217 In addition, several novel TFs were also enriched, namely, FOSL2 and EGR2, and are
218 potential regulators of NK cell differentiation. These TFs were assessed along with the other
219 enriched TF families from the motif analysis to identify which gene in the family was expressed
220 or differentially expressed during NK differentiation to further filter out true regulators. At each
221 time point, we plotted both the expression value (colored from red to green to represent high to
222 low expression in the gene microarray profile) and the motif enrichment score (shown by the circle
223 size representing the $-\log(\text{P-value})$ of the enrichment) in the same figure (**Figure 3C**), and we
224 observed that the known regulators were highly expressed and enriched at different stages, same
225 as the genes *FOSL2* and *EGR2*. By integrating results from the above three analyze (**Figure 3A-**
226 **C**), we predict FOSL2 and EGR2 as potential regulators.

227 DNA sequences that are directly occupied by DNA-binding proteins are protected from
228 transposition, and therefore the resulting sequence “footprint” reveals the presence of a DNA-
229 binding protein at each site, analogous to DNase digestion footprints. TF “footprint” analysis of
230 our ATAC-seq profile provided further evidence of direct occupancy of a TF candidate on genomic
231 DNA and thus refined the prediction of potential regulators. We illustrated the “footprints” of 2
232 known regulators, STAT5 and T-bet, and observed higher DNA accessibility and deeper
233 “footprints” flanking their motifs in the late compared with the early stage during NK cell
234 differentiation (**Figure 3D**). Similarly, “footprints” of the TFs FOSL2 and EGR2 were also deeper
235 and more accessible at the late stage, suggesting not only that the motifs of these 2 TFs were
236 enriched at stage-specific peaks but that they were most likely physically bound to those accessible
237 chromatin sites, indicating that they are functional regulators of NK differentiation (**Figure 3D**).
238 Overall, the results from footprint analysis were also consistent with that from the HOMER and
239 Genomica’s motif enrichment analysis.

240 By combining the TF motif and “footprint” analysis, we can theoretically predict genes
241 that are regulated by any TF of interest. We performed this analysis on the TFs RUNX, *T-bet*,
242 FOSL2 and EGR2 and integrated the gene expression profiles at each time point (**Figure S5B**,
243 **Table S2**). We found that genes regulated by each of these TFs were also highly expressed at late
244 stages and were significantly enriched in GO terms of immune system construction and other
245 related functions. Transcription factors ETS1 was reported to drive early stages of NK cell

246 differentiation [53], and we found that there was a FOSL2 binding site in a dynamical accessible
247 sites on the gene body of *ETS1*, suggesting FOSL2 might regulate NK cell differentiation through
248 *ETS1* (**Figure 3E**). Transcription factors GATA3 was found regulating liver-specific NK cells,
249 IFN- γ production, and T-bet expression in mice [54]. Similarly, we found that both FOSL2 and
250 EGR2 bind to the gene body of *GATA3* (**Figure 3E**), suggesting that the TFs FOSL2 and EGR2
251 may also regulate NK cell differentiation through regulation of *GATA3*.

252

253 **FOSL2 and EGR2 are necessary for NK cell differentiation**

254

255 *Fosl2* belongs to the *Fos* gene family, which encodes leucine zipper proteins that combines
256 to the TF complex AP-1 in the form of a dimer with JUN family. Thus, FOS proteins have been
257 suggested as key regulators of transformation, differentiation and cell proliferation. *FOSL2* (FOS-
258 like 2) is a protein-coding gene. The GO annotations of *FOSL2* include sequence-specific DNA
259 binding, transcription factor activity and RNA polymerase II specific DNA binding. A previous
260 report has shown that *FOSL2* is constitutively expressed in adult T-cell leukemia (ATL) and up-
261 regulates *CCR4* and promotes ATL cell growth, together with *JUND* [55]. *EGR2* (early growth
262 response 2) is also a protein-coding gene that contains three tandem C2H2-type zinc fingers. GO
263 annotations of this gene include ligase activity, sequence-specific DNA binding and transcription
264 factor activity. Previous reports have shown that *Egr2* regulate T-cell and B-cell function in
265 homeostasis and adaptive immune responses by controlling inflammation and promoting antigen
266 receptor signaling [56, 57].

267 Since their regulatory functions in NK cell differentiation have not been well characterized,
268 we performed loss-of-function experiments to assess the effects of *FOSL2* and *EGR2* on NK cell
269 differentiation. We first validated their expression levels with real-time PCR, and we found that
270 their expression gradually increased in later stages of NK cell differentiation, consistent with our
271 bioinformatics exploration (**Figure S6A**). Subsequently, we infected the cultured cells with TF-
272 specific shRNA- and control shRNA-containing lentivirus, respectively, which are represented by
273 GFP expression (**Figure 4A**). We then sorted the GFP-positive cells and observed dramatically
274 reduced FOSL2 and EGR2 expression levels (**Figure 4B**), suggesting positive targeting of the TF-
275 specific shRNA. During NK cell differentiation, we observed a nearly 30% reduction in the
276 proportion of differentiated NK cells in GFP-positive cells at days 28 and 35 (**Figure 4C-D**).

277 However, in the non-transfected (GFP-negative) cells, the proportion of differentiated NK cells
278 was not affected at the same time (**Figure 4E-F**). These results indicate that knockdown of *FOSL2*
279 and *EGR2* expression, but not viral infection, inhibit NK cell differentiation, suggesting that
280 *FOSL2* and *EGR2* are necessary for NK cell differentiation. We then tested another important
281 marker of NK cell maturation CD11b [1], and found that it was significantly reduced in GFP-
282 positive NK cells, suggesting that *FOSL2* and *EGR2* might affect NK cell maturation (**Figure 4G**).
283 Overall, we predicted these two TFs *FOSL2* and *EGR2* as key regulators based on ATAC-seq and
284 gene microarray profile analysis, and then experimentally verified that they were indeed necessary
285 for normal NK cell differentiation.

286 In addition, we attempted to identify the signaling pathways through which *FOSL2* and
287 *EGR2* were involved in driving NK cell differentiation by performing module map analysis of all
288 the differential peaks. 18 modules were identified according to the patterns of their accessibility
289 (**Table S3**), several of which were significantly enriched with known TFs that regulate NK cell
290 differentiation (**Figure S6B**), including *FOSL2* and *EGR2*, and many genes in the JAK-STAT
291 pathway, such as the STAT family, *IRF2*, *TBX21*, and *GATA3*. These results suggest that *FOSL2*
292 and *EGR2*, may regulate NK cell differentiation through the JAK-STAT pathway (**Figure S6C**).

293

294 **Transcriptional regulatory network dynamics during NK cell differentiation**

295

296 The dramatic chromatin accessibility differences during NK cell differentiation prompted
297 us to check the time point-specific transcriptional regulatory network. Although some TFs have
298 been discovered to regulate NK cell differentiation, little is known about the dynamics of the entire
299 regulatory network during this process. Since the TF footprint pattern from the ATAC-seq reads
300 can simultaneously directly reveal the binding profiles of hundreds of TFs with known cognate
301 motifs [22], together with gene expression profiles, we can construct a regulatory network at each
302 time point and assess how it changes during NK cell differentiation. First, we used HOMER to
303 identify enriched TFs that bound to the cluster I and cluster II peaks shown in **Figure 3A** ($P < 0.05$).
304 We then examined the gene expression profiles of these TFs and found that 120 TFs were
305 expressed in at least one stage during the differentiated process. By applying differential analysis,
306 we obtained 14-26 TFs that were distinctly expressed at each stage (fold change > 1.5) (**Figure S7**),
307 and defined them as the nodes of the regulatory network. The connections (edges) between any 2

308 TFs were defined as follows: If the TF A motif is on the promoter of TF B, then we say TF A
309 regulates TF B and draw an arrow from TF A to TF B. Here, only TFs that were expressed at the
310 specific time point were consideration [58]. Using this method, we constructed the transcriptional
311 regulatory network at each time points with both the enrichment (P-value) and expression
312 information for all the relevant TFs (**Figure 5A-E**). Interestingly, day 7 specific TFs were densely
313 interconnected at the beginning, and quickly vanished after two weeks (**Figure 5A**). In contrast,
314 the day 35 specific network gradually grew out through the induction of relevant TFs. Many known
315 regulators, such as EOMES, TBX21, ETS1, PRDM1, and GATA3 and also FOLS2 and EGR2
316 were increasingly enriched on the network (**Figure 5E**). Similar phenomena were also observed
317 on other networks (**Figure 5B-D**). We believe the dynamics of the transcriptional regulatory
318 network explain the increase in the proportion and the differentiation of NK cells.

319

320 **Discussion**

321

322 NK cells are important innate immune cells that have been recognized as efficient effector
323 cells to treat tumors. To better understand the differentiation machinery of NK cells and identify
324 new regulators in the process, we studied the landscape of active elements during NK cell
325 differentiation using the sensitive ATAC-seq method. As shown by the ATAC-seq results, three
326 distinct clusters of DNA accessible elements were found. Further analysis showed that the
327 chromatin accessibility correlated well with the level of the expression of the corresponding genes.
328 In short, this study provides an epigenomic landscape and dynamics of NK cell differentiation and
329 presents foundational profiles for studying the relationship between chromatin accessibility, gene
330 expression and cell growth during this process (**Figure 2**).

331 TFs bind to their motifs and are often obligate nucleosome evictors and the creators of
332 accessible DNA sites, and therefore we can use ATAC-seq to predict critical regulators in NK cell
333 differentiation [26]. By motif analysis in HOMER, we found that several known TFs were enriched
334 at different stages during NK cell differentiation. Similar results were also obtained using
335 Genomica (**Figure 3B**). The discovery of known regulators strongly suggested the reliability of
336 our analysis Furthermore, by integrating results from HOMER, Genomica and motif footprint
337 analysis, we identified two novel TFs, FOSL2 and EGR2 that were essential for NK cell maturation.
338 Knockdown of either of these two TFs significantly inhibited NK cell transformation in the *in vitro*

339 NK cell differentiation system. Module map analysis suggested these two TFs may regulate NK
340 cell through the JAK-STAT pathway, and therefore further studies of this pathway may facilitate
341 the generation of NK cells and thus promote the NK cell-based immunotherapy. Overall, this study
342 also provided a framework to identify new regulators from chromatin accessible data for NK cell
343 differentiation.

344 TFs do not usually function alone, they always interact with other molecules to fulfill their
345 unique roles. Hence, we depicted the transcriptional regulatory networks at different stages during
346 NK cell differentiation. In order to construct a stable transcriptional regulatory network, we
347 performed a rigorous screening of TFs to avoid stochastic fluctuations, and integrated both the
348 enrichment (P-value) and expression information for all the relevant TFs. Therefore, even a change
349 of either the enrichment or gene expression cutoff may result in different networks, the most
350 critical TFs to the regulatory process still remain. However, since the screening of TFs mainly rely
351 on the gene microarray, which is not as accurate as RNA-seq, the structure of these predicted
352 network may not as robust as well.

353 From our previous study, we noticed that with a minimal cytokine cocktail, we can generate
354 sufficient number of functional NK cells that express the cytokines necessary for NK cells and
355 have a high effect on tumor cells [19]. Although there may also be a certain proportion of other
356 lymphocytes, *in vitro* expansion of NK cells from peripheral blood (PB) or UCB cells has been
357 successfully performed and developed in several clinical strategies to treat cancer [4, 5, 59-61].
358 Therefore, a comprehensive understanding of the regulatory patterns at each differentiated stage
359 of the *in vitro*-derived NK cells in this system will help to uncover the underlying mechanisms of
360 NK cell differentiation. The transcriptional regulatory network revealed in this study will lay the
361 foundation for faster and better *in vitro* production of effective NK cells, thus facilitating NK cell-
362 based immunotherapy. Moreover, we have identified two novel TFs, FOSL2 and EGR2, as
363 essential regulators in controlling NK cell maturation and function. We also predicted the potential
364 signaling pathways in which these two TFs may be involved and illustrated the dynamics of the
365 transcriptional regulatory network during NK cell differentiation. In spite of the advantages of our
366 strategy, there are two main limitations of this study: (1) Although induced NK cells are very
367 similar to those produced *in vivo*, these two types of NK cells are still not identical. We observed
368 certain differences between the induced NK cells and the NK cells produced *in vivo* in terms of
369 chromatin accessibility (data not shown). (2) Before NK cells are fully developed, there is always

370 a mixture of cell populations with stem cells, NK progenitor cells, immature NK cells and mature
371 NK cells and other cell types that we are incapable to delineate at this moment, i.e. any bulk cell
372 based analysis may neglect the huge heterogeneity between cells by default. Therefore, to fully
373 uncover the regulatory mechanism, single cell technologies are required in the future to further
374 delineate the cell-to-cell heterogeneity and regulatory dynamics at single cell precision.

375

376 **Materials and Methods**

377

378 **Samples:** Umbilical cord blood (UCB) samples were collected at birth from women with normal,
379 full-term deliveries at Anhui Provincial Hospital, Hefei, after receiving informed consent.
380 Approval was obtained from the Ethics Committee of the University of Science and Technology
381 of China. The culture procedure for NK cell differentiation from UCB CD34⁺ cells has been
382 previously described [19].

383 **Immunofluorescence staining:** The cultured cells were post-fixed in 4% paraformaldehyde,
384 blocked with 10% goat serum and stained with primary antibodies at 4°C overnight. Secondary
385 antibodies were stained at 37°C for 1 h. Cell nuclei were stained with DAPI for 5 min at room
386 temperature. Confocal images were acquired using a Zeiss LS710 microscope.

387 **RNA isolation and real-time PCR:** Cultured cells were lysed in TRIzol reagent (Invitrogen), and
388 total RNA was extracted following the manufacturer's instructions. cDNA was synthesized with
389 Moloney murine leukemia virus reverse transcriptase (Invitrogen) and oligo (dT) 20 primers. Then,
390 SYBR Premix Ex Taq (TaKaRa) was used for real-time PCR (primer sequences, **Table S4**). The
391 data were analyzed using the 2- $\Delta\Delta C_t$ method.

392 **Lentivirus production and transduction:** To produce lentiviral particles, 293T cells were
393 transfected with the plasmids PLKO.1, pRRE, pREV and pVSV-G via Lipofectamine 2000
394 (Invitrogen) following the manufacturer's protocol. Then, we harvested the supernatants 48 and
395 72 h post-transfection. To remove cell debris, supernatants were centrifuged (3000 rpm, 10 min),
396 and then, the lentivirus particles were concentrated by ultracentrifugation at 50000 g for 2 h at 4°C.
397 Finally, the virus particles were gently resuspended in HBSS and stored at -80°C. After UCB
398 CD34⁺ cells were cultured with multiple cytokines for 14-18 d, we incubated the lentivirus and the
399 cultured cells with polybrene (5 μ g/ml) and centrifuged them at 1000 rpm for 70 min at 10°C.

400 **Statistical analysis:** Statistical significance was analyzed using unpaired two-tailed t-tests. P-
401 values less than 0.05 were considered statistically significant.

402 **Microarray:** The microarray was performed using an Affymetrix GeneChip® PrimeView Human
403 Gene Expression Array. The signal values of the samples were normalized using RMA.

404 Differential analysis was performed using Student's t-test as previously described [19]. The
405 microarray data were under accession number GEO: GSE87787.

406 **ATAC-seq:** ATAC-seq was performed as previously described [22], and 2x150 paired-end
407 sequencing was performed on an Illumina HiSeq X-10 to yield, on average, 78 M reads/sample.

408 **ATAC-seq analysis:** Sample reads from biological replicates were then grouped together and
409 divided into 8 categories: day 7, day 14, day19, day 21, day 24, day 26, day 28 and day 35. Intrinsic
410 analysis and other ATAC-seq analysis was performed same as our previous work [26, 62].

411 **Difference analysis:** All samples were grouped into 8 categories (16 samples): day 7, 14, 19, 21,
412 24, 26, 28 and day 35. Data normalization and significance analysis were performed via pairwise
413 comparison between the 8 categories using DESeq2 [63], with a $P < 0.01$, \log_2 -fold change > 5 ,
414 $FDR < 0.01$, and intrinsic analysis [26] and with a z-score > 1 . We finally obtained 6401 differential
415 peaks. Unsupervised clustering was performed using Cluster 3.0 and visualized in Treeview.
416 GREAT [33] was used to predict enriched functions and Gene Ontology.

417 **Stage-specific peaks:** Each stage (e.g., day 7) consists of genes that were induced (> 1.5 -fold
418 change) at that stage compared with all the other stages. We defined peaks that were accessible
419 only in one stage as stage-specific peaks, and those that were accessible at all stages as
420 conservative peaks.

421 **Define promoter:** We define the range of 2kb around the transcription start site as a promoter.

422 **TF motif enrichment analysis:** HOMER [29] was used to perform the TF enrichment analysis
423 with the following options: findmotifs.pl input.fa fasta output/ for **Figure 3A** and findmotifs.pl
424 input.fa human uotputdir -fasta bg.fa for **Figure S5A**. TF enrichment analysis using Genomica
425 and TF footprinting analysis was performed the same as our previous work [26].

426 **Gene Module Analysis:** Gene module analyses were performed using WGCNA [64] in **Figure**
427 **S6B** with the options SoftPower = 20, minModuleSize = 30.

428 **STRING analysis:** Protein-protein interaction analyses were performed using STRING [65] in
429 **Figure S6C**.

430 **Transcriptional regulatory network:** We used HOMER to find the transcription factors that bind
431 to cluster I and cluster II and obtained a total of 253 transcription factors that could regulate these
432 differentially expressed elements ($P < 0.05$), of which 120 were expressed in the microarray profile
433 of the corresponding sample. If TF A bound to the promoter of TF B, we defined TF A as a
434 regulator of TF B and then constructed a transcriptional regulatory network. Networks of TFs were
435 assembled from this set of 120 TFs that were expressed in at least one sample. An edge between
436 TF A and TF B indicated that TF A bound to the promoter of TF B [58].

437

438 **Data Availability**

439

440 **GSA accession:** The raw sequence data have been deposited in the Genome Sequence Archive
441 [66] in BIG Data Center [67], under the accession number CRA000846 and are publicly accessible
442 at <http://bigd.big.ac.cn/gsa/s/u7fdeNV3>.

443

444 **Acknowledgements**

445

446 This work was supported by the National Natural Science Foundation of China grant 81788101
447 (to Z.T.), the National Key R&D Program of China 2017YFA0102903 (to K.Q.), the National
448 Natural Science Foundation of China grant 91640113 (to K.Q.) and 31771428 (to K.Q.), the key
449 project of Natural Science Foundation of China 81330071 (to H.W.). We thank the USTC
450 supercomputing center and the School of Life Science Bioinformatics Center for providing
451 supercomputing resources for this project.

452

453 **Authors' contributions**

454

455 KQ and HW conceived the project, YW and YL performed all the cell sorting and ATAC-seq
456 library experiments, KL performed most of the data analysis, QY analyzed the microarray data,
457 HW and ZT discussed the results and provided advices, KL and KQ wrote the manuscript.

458

459 **Conflict interests**

460

461 All authors have no competing financial interests.

462

463 **References**

464

465 [1] Sun JC. Transcriptional Control of NK Cells. *Curr Top Microbiol Immunol* 2016;395:1-36.

466 [2] Miller JS, Soignier Y, Panoskaltsis-Mortari A, McNearney SA, Yun GH, Fautsch SK, et al.
467 Successful adoptive transfer and in vivo expansion of human haploidentical NK cells in patients
468 with cancer. *Blood* 2005;105:3051-7.

469 [3] Miller JS, Verfaillie C, McGlave P. The generation of human natural killer cells from
470 CD34+/DR- primitive progenitors in long-term bone marrow culture. *Blood* 1992;80:2182-7.

471 [4] Luevano M, Madrigal A, Saudemont A. Generation of natural killer cells from hematopoietic
472 stem cells in vitro for immunotherapy. *Cell Mol Immunol* 2012;9:310-20.

473 [5] Cheng M, Chen Y, Xiao W, Sun R, Tian Z. NK cell-based immunotherapy for malignant
474 diseases. *Cell Mol Immunol* 2013;10:230-52.

475 [6] Dolstra H, Roeven MWH, Spanholtz J, Hangalapura BN, Tordoir M, Maas F, et al. Successful
476 Transfer of Umbilical Cord Blood CD34(+) Hematopoietic Stem and Progenitor-derived NK Cells
477 in Older Acute Myeloid Leukemia Patients. *Clin Cancer Res* 2017;23:4107-18.

478 [7] Ruggeri L, Capanni M, Urbani E, Perruccio K, Shlomchik WD, Tosti A, et al. Effectiveness of
479 donor natural killer cell alloreactivity in mismatched hematopoietic transplants. *Science*
480 2002;295:2097-100.

481 [8] Guillerey C, Huntington ND, Smyth MJ. Targeting natural killer cells in cancer
482 immunotherapy. *Nat Immunol* 2016;17:1025-36.

483 [9] Sivori S, Cantoni C, Parolini S, Marcenaro E, Conte R, Moretta L, et al. IL-21 induces both
484 rapid maturation of human CD34+ cell precursors towards NK cells and acquisition of surface
485 killer Ig-like receptors. *Eur J Immunol* 2003;33:3439-47.

486 [10] Spanholtz J, Tordoir M, Eissens D, Preijers F, van der Meer A, Joosten I, et al. High log-scale
487 expansion of functional human natural killer cells from umbilical cord blood CD34-positive cells
488 for adoptive cancer immunotherapy. *PLoS One* 2010;5:e9221.

489 [11] Kao IT, Yao CL, Kong ZL, Wu ML, Chuang TL, Hwang SM. Generation of natural killer
490 cells from serum-free, expanded human umbilical cord blood CD34+ cells. *Stem Cells Dev*
491 2007;16:1043-51.

492 [12] Haraguchi K, Suzuki T, Koyama N, Kumano K, Nakahara F, Matsumoto A, et al. Notch
493 activation induces the generation of functional NK cells from human cord blood CD34-positive
494 cells devoid of IL-15. *J Immunol* 2009;182:6168-78.

495 [13] Yoon SR, Lee YS, Yang SH, Ahn KH, Lee JH, Lee JH, et al. Generation of donor natural
496 killer cells from CD34(+) progenitor cells and subsequent infusion after HLA-mismatched
497 allogeneic hematopoietic cell transplantation: a feasibility study. *Bone Marrow Transplant*
498 2010;45:1038-46.

499 [14] Shah N, Martin-Antonio B, Yang H, Ku S, Lee DA, Cooper LJ, et al. Antigen presenting cell-
500 mediated expansion of human umbilical cord blood yields log-scale expansion of natural killer
501 cells with anti-myeloma activity. *PLoS One* 2013;8:e76781.

502 [15] Vitale C, Cottalasso F, Montaldo E, Moretta L, Mingari MC. Methylprednisolone induces
503 preferential and rapid differentiation of CD34+ cord blood precursors toward NK cells. *Int*
504 *Immunol* 2008;20:565-75.

- 505 [16] Carayol G, Robin C, Bourhis JH, Bennaceur-Griscelli A, Chouaib S, Coulombel L, et al. NK
506 cells differentiated from bone marrow, cord blood and peripheral blood stem cells exhibit similar
507 phenotype and functions. *Eur J Immunol* 1998;28:1991-2002.
- 508 [17] Yu Y, Hagihara M, Ando K, Gansuud B, Matsuzawa H, Tsuchiya T, et al. Enhancement of
509 human cord blood CD34+ cell-derived NK cell cytotoxicity by dendritic cells. *J Immunol*
510 2001;166:1590-600.
- 511 [18] Cany J, van der Waart AB, Spanholtz J, Tordoir M, Jansen JH, van der Voort R, et al.
512 Combined IL-15 and IL-12 drives the generation of CD34(+)-derived natural killer cells with
513 superior maturation and alloreactivity potential following adoptive transfer. *Oncoimmunology*
514 2015;4:e1017701.
- 515 [19] Wu Y, Li Y, Fu B, Jin L, Zheng X, Zhang A, et al. Programmed differentiated natural killer
516 cells kill leukemia cells by engaging SLAM family receptors. *Oncotarget* 2017;8:57024-38.
- 517 [20] Colucci F, Samson SI, DeKoter RP, Lantz O, Singh H, Di Santo JP. Differential requirement
518 for the transcription factor PU.1 in the generation of natural killer cells versus B and T cells. *Blood*
519 2001;97:2625-32.
- 520 [21] Gordon SM, Chaix J, Rupp LJ, Wu J, Madera S, Sun JC, et al. The transcription factors T-bet
521 and Eomes control key checkpoints of natural killer cell maturation. *Immunity* 2012;36:55-67.
- 522 [22] Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native
523 chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins
524 and nucleosome position. *Nat Methods* 2013;10:1213-8.
- 525 [23] Sen DR, Kaminski J, Barnitz RA, Kurachi M, Gerdemann U, Yates KB, et al. The epigenetic
526 landscape of T cell exhaustion. *Science* 2016;354:1165-9.
- 527 [24] Pauken KE, Sammons MA, Odorizzi PM, Manne S, Godec J, Khan O, et al. Epigenetic
528 stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. *Science*
529 2016;354:1160-5.
- 530 [25] Denny SK, Yang D, Chuang CH, Brady JJ, Lim JS, Gruner BM, et al. Nfib Promotes
531 Metastasis through a Widespread Increase in Chromatin Accessibility. *Cell* 2016;166:328-42.
- 532 [26] Qu K, Zaba LC, Giresi PG, Li R, Longmire M, Kim YH, et al. Individuality and variation of
533 personal regulomes in primary human T cells. *Cell Syst* 2015;1:51-61.
- 534 [27] Qu K, Zaba LC, Satpathy AT, Giresi PG, Li R, Jin Y, et al. Chromatin Accessibility
535 Landscape of Cutaneous T Cell Lymphoma and Dynamic Response to HDAC Inhibitors. *Cancer*
536 *Cell* 2017;32:27-41 e4.
- 537 [28] Lau CM, Adams NM, Geary CD, Weizman OE, Rapp M, Pritykin Y, et al. Epigenetic control
538 of innate and adaptive immune memory. *Nat Immunol* 2018;19:963-72.
- 539 [29] Heinz E, Born D, Zieger G, May T, Krause T, Kruger A, et al. Progress report on Safe
540 VISITOR: approaching a practical instrument for terahertz security screening. *Passive Millimeter-*
541 *Wave Imaging Technology Xiii* 2010;7670.
- 542 [30] Yu J, Freud AG, Caligiuri MA. Location and cellular stages of natural killer cell development.
543 *Trends Immunol* 2013;34:573-82.
- 544 [31] Montaldo E, Del Zotto G, Della Chiesa M, Mingari MC, Moretta A, De Maria A, et al. Human
545 NK cell receptors/markers: a tool to analyze NK cell development, subsets and function.
546 *Cytometry A* 2013;83:702-13.
- 547 [32] Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits
548 of human breast tumours. *Nature* 2000;406:747-52.
- 549 [33] McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves
550 functional interpretation of cis-regulatory regions. *Nature Biotechnology* 2010;28:495-U155.

- 551 [34] Grewal SI, Moazed D. Heterochromatin and epigenetic control of gene expression. *Science*
552 2003;301:798-802.
- 553 [35] Ohno S, Sato T, Kohu K, Takeda K, Okumura K, Satake M, et al. Runx proteins are involved
554 in regulation of CD122, Ly49 family and IFN-gamma expression during NK cell differentiation.
555 *Int Immunol* 2008;20:71-9.
- 556 [36] Yamanaka R, Lekstrom-Himes J, Barlow C, Wynshaw-Boris A, Xanthopoulos KG.
557 CCAAT/enhancer binding proteins are critical components of the transcriptional regulation of
558 hematopoiesis (Review). *Int J Mol Med* 1998;1:213-21.
- 559 [37] Chang HC, Zhang S, Thieu VT, Slee RB, Bruns HA, Laribee RN, et al. PU.1 expression
560 delineates heterogeneity in primary Th2 cells. *Immunity* 2005;22:693-703.
- 561 [38] Dakic A, Metcalf D, Di Rago L, Mifsud S, Wu L, Nutt SL. PU.1 regulates the commitment
562 of adult hematopoietic progenitors and restricts granulopoiesis. *J Exp Med* 2005;201:1487-502.
- 563 [39] Nutt SL, Metcalf D, D'Amico A, Polli M, Wu L. Dynamic regulation of PU.1 expression in
564 multipotent hematopoietic progenitors. *J Exp Med* 2005;201:221-31.
- 565 [40] Carotta S, Dakic A, D'Amico A, Pang SH, Greig KT, Nutt SL, et al. The transcription factor
566 PU.1 controls dendritic cell development and Flt3 cytokine receptor expression in a dose-
567 dependent manner. *Immunity* 2010;32:628-41.
- 568 [41] Iwasaki H, Somoza C, Shigematsu H, Duprez EA, Iwasaki-Arai J, Mizuno S, et al. Distinctive
569 and indispensable roles of PU.1 in maintenance of hematopoietic stem cells and their
570 differentiation. *Blood* 2005;106:1590-600.
- 571 [42] Boos MD, Yokota Y, Eberl G, Kee BL. Mature natural killer cell and lymphoid tissue-
572 inducing cell development requires Id2-mediated suppression of E protein activity. *Journal of*
573 *Experimental Medicine* 2007;204:1119-30.
- 574 [43] Eckelhart E, Warsch W, Zebedin E, Simma O, Stoiber D, Kolbe T, et al. A novel Ncr1-Cre
575 mouse reveals the essential role of STAT5 for NK-cell survival and development. *Blood*
576 2011;117:1565-73.
- 577 [44] Collins A, Littman DR, Taniuchi I. RUNX proteins in transcription factor networks that
578 regulate T-cell lineage choice. *Nat Rev Immunol* 2009;9:106-15.
- 579 [45] Mathelier A, Fornes O, Arenillas DJ, Chen CY, Denay G, Lee J, et al. JASPAR 2016: a major
580 expansion and update of the open-access database of transcription factor binding profiles. *Nucleic*
581 *Acids Res* 2016;44:D110-5.
- 582 [46] O'Shea JJ, Schwartz DM, Villarino AV, Gadina M, McInnes IB, Laurence A. The JAK-STAT
583 pathway: impact on human disease and therapeutic intervention. *Annu Rev Med* 2015;66:311-28.
- 584 [47] Tamura T, Yanai H, Savitsky D, Taniguchi T. The IRF family transcription factors in
585 immunity and oncogenesis. *Annu Rev Immunol* 2008;26:535-84.
- 586 [48] Taniguchi T, Ogasawara K, Takaoka A, Tanaka N. IRF family of transcription factors as
587 regulators of host defense. *Annual Review of Immunology* 2001;19:623-55.
- 588 [49] Lohoff M, Duncan GS, Ferrick D, Mittrucker HW, Bischof S, Prechtel S, et al. Deficiency in
589 the transcription factor interferon regulatory factor (IRF)-2 leads to severely compromised
590 development of natural killer and T helper type 1 cells. *J Exp Med* 2000;192:325-36.
- 591 [50] Taki S, Nakajima S, Ichikawa E, Saito T, Hida S. IFN Regulatory Factor-2 Deficiency
592 Revealed a Novel Checkpoint Critical for the Generation of Peripheral NK Cells. *The Journal of*
593 *Immunology* 2005;174:6005-12.
- 594 [51] Szabo SJ, Kim ST, Costa GL, Zhang X, Fathman CG, Glimcher LH. A novel transcription
595 factor, T-bet, directs Th1 lineage commitment. *Cell* 2000;100:655-69.

- 596 [52] Townsend MJ, Weinmann AS, Matsuda JL, Salomon R, Farnham PJ, Biron CA, et al. T-bet
597 regulates the terminal maturation and homeostasis of NK and V alpha 14i NKT cells. *Immunity*
598 2004;20:477-94.
- 599 [53] Ramirez K, Chandler KJ, Spaulding C, Zandi S, Sigvardsson M, Graves BJ, et al. Gene
600 deregulation and chronic activation in natural killer cells deficient in the transcription factor ETS1.
601 *Immunity* 2012;36:921-32.
- 602 [54] Samson SI, Richard O, Tavian M, Ranson T, Vosshenrich CA, Colucci F, et al. GATA-3
603 promotes maturation, IFN-gamma production, and liver-specific homing of NK cells. *Immunity*
604 2003;19:701-11.
- 605 [55] Higuchi T, Nakayama T, Arao T, Nishio K, Yoshie O. SOX4 is a direct target gene of FRA-
606 2 and induces expression of HDAC8 in adult T-cell leukemia/lymphoma. *Blood* 2013;121:3640-
607 9.
- 608 [56] Li S, Miao T, Sebastian M, Bhullar P, Ghaffari E, Liu M, et al. The transcription factors Egr2
609 and Egr3 are essential for the control of inflammation and antigen-induced proliferation of B and
610 T cells. *Immunity* 2012;37:685-96.
- 611 [57] Okamura T, Fujio K, Sumitomo S, Yamamoto K. Roles of LAG3 and EGR2 in regulatory T
612 cells. *Ann Rheum Dis* 2012;71 Suppl 2:i96-100.
- 613 [58] Novershtern N, Subramanian A, Lawton LN, Mak RH, Haining WN, McConkey ME, et al.
614 Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell*
615 2011;144:296-309.
- 616 [59] Suck G, Koh MB. Emerging natural killer cell immunotherapies: large-scale ex vivo
617 production of highly potent anticancer effectors. *Hematol Oncol Stem Cell Ther* 2010;3:135-42.
- 618 [60] Smyth MJ, Hayakawa Y, Takeda K, Yagita H. New aspects of natural-killer-cell surveillance
619 and therapy of cancer. *Nat Rev Cancer* 2002;2:850-61.
- 620 [61] Fujisaki H, Kakuda H, Shimasaki N, Imai C, Ma J, Lockey T, et al. Expansion of highly
621 cytotoxic human natural killer cells for cancer cell therapy. *Cancer Res* 2009;69:4010-7.
- 622 [62] Zuo Z, Jin Y, Zhang W, Lu Y, Li B, Qu K. ATAC-pipe: general analysis of genome-wide
623 chromatin accessibility. *Brief Bioinform* 2018.
- 624 [63] Anders S, Huber W. Differential expression analysis for sequence count data. *Genome*
625 *Biology* 2010;11.
- 626 [64] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis.
627 *BMC Bioinformatics* 2008;9:559.
- 628 [65] Snel B, Lehmann G, Bork P, Huynen MA. STRING: a web-server to retrieve and display the
629 repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res* 2000;28:3442-4.
- 630 [66] Wang Y, Song F, Zhu J, Zhang S, Yang Y, Chen T, et al. GSA: Genome Sequence Archive.
631 *Genomics Proteomics Bioinformatics* 2017;15:14-8.
- 632 [67] Members BIGDC. Database Resources of the BIG Data Center in 2018. *Nucleic Acids Res*
633 2018;46:D14-D20.
- 634

635 **Figures legends**

636

637 **Figure 1. Landscape of DNA accessibility during NK cell differentiation**

638 **A** Confocal microscopy images of membrane CD56 (red) in the cultured cells at days 7, 14, 21,
639 28 and 35. Scale bars, 30 μ m. Nuclei are stained with DAPI.

640 **B** Schematic representation of the overall experimental design of this study. Gene expression and
641 chromosome opening at different time points were assessed using microarray and ATAC-seq data
642 respectively. The bioinformatics pipeline for data analysis is shown in the bottom.

643 **C** The gene expression profiles of *SPI1* and *TBX21* at different stages of NK cell differentiation.

644 **D** Flow cytometric measurement of PU.1 (*SPI1*-encoded protein) and T-bet (*TBX21*-encoded
645 protein) expression using in cultured cells during a 35-day time course.

646 **E** Normalized ATAC-seq profiles of the *SPI1* (top) and *TBX21* (bottom) gene loci at different
647 stages during NK cell differentiation. These two genes are known to regulate NK cell
648 differentiation, and ATAC-seq signals were obtained from the UCSC genome browser.

649

650 **Figure 2. Epigenomic signatures of different stages during NK cell differentiation**

651 **A** Heatmap of 6401 differentially expressed regulatory elements during NK cell differentiation.
652 Each column is a sample, and each row is a peak. Samples and peaks are organized by two-
653 dimensional unsupervised hierarchical clustering. Color scale showing the relative ATAC-seq
654 peak intensity centered by each peak summit. Top: samples at all time points were categorized into
655 three groups, early stage: days 7~21 (orange); interim stage: days 24~28 (yellow) and late stage:
656 days 28~35 (green). Samples from the same cluster are labeled with the same color. Left:
657 differential peaks are categorized into 3 clusters.

658 **B** The top ten most significant gene ontology (GO) terms enriched in cluster I and cluster II peaks.

659 **C** Box-plots of the mRNA expression levels of the genes that are more accessible in early, interim
660 and late stages during NK cell differentiation. 1,2 represents biological replicates 1 and 2,
661 respectively. P-values were estimated from Student's t-test.

662 **D** The changes in ATAC-seq signals (red), gene expression (orange) and number of NK cells
663 (green) at different time points during NK cell differentiation.

664

665 **Figure 3. Transcription factor occupancy network during NK cell differentiation**

666 **A** TF motifs enriched in cluster I (left) and cluster II (right) peaks, with enrichment P-values
667 estimated from HOMER. TFs known to regulate NK cell differentiation are colored red.

668 **B** Enrichment of known TF motifs in differential accessible elements in all samples. Each row is
669 a motif, and each column is a sample. Values in the matrix represent the significance level in terms
670 of the $-\log(\text{P-value})$ of the enrichment estimated from Genomica. Red indicates that the motif is
671 enriched in the corresponding sample, and blue indicates depletion. Red box indicates known key
672 TFs that regulates NK cell development, the pink box indicates the new TF whose function will
673 be experimentally tested later in Figure 4.

674 **C** Prediction of TFs that may regulate NK cell differentiation. TFs known to regulate NK cell
675 differentiation are shown at the top, TFs predicted to regulate early stages of the differentiation
676 process are shown in the bottom, and those that regulate the late stage are shown at the middle.
677 The color of each circle represents the expression level of the gene encoding the corresponding
678 TF, while the size of the circle represents the significance of the motifs estimated by P-values.

679 **D** Visualization of the ATAC-seq footprint for STAT5, TBET, FOSL2 and EGR2 motifs during
680 five stages of NK cell differentiation: day 7 (blue), day 14 (green), day 21 (red), day 28 (cyan),
681 day 35 (purple). ATAC-seq signals across all these motif binding sites in the genome were aligned
682 on the motif and averaged.

683 **E** Normalized ATAC-seq profiles of the *ETS1* (left) and *GATA3* (right) gene loci at different stages
684 during NK cell differentiation. These two genes are known to regulate NK cell differentiation, and
685 ATAC-seq signals were obtained from the UCSC genome browser. Bottom: The motif of FOSL2
686 and EGR2.

687

688 **Figure 4. FOSL2 and EGR2 are necessary for NK cell differentiation**

689 **A** The gating strategy for cultured cells transduced with lentiviruses expressing either shFOSL2,
690 shEGR2 or control shRNA via detection of GFP expression.

691 **B** Knockdown efficiency of *FOSL2* and *EGR2* with shFOSL2 and shEGR2.

692 **C** Flow cytometric analysis of CD56 in the shControl-, shFOSL2- and shEGR2-positive cells at
693 days 28 and 35.

694 **D** Quantification of CD56-positive cells in the GFP-positive cell population at days 28 and 35
695 (n=5).

696 **E** Flow cytometric analysis of CD56 in the GFP-negative cell population at days 28 and 35.

697 **F** Quantification of CD56-positive cells in the GFP-negative cell population at days 28 and 35
698 (n=5).

699 **G** Quantification of CD11b-positive cells in the CD56⁺ cell population at day 35 (n=5). The results
700 from five biological replicates are presented as the mean \pm SEM. *P<0.05, **P<0.001, and
701 ***P<0.0005.

702

703 **Figure 5. Transcriptional regulatory network dynamics during NK cell differentiation**

704 **A~E** Cis-regulatory networks between TFs (nodes) enriched in at least one gene set and
705 specifically expressed (fold change>1.5) at day 7 (**A**), day 14 (**B**), day 21 (**C**), day 28 (**D**) and day
706 35 (**E**). The arrow at the edge from node A to node B indicates that TF A regulates TF B by binding
707 to the promoter site of the latter. The color of each node indicates the expression value of the gene
708 encoding that TF, and the size of the circle represents the significance of the motif enrichment
709 according to the P-value. The types of edges indicate the Pearson correlation between the gene
710 expression profiles of the connected TFs: positively correlated (correlation coefficient>0.4);
711 negatively correlated (correlation coefficient<-0.4); no correlation (correlation coefficient between
712 -0.4 to 0.4).

713

714 **Supplementary Figure S1. Landscape of DNA accessibility during NK cell differentiation**

715

716 **A** Fractions of CD56⁺CD3⁻ cells in the total gated cells during a 35-day time course.

717 **B-C** Quality control analysis of ATAC-seq data. **B**: The TSS enrichment score for all samples. **C**:
718 The fragment length distribution of all the mapped reads.

719 **D** Known cell surface markers: Normalized ATAC-seq profiles of the *CD34*, *KIT*(CD117),
720 *KLRD1*(CD94) and *NCAMI*(CD56) gene loci at different stages during NK cell differentiation.

721 **E** Normalized ATAC-seq profiles of the *EOMES* (left) and *GATA3* (right) gene loci at different
722 stages during NK cell differentiation. These genes are known to regulate NK cell differentiation,
723 and ATAC-seq signals were obtained from the UCSC genome browser.

724

725 **Supplementary Figure S2. The correlation analysis on the samples**

726 **A** The Pearson correlation analysis on the replicates at each time point. R at the top is the Pearson
727 correlation coefficient.

728 B Heatmap of the Pearson correlation between all the samples with unsupervised clustering
729 performed in Cluster 3.0.

730

731 **Supplementary Figure S3. Epigenomic signatures of NK cell differentiation at different**
732 **stages**

733 A Venn diagram of peaks identified at each stage of NK cell differentiation. Specific peaks were
734 defined as peaks that were identified only at a specific time point, and conserved peaks were those
735 identified at all stages during the process.

736 B The top most significant GO terms of all the stage-specific and conservative peaks.

737

738 **Supplementary Figure S4. Epigenomic signatures of different stages during NK cell**
739 **differentiation**

740 A Heatmap of all the 6401 differential regulatory elements in all the samples. Each column is a
741 sample, and each row is a peak. Samples and peaks were organized by two-dimensional
742 unsupervised hierarchical clustering. The color scale shows the relative ATAC-seq signal
743 intensities as indicated. Top: samples at all time points were categorized into three groups, early
744 stage: days 7~21 (orange); interim stage: days 24~28 (yellow) and late stage: days 28~35 (green).
745 Samples from the same cluster are labeled with the same color. Left: differential peaks are
746 categorized into 3 clusters.

747 B Distance from all the peaks in cluster I, II, and III to their nearest genes. Known TFs regulating
748 NK cell differentiation are labeled.

749 C Principle component analysis of chromatin accessibility during NK cell differentiation. Three
750 clusters were also identified: early (days 7~21), interim (days 24 ~28), and late (day 35).

751

752 **Supplementary Figure S5. Transcription factor occupancy network during NK cell**
753 **differentiation**

754

755 A The top ten TF motifs enriched in cluster I (left) and cluster II (right) peaks, with enrichment P-
756 values estimated from HOMER. TFs known to regulate NK cell differentiation are color-coded.

757 B Left: heatmaps of the gene expression profiles of the genes predicted to be regulated by RUNX,
758 T-bet, EGR2 and FOSL2. Unsupervised hierarchical clustering was performed. Right: the top ten

759 most significant GO terms enriched in up-regulated (orange) and down-regulated (green) genes
760 predicted to be regulated by each TF on the left.

761

762 **Supplementary Figure S6. EGR2 and FOSL2 are necessary for NK cell differentiation**

763 **A** Real-time qPCR analysis (n=3) of the genes *EGR2* (left) and *FOSL2* (right). The results from
764 three replicates are presented as the mean \pm SEM.

765 **B** Module map analysis in Genomica: representative modules from module map analysis in
766 Genomica. Top: the most significantly enriched TFs in each module; bottom: the chromatin
767 accessibility changes of the peaks in the corresponding module compared with day 7. Right: genes
768 associated with peaks in that module.

769 **C** Signaling pathways of known and predicted TFs that regulate NK cell differentiation.

770

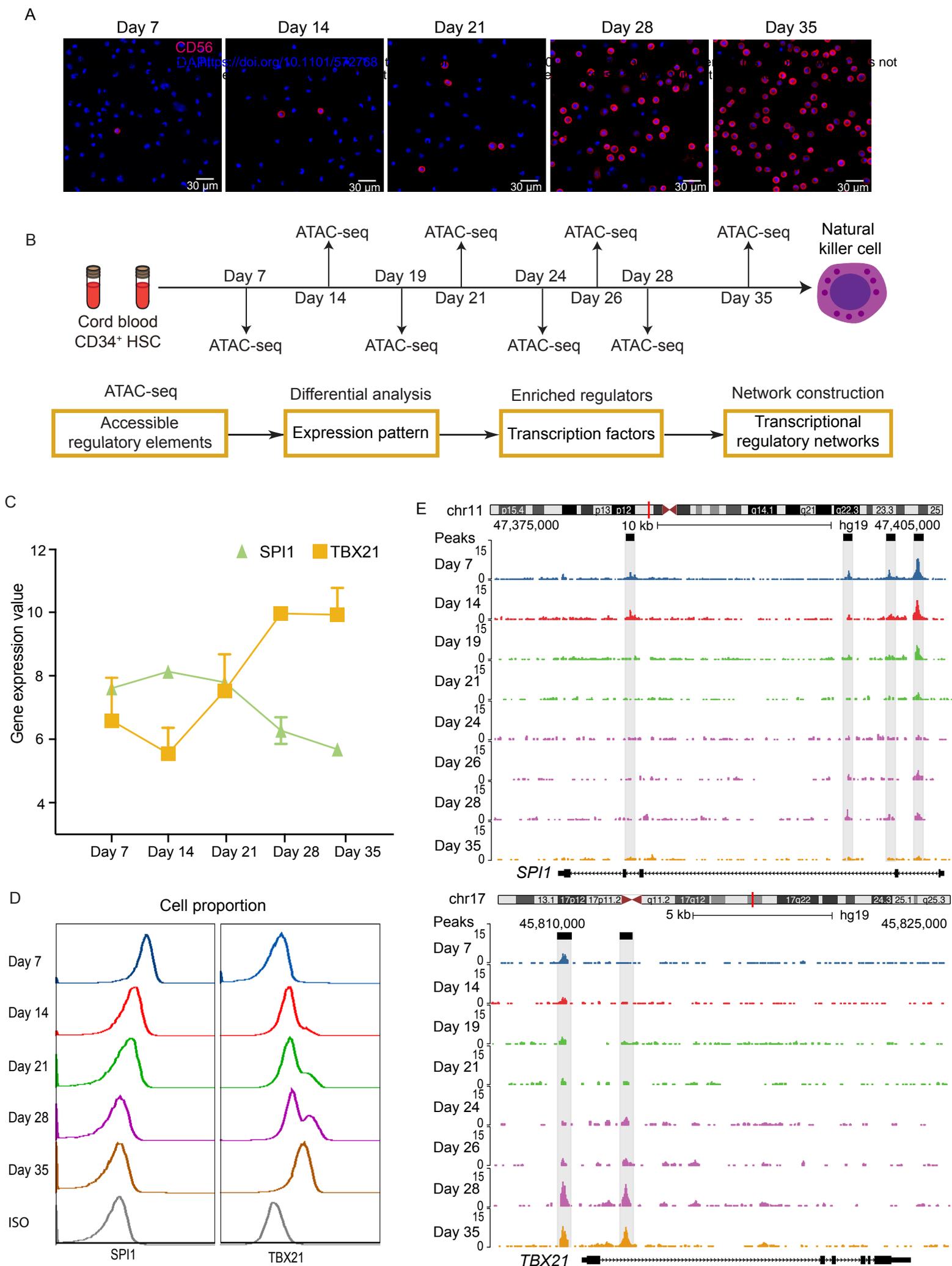
771 **Supplementary Figure S7. Define time-specific TFs based on differential expression analysis**

772 **A:** Distribution of all TF's fold change. X axis represents the fold change. y axis represents the
773 density of TF under different fold change. Std indicates Standard Deviation. Foldchange1.5 is 2
774 standard deviations away. Colors represent different time points.

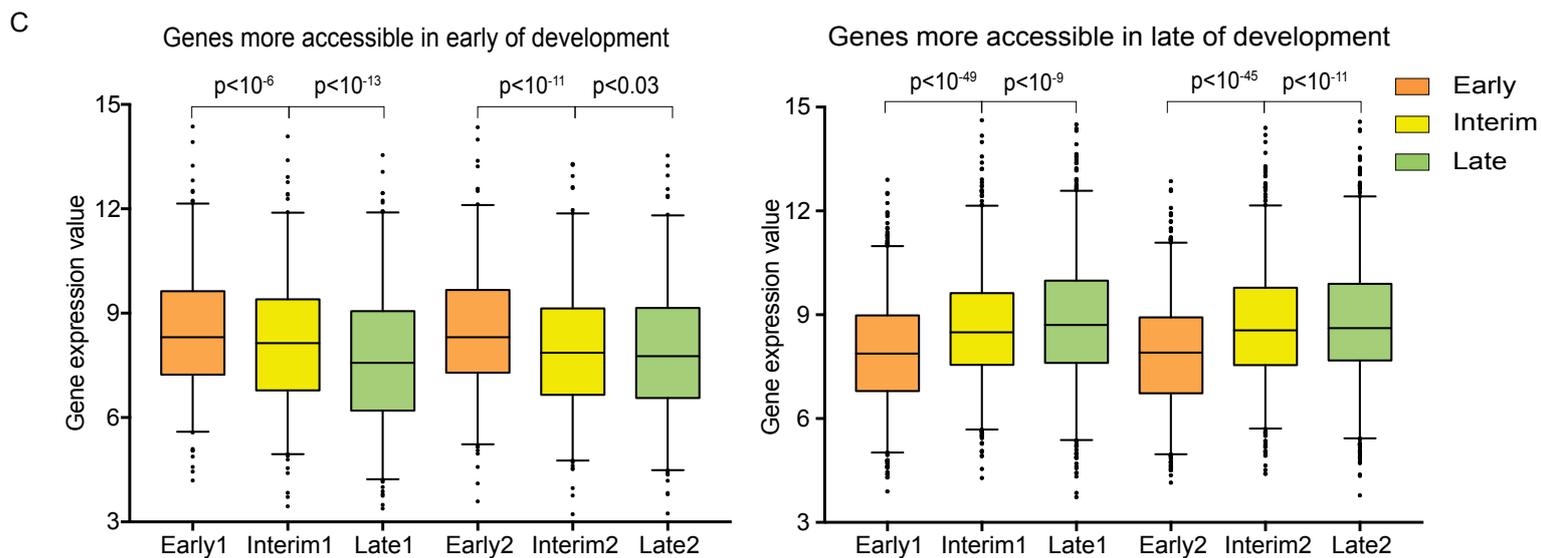
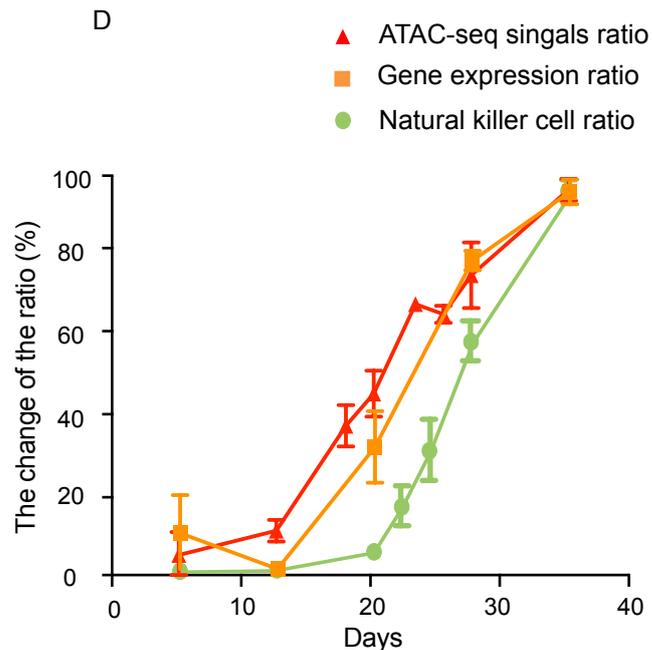
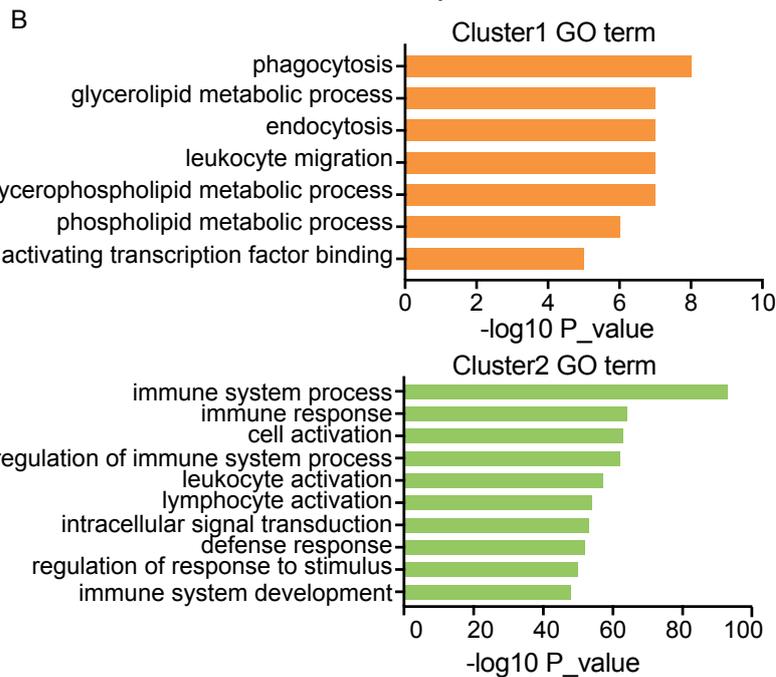
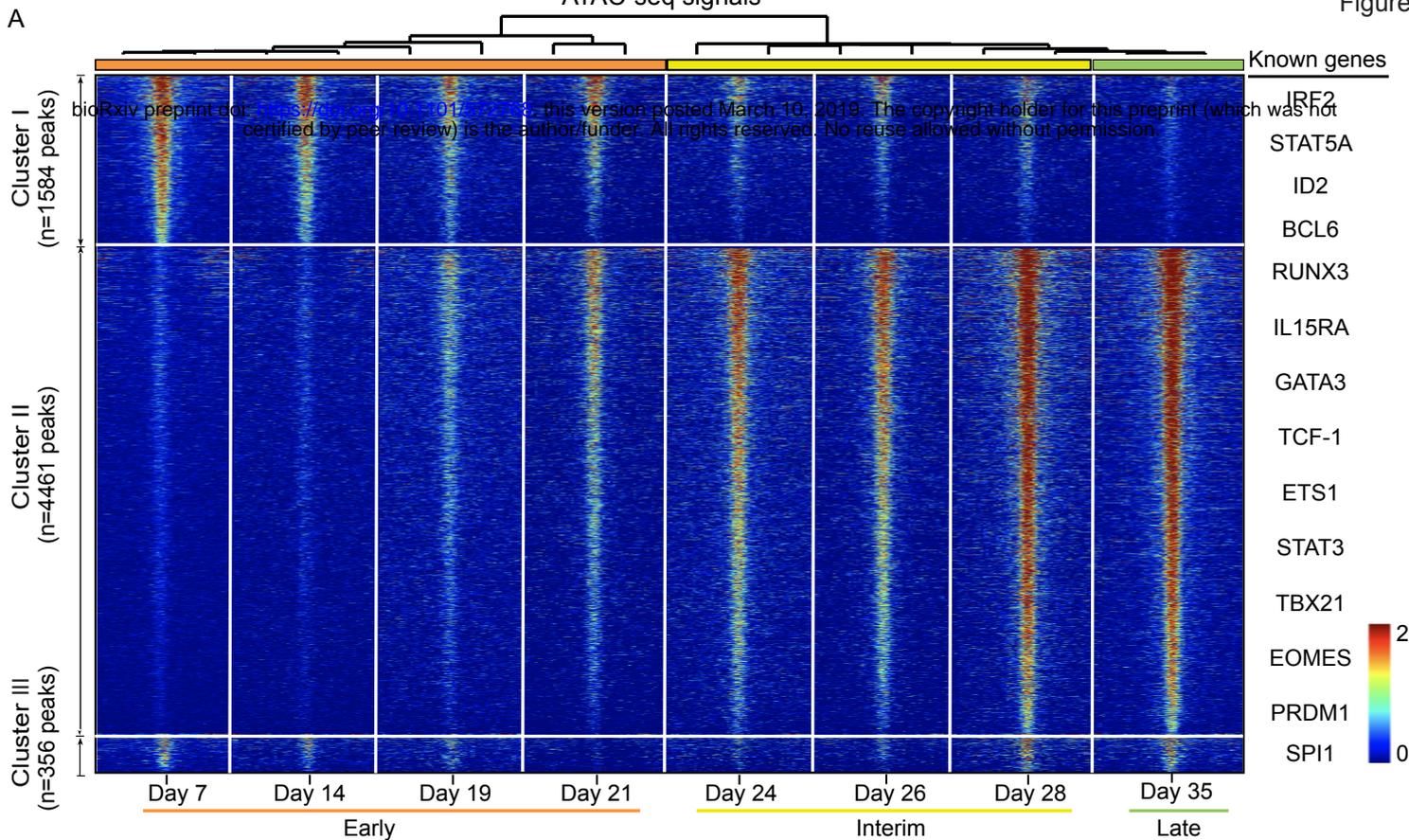
775 **B:** The number of TFs that qualify fold changes. X axis represents different FDC (fold change).
776 The y axis represents the number of TFs that pass the fold change cutoff.

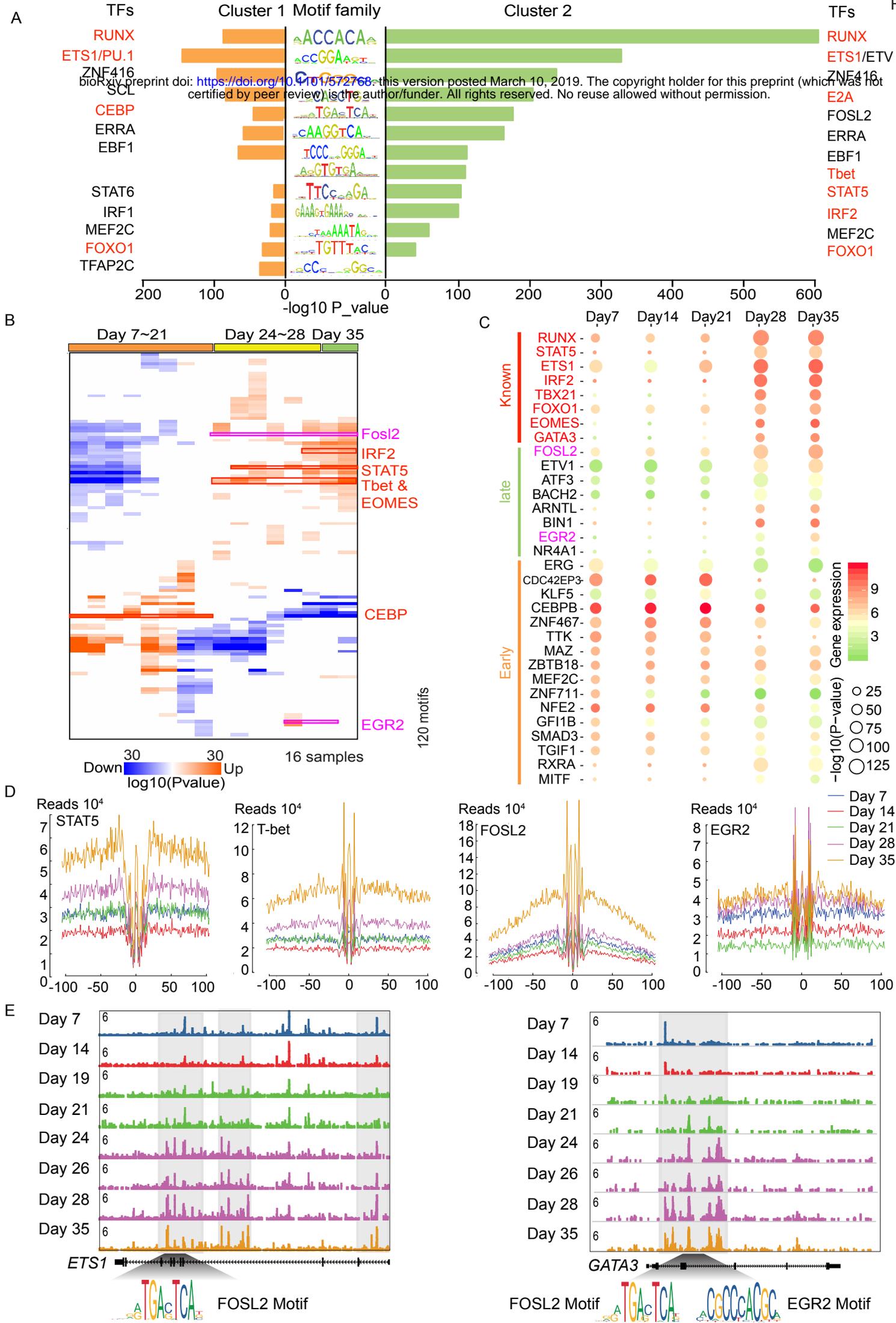
777 **C:** The ratio of known TFs versus all known TFs that qualify the different fold change cutoffs.

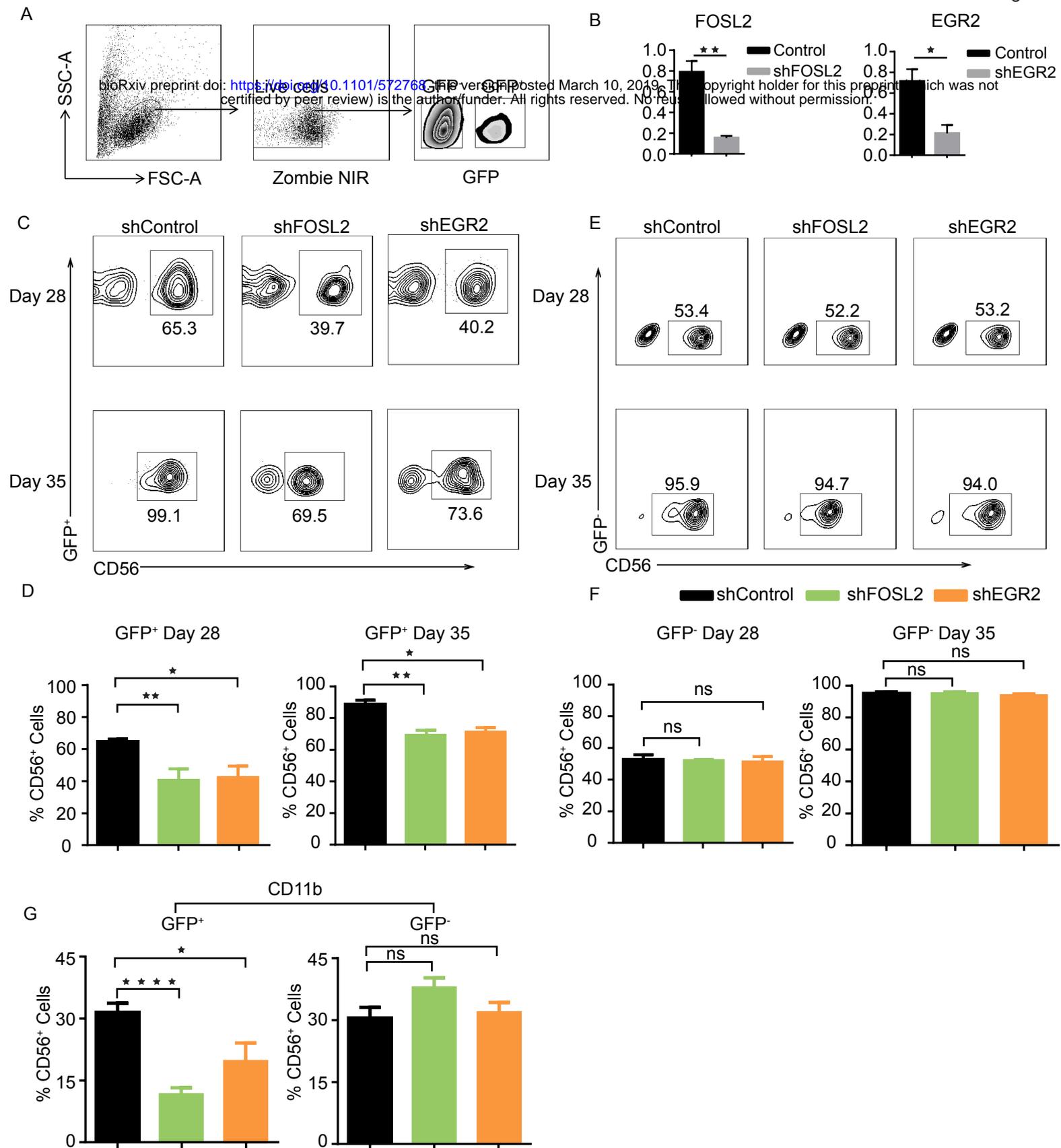
778 Known TFs: TFs regulating NK cell development reported in the literature.



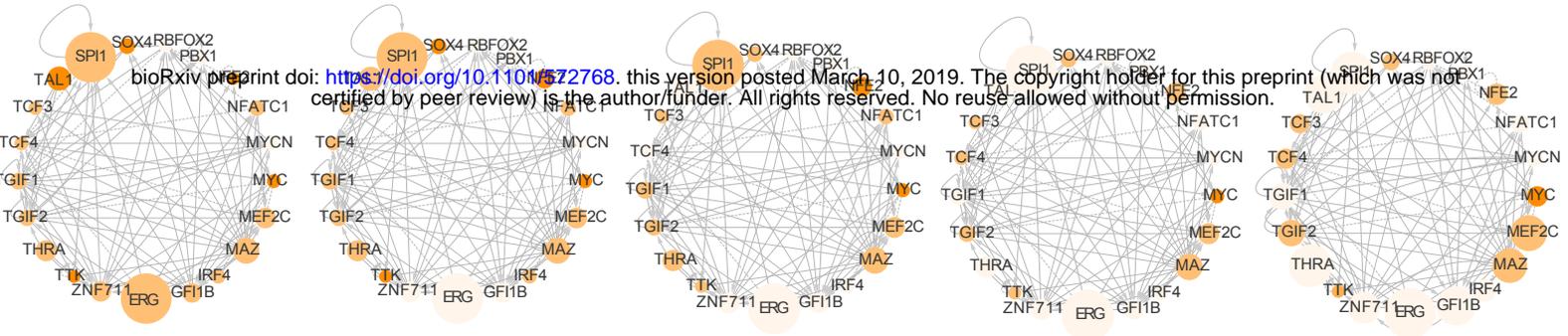
ATAC-seq signals



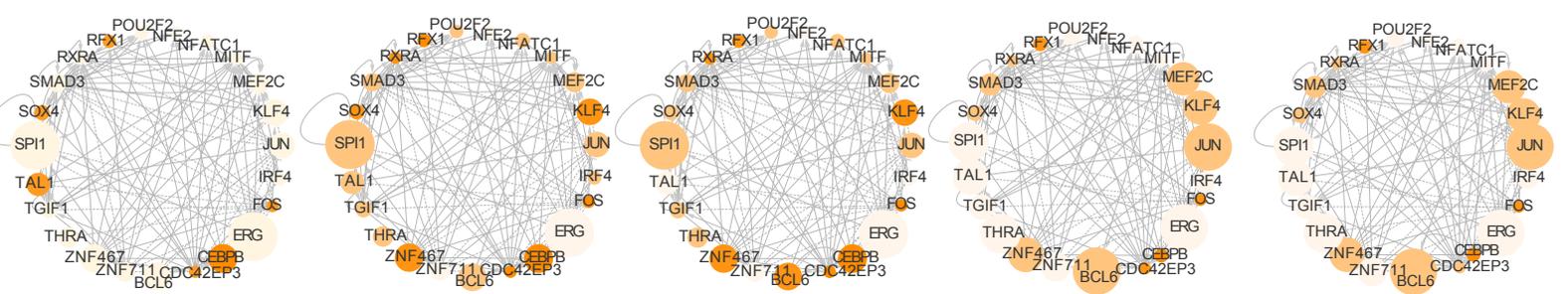




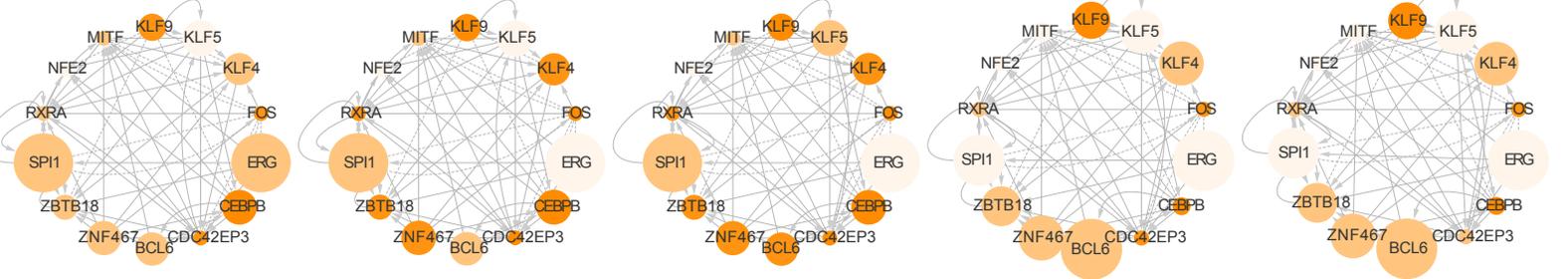
A Day 7 network



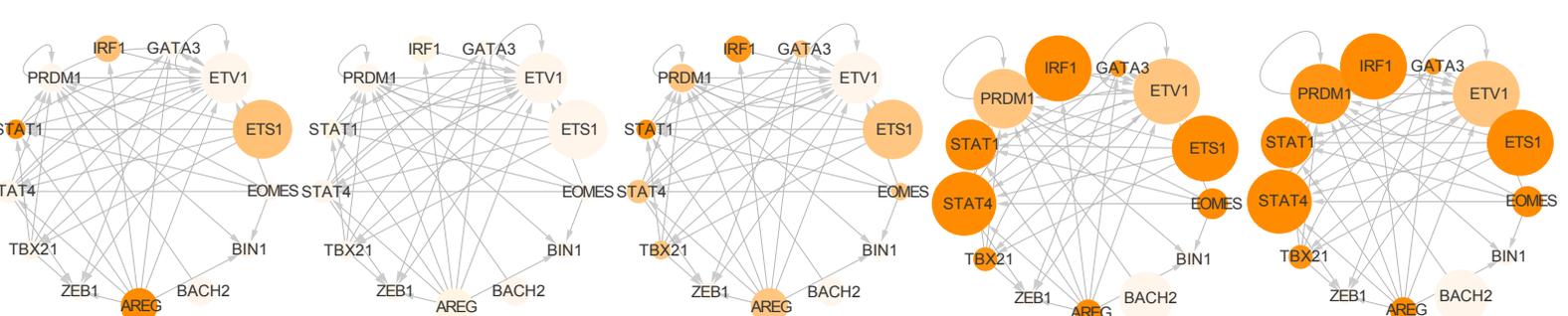
B Day 14 network



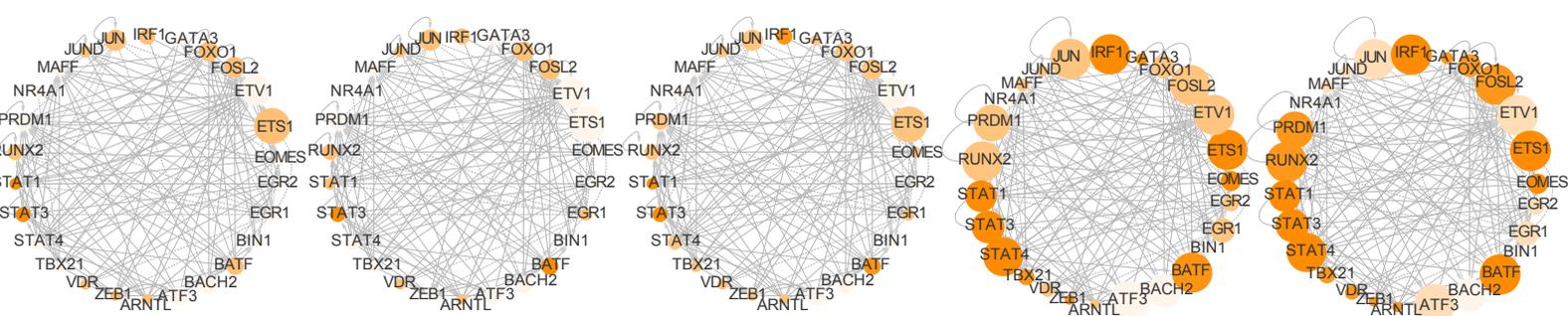
C Day 21 network



D Day 28 network



E Day 35 network



Day 7 → Day 14 → Day 21 → Day 28 → Day 35

→ positive correlation

⋯ no correlation

⊥ negative correlation

$-\log_{10}(P\text{-value})$
 ○ 0 ○ 40 ○ 80 ○ >100

Gene expression
 4 12