Characteristics of functional enrichment and gene expression level of human putative transcriptional target genes

Naoki Osato[1*]

[1]Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University, Osaka 565-0871, Japan

[*]Corresponding author

E-mail address: naokiosato11@gmail.com, nosato@ist.osaka-u.ac.jp

Abstract

Background: Transcriptional target genes show functional enrichment of genes. However, how many and how significantly transcriptional target genes include functional enrichments are still unclear. To address these issues, I predicted human transcriptional target genes using open chromatin regions, ChIP-seq data and DNA binding sequences of transcription factors in databases, and examined functional enrichment and gene expression level of putative transcriptional target genes.

Results: Gene Ontology annotations showed four times larger numbers of functional enrichments in putative transcriptional target genes than gene expression information alone, independent of transcriptional target genes. To compare the number of functional enrichments of putative transcriptional target genes between cells or search conditions, I normalized the number of functional enrichment by calculating its ratios in the total number of transcriptional target genes. With this analysis, native putative transcriptional target genes showed the largest normalized number of functional enrichments, compared with target genes including 5 – 60% of randomly selected genes. The normalized number of functional enrichments was changed according to the criteria of enhancer-promoter interactions such as distance from transcriptional start sites and orientation of CTCF-binding sites. Forward-reverse orientation of CTCF-binding sites showed significantly higher normalized number of functional enrichments than the other orientations. Journal papers showed that the top five frequent functional enrichments were related to the cellular functions in the three cell types. The median

expression level of transcriptional target genes changed according to the criteria of enhancer-promoter assignments (i.e. interactions) and was correlated with the changes of the normalized number of functional enrichments of transcriptional target genes.

Conclusions: Human putative transcriptional target genes showed significant functional enrichments. Functional enrichments were related to the cellular functions. The normalized number of functional enrichments of human putative transcriptional target genes changed according to the criteria of enhancer-promoter assignments and correlated with the median expression level of the target genes. These analyses and characters of human putative transcriptional target genes would be useful to examine the criteria of enhancer-promoter assignments and to predict the novel mechanisms and factors such as DNA binding proteins and DNA sequences of enhancer-promoter interactions.

Keywords: transcriptional target genes, functional enrichment, gene expression, transcription factors, enhancer, open chromatin regions, ChIP-seq, CTCF

Background

More than 400 types of cells have been found in the human body. Human development is accompanied by the differentiation of stem cells into various cell types, leading to a diversification of their phenotypes and functions. For example, the development of the immune system involves differentiation and diversification of stem

55 cells into various types of mature immune cells. The functions of monocytes include
56 phagocytosis and antigen presentation. $CD4^+$ T cells, however, play a central role in
57 cell-mediated immunity and are involved in the activation of phagocytes and
58 antigen-specific cytotoxic T-lymphocytes, and the release of various cytokines in
59 response to an antigen. The $CD20^+$ B cells are involved in the production of antibodies
60 against antigens.

61 Differentiation of cells is often triggered by the expression of transcription factors
62 (TF) followed by the expression of their target genes, which results in the
63 transformation of cells into other cell types. For example, the transcription factors PU.1
64 and CCAAT enhancer-binding protein α (C/EBPα) play a critical role in the expression
65 of myeloid-specific genes and the generation of monocytes and macrophages [1, 2]. The
66 transcription factor GATA-3 is essential for early T cell development and the
67 differentiation of naive $CD4^+$ T cells into Th2 effector cells [3]. E2A, EBF1, PAX5, and
68 Ikaros are among the most important transcription factors that control early
69 development in mice, thereby conditioning homeostatic B cell lymphopoiesis [4].

70 We previously examined the differentiation of monocytes and macrophages in
71 mice, and discovered that the transcription factor IRF8 was essential for cellular
72 differentiation [5]. An analysis of transcription factor-binding sites (TFBS) revealed that
73 IRF8 regulated the expression of KLF4 through the IRF8 transcriptional cascade.
74 Functional enrichment analyses revealed that the target genes of IRF8 showed
75 functional enrichment for antigen presentation, whereas those of KLF4 showed
76 functional enrichments for phagocytosis and locomotion. These results suggested that

77 the transcriptional cascades of IRF8 and KLF4 included different functional modules of
78 target genes.

79 Functional enrichments of transcriptional cascades of IRF8 and KLF4 appeared to
80 be related to the cellular functions of monocytes and macrophages. Although several
81 transcription factors were expressed in monocytes and macrophages, the number of
82 these transcriptional target genes that resulted in functional enrichments remains
83 unknown. Whether transcriptional target genes in other human cells show functional
84 enrichments remain unclear. If the transcriptional target genes showed significant
85 functional enrichment, analyzing transcriptional target genes would be useful in
86 identifying genes involved in a specific cellular function. Using the budding yeast,
87 previous studies examined the functional enrichments on a genome-scale genetic
88 interaction map using the GeneMANIA algorithm [6-8]. Using bacterial systems, the
89 analyses of functional enrichments of predicted regulatory networks were performed
90 using Gene Ontology annotations [9]. Various databases of functional annotations of
91 genes and pathways exist. Analysis of functional enrichments is expected to be useful
92 for understanding the association of genes involved in similar functions and same
93 pathways, and for predicting unknown gene functions such as non-protein-coding
94 RNAs. In addition, the extent of enhancer contribution to functional enrichments of
95 transcriptional target genes remains unknown.

96 In this study, transcriptional target genes were predicted using public databases of
97 open chromatin regions of human monocytes, naive $CD4^+$ T, $CD20^+$ B cells, HUVEC,
98 IMR90, MCF-7, HMEC, H1-hESC, iPSC, and ChIP-seq data of human H1-hESC cells

99 and known transcription factor binding sequences. Functional enrichment analyses of
100 putative transcriptional target genes were conducted using 10 different annotation
101 databases of functional annotations and pathways. The gene expression level of
102 transcriptional target genes was examined in the cells.
103

104 ## Results

105 **Prediction of transcriptional target genes**

106 To examine functional enrichments of transcriptional target genes in a genome
107 scale, transcriptional target genes were predicted in human monocytes, $CD4^+$ T cells,
108 and $CD20^+$ B cells. Searches for known transcription factor binding sequences, which
109 were collected from various databases and papers, were conducted in open chromatin
110 regions of the promoter sequences of RefSeq transcripts (Figure 1, see Methods).
111 Among 6,277 transcription factor binding sequences derived from vertebrates, 4,373
112 were linked to 1,018 TF transcripts computationally (see Methods). To maintain the
113 sensitivity of the searches for transcription factor binding sites and as some transcription
114 factors will recognize multiple distinctly different sequence motifs, transcription factor
115 binding sequences that targeted the same genes were recognized as redundant, and one
116 of the sequences was used [10] (see Methods). In total, 3,337 transcription factor
117 binding sequences in human monocytes, 3,652 in $CD4^+$ T cells, and 3,187 in $CD20^+$ B
118 cells were identified with their target genes, which were selected from highly expressed
119 genes in a cell (top 30% expression level, see Methods).

120 The total numbers of unique highly expressed target genes of transcription factor

121 binding sequences were 4,481, 7,558, and 4,753 in monocytes, $CD4^+$ T cells, and
122 $CD20^+$ B cells respectively using promoters. The mean target genes of a transcription
123 factor were 124, 164, and 144 in monocytes, $CD4^+$ T cells, and $CD20^+$ B cells,
124 respectively, with the corresponding medians being 24, 33, and 24, respectively. With
125 regard to the genomic localizations of TFBS, 51%, 65%, and 61% of TFBS were
126 located within promoters (±5 kb of TSS) of target genes in monocytes, $CD4^+$ T cells,
127 and $CD20^+$ B cells, respectively (according to association rule 1, see Methods).
128
129 **Functional enrichments of putative transcriptional target genes**
130 Functional enrichments of the putative target genes were examined. The
131 distribution of functional enrichments in transcriptional target genes was predicted
132 using genome sequences of promoters in the three cell types (Figure 1 and Table 1, see
133 Methods). Furthermore, the effect of transcriptional target genes including randomly
134 selected genes on functional enrichments was investigated using DNase-DGF data of
135 monocytes, $CD4^+$ T and $CD20^+$ B cells, HUVEC, IMR90, MCF-7, HMEC, and
136 ChIP-seq data of H1-hESC (Figure 2A and B, and Figure S1 and S2, see Methods). The
137 native putative transcriptional target genes not including randomly selected genes
138 showed the highest functional enrichments using Gene Ontology, GO Slim, KEGG,
139 Pathway Commons, WikiPathways, InterPro and UniProt functional regions (Domains)
140 in both DNase-DGF and ChIP-seq data of the five types of cells. Of the 10 databases
141 used in this analysis, the Gene Ontology database consists of three types of functional
142 annotations, i.e., 20,836 biological processes, 9,020 molecular functions, and 2,847

143 cellular components. The numbers of functional enrichments of Gene Ontology
144 annotations in target genes of a transcription factor were 2,902, 4,077, and 2,778 in
145 monocytes, $CD4^+$ T cells, and $CD20^+$ B cells, respectively. An examination of
146 functional enrichments of highly expressed genes (top 30% expression level),
147 independent of transcriptional target genes, revealed 237, 301, and 239 'unique' Gene
148 Ontology annotations in monocytes, $CD4^+$ T cells, and $CD20^+$ B cells, respectively
149 (Table 1). Further, the examination of functional enrichments of highly expressed target
150 genes (top 30% expression level) in target genes revealed 1,271, 1,654, and 1,192
151 'unique' Gene Ontology annotations in monocytes, $CD4^+$ T cells, and $CD20^+$ B cells,
152 respectively i.e., These numbers were four times larger than functional enrichments
153 identified by gene expression information alone, independent of transcriptional target
154 genes, suggesting that transcriptional target genes were frequently associated with
155 similar functions or pathways (Table S3 and S4).

156 Functional enrichments of transcriptional target genes from other databases were
157 also examined (Table 1). KEGG, Target genes of transcription factors, Disease Ontology,
158 GO Slim, Pathway Commons, Cellular biomarkers, Target genes of microRNAs,
159 Protein domains, and WikiPathways had 95, 16, 127, 12, 242, 17, 97, 303, and 105
160 unique functional annotations, respectively. The numbers of functional enrichments of
161 transcriptional target genes in the other annotation databases except for microRNAs and
162 Protein domains were significantly higher than gene expression information alone,
163 independent of transcriptional target genes, as well as Gene Ontology annotations
164 (Table 1). The functional enrichments of transcriptional target genes from Pathway

165 Commons for monocytes, $CD4^+$ T cells, and $CD20^+$ B cells are shown in Table 2 and
166 Table S5. Functional enrichments were found to be related to cellular functions, e.g.,
167 interferon signaling, GMCSF (Granulocyte-macrophage colony-stimulating factor, a
168 kind of cytokine)-mediated signaling events, antigen processing-cross presentation in
169 monocytes; TCR (T-cell receptor) signaling in naive $CD4^+$ T cells, IL-12 (Interleukin-12,
170 a kind of cytokine)-mediated signaling events, and downstream signaling in naive $CD8^+$
171 T cells in $CD4^+$ T cells; interferon alpha/beta signaling, IL8- and CXCR2 (Chemokine
172 receptor type 2, a kind of cytokine)-mediated signaling events, and BCR (B cell antigen
173 receptor) signaling pathway in $CD20^+$ B cells. WikiPathways, KEGG and GO also
174 revealed that functional enrichments were associated with cellular functions (Table S6,
175 S7 and S8).
176
177 **Effect of enhancer-promoter association rules on functional enrichments**
178 To understand the effect of 'promoter and extended regions for
179 enhancer-promoter association (EPA)' on the functional enrichments of target genes, the
180 rule of extended regions was modified according to four criteria (Figure 3A and see
181 Methods) [11], and functional enrichments were investigated.
182 According to the association rule (1), the means of target genes were 177, 217,
183 and 175 in monocytes, $CD4^+$ T cells, and $CD20^+$ B cells, respectively, whereas the
184 corresponding medians were 55, 58, and 37, respectively (Table S9). The numbers of
185 functional enrichments of Pathway Commons annotations using promoter regions were
186 1,005, 1,806, and 821 in monocytes, $CD4^+$ T cells, and $CD20^+$ B cells, respectively

187 (Table S10). With the use of EPA (association rule 1), the numbers of functional
188 enrichments of Pathway Commons annotations were 3,087, 7,216, and 3,900,
189 representing 3.07-, 4.00-, and 4.75-fold increases, respectively, in the three cells types.
190 Additionally, the numbers of 'unique' Pathway Commons annotations with promoter
191 regions were 321, 415, and 329 in monocytes, $CD4^+$ T cells, and $CD20^+$ B cells,
192 respectively; the corresponding numbers with the use of EPA (association rule 1) were
193 364, 437, and 364, representing 1.13-, 1.05-, and 1.11-fold increases, respectively, in the
194 three cell types. The normalized numbers of functional enrichments of Pathway
195 Commons annotations were 44.75, 84.51, and 59.32, representing 1.84-, 2.80-, and
196 3.32-fold increases, respectively, in the three cell types (association rule 1, Table 3).
197 Other cell types also showed the same tendencies (Table 3).

198 The normalized numbers of the functional enrichments of transcriptional target
199 genes showed association rule (4) as the highest number, followed by association rule
200 (1) and (2) in the three cell types. Although association rule (3) was the longest among
201 the four criteria, it showed the lowest number of functional enrichments in the three cell
202 types (Figure 3A and Table 3). ChIP-seq data of 19 TF in H1-hESC (Human embryonic
203 stem cells) also showed almost the same tendency (difference between association rule
204 (4) and (1) was not statistically significant, probably due to a large number of
205 transcriptional target genes predicted using 19 TF ChIP-seq data. Several thousands of
206 target genes of each TF were predicted. Some of them would be indirect interactions
207 between TF and genome DNA, which were identified by ChIP-seq experiments. (Table
208 S11, see Additional file).

209 Differences in functional enrichments using Pathway Commons were examined
210 between promoters *versus* EPA (association rule 1) (Table 4 and Table S12). A
211 comparison of 321 and 364 functional enrichments using the promoters and EPA,
212 respectively, in monocytes revealed that 152 (47% in promoters, 42% in extended
213 regions) of them were common. For example, IFN-gamma (Interferon gamma) pathway,
214 GMCSF (Granulocyte-macrophage colony-stimulating factor, a kind of
215 cytokine)-mediated signaling events, and PDGF (Platelet-derived growth factor)
216 receptor signaling network were enriched using extended regions (association rule 1) as
217 opposed to promoters (Table S12). The comparison of 415 (promoters) and 437
218 (extended regions) functional enrichments in $CD4^+$ T cells revealed that 163 of them
219 (39% in promoters, 37% in extended regions) were common. IFN-gamma pathway,
220 TCR (T-cell receptor) signaling in naive $CD4^+$ T cells, and IL3 (Interleukin-3, a kind of
221 cytokine)-mediated signaling events were enriched using extended regions. The
222 comparison of 329 (promoters) and 364 (extended regions) functional enrichments in
223 $CD20^+$ B cells revealed that 171 of them (52% in promoters, 47% in extended regions)
224 were common. IL5-mediated signaling events, IL4-mediated signaling events, and
225 cytokine signaling in immune system were enriched in $CD20^+$ B cells using extended
226 regions. Only about 40% of functional enrichments of Pathway Commons annotations
227 were unchanged between promoters and EPA. EPA significantly affected the functional
228 enrichments of transcriptional target genes. Journal papers showed that frequent
229 functional enrichments were related to the cellular functions in the three cell types
230 (Table 4). These results showed that new functional enrichments related to cellular

231 functions were identified using extended regions for enhancer-promoter association.

232

233 **Effect of CTCF-binding sites on functional enrichments**

234 CTCF have the activity of insulators to block the interaction between enhancers
235 and promoters [12]. Recent studies identified a correlation between the orientation of
236 CTCF-binding sites and chromatin loops (Figure 3B) [13]. Forward–reverse (FR)
237 orientation of CTCF-binding sites are frequently found in chromatin loops. To examine
238 the effect of forward–reverse orientation of CTCF-binding sites on functional
239 enrichments of target genes, 'promoter and extended regions for enhancer-promoter
240 association (EPA)' were shortened at the genomic locations of forward–reverse
241 orientation of CTCF-binding sites, and transcriptional target genes were predicted from
242 the shortened regions using TFBS (see Methods). The numbers of functional
243 enrichments of target genes were investigated. According to EPA (association rule 4)
244 that were shortened at genomic locations of forward–reverse orientation of
245 CTCF-binding sites, the means of target genes were 67, 64, and 77 in monocytes, $CD4^+$
246 T cells, and $CD20^+$ B cells, respectively, whereas the corresponding medians were 23,
247 21, and 20, respectively (Table S13). The normalized numbers of functional
248 enrichments of Pathway Commons annotations using EPA were 71.42, 108.08, and
249 90.99 in monocytes, $CD4^+$ T cells, and $CD20^+$ B cells, respectively (Table 5). With the
250 use of EPA shortened at forward–reverse orientation of CTCF-binding sites, the
251 normalized numbers of functional enrichments of Pathway Commons annotations were
252 196.58, 220.54, and 220.77, representing 2.75-, 2.04-, and 2.43-fold increases,

253 respectively, in the three cells types. Additionally, the normalized numbers of functional
254 enrichments of 'unique' Pathway Commons annotations with EPA were 5.09, 5.34, and
255 6.00 in monocytes, $CD4^+$ T cells, and $CD20^+$ B cells, respectively; the corresponding
256 normalized numbers with the use of EPA shortened at forward–reverse orientation of
257 CTCF-binding sites were 9.88, 10.72, and 9.10, representing 1.94-, 2.01-, and 1.52-fold
258 increases, respectively, in the three cell types (Table S14). Other cell types also showed
259 the same tendencies (Table 5). The normalized numbers of functional enrichments were
260 significantly increased between EPA and EPA shortened at forward–reverse orientation
261 of CTCF-binding sites in Gene Ontology, Disease Ontology, Pathway Commons, GO
262 Slim, WikiPathways, KEGG, InterPro and UniProt functional regions (Domains)
263 annotations. These increases were also significant, compared with EPA shortened at
264 CTCF-binding sites without the consideration of their orientation. Transcriptional target
265 genes predicted from EPA shortened at forward–reverse orientation of CTCF-binding
266 sites tend to include similar function of genes significantly.

267 Differences in functional enrichments obtained using EPA *versus* EPA shortened
268 at forward–reverse orientation of CTCF-binding sites were examined using the
269 functional enrichments of Pathway Commons (Table 6 and Results in Additional file).
270 Transcriptional target genes predicted from EPA shortened at the CTCF-binding sites
271 tended to include the similar function of genes. About 40 – 80% of functional
272 enrichments were unchanged between promoters and EPA shortened at forward–reverse
273 orientation of CTCF-binding sites, and the functional enrichments observed in EPA
274 shortened at forward–reverse orientation of CTCF-binding sites as opposed to

275 promoters included various immunological terms. Journal papers showed that the top
276 five frequent functional enrichments were related to the cellular functions in the three
277 cell types (Table 6). These results showed that new functional enrichments related to
278 cellular functions were identified using forward–reverse orientation of CTCF-binding
279 sites.
280
281 **Comparison of expression levels of putative transcriptional target genes**
282 To examine the relationship between functional enrichments and expression
283 levels of target genes, the expression levels of target genes predicted from promoters
284 and three types of 'promoter and extended regions for enhancer-promoter assignment
285 (EPA)' were investigated in monocytes, $CD4^+$ T, H1-hESC and iPSC (Figure 4). Median
286 expression levels of the target genes of the same transcription factor binding sequences
287 were compared between promoters and three types of EPA. Red and blue dots in Figure
288 4 show statistically significant difference of the distribution of expression levels of
289 target genes between promoters and EPA. Additionally, "red dots" show the median
290 expression level of target genes of a TFBS was 'higher' in EPA than promoters, and
291 "blue dots" show the median expression level of target genes of a TFBS was 'lower' in
292 EPA than promoters. The ratios of red dots were higher in EPA (association rule 4) that
293 were shortened at forward–reverse orientation of CTCF-binding sites *versus* promoters
294 (left graph in Figure 4) than EPA (association rule 4) *versus* promoters (right graph) in
295 monocytes and $CD4^+$ T cells. The ratios of blue dots were higher in EPA (association
296 rule 4) that were shortened at forward–reverse orientation of CTCF-binding sites *versus*

297 promoters (left graph) than EPA (association rule 4) *versus* promoters (right graph) in
298 H1-hESC and iPSC. Moreover, the ratio of the sum of median expression levels
299 between the three types of EPA and promoters in monocytes and $CD4^+$ T cells was the
300 highest in EPA shortened at forward–reverse orientation of CTCF-binding sites (Table
301 S16). Conversely, the ratio of the sum of median expression levels between the three
302 types of EPA and promoters in H1-hESC and iPSC was the lowest in EPA shortened at
303 forward–reverse orientation of CTCF-binding sites.

304 EPA shortened at forward–reverse orientation of CTCF-binding sites changed (i.e.
305 increased or decreased) the expression levels of target genes more than the other types
306 of EPA. This implied that gene expression tended to be activated in monocytes and
307 $CD4^+$ T cells, but repressed in H1-hESC and iPSC by enhancers. EPA shortened at
308 forward–reverse orientation of CTCF-binding sites also showed the highest normalized
309 number of functional enrichments of transcriptional target genes, as shown in the
310 previous paragraphs.

311

312 **Discussion**

313 Genome-wide functional enrichments and gene expression levels of putative
314 target genes of human transcription factors were investigated. Human putative
315 transcriptional target genes showed significantly larger numbers of functional
316 enrichments than gene expression information alone, independent of transcriptional
317 target genes. Moreover, when the number of functional enrichments of human putative
318 transcriptional target genes was normalized by the total number of transcriptional target

319 genes, native putative transcriptional target genes showed the highest ratio of functional
320 enrichments, compared with target genes partially including randomly selected genes.
321 The ratio of functional enrichments was decreased according to the increase of the ratio
322 of randomly selected genes in target genes. These tendencies were observed in putative
323 transcriptional target genes predicted from both open chromatin regions and ChIP-seq
324 data of transcription factors. Prediction of transcriptional target genes from open
325 chromatin regions includes false positives, since DNase I cleavage bias affects the
326 computational analysis of DNase-seq experiments [14]. However, the detection of
327 ChIP-seq peaks is also changed depending on the methods to identify them and the
328 depth of DNA sequencing of ChIP-seq experiments [15]. Though human putative
329 transcriptional target genes include false positives, they showed significantly the largest
330 number of functional enrichments, compared with target genes including 5 – 60% of
331 randomly selected genes (Figure 2).

332 The median expression level of human putative transcriptional target genes was
333 changed according to the criteria of enhancer-promoter assignments, and was correlated
334 with the normalized number of functional enrichments. The median expression level of
335 transcriptional target genes was 'decreased' significantly in transcriptional target genes
336 predicted using enhancers, compared with those predicted using promoters in H1-hESC
337 and iPSC, and the median expression level was 'increased' significantly in target genes
338 predicted using enhancers, compared with those predicted using promoters in
339 monocytes and $CD4^+$ T cells. These results implied that transcription factors bound in
340 enhancers act as repressors in H1-hESC (ES) and iPSC, but those act as activators in

341 monocytes and $CD4^+$ T cells. The change of functional roles of transcription factors
342 depending on the cell types would be analyzed and reported elsewhere.

343 The median expression level was increased significantly in target genes predicted
344 using enhancers, compared with those predicted from promoters in immune cells using
345 gene expression data (Blueprint RNA-seq RPKM data; GSE58310), but smaller number
346 of target genes showed the increase of median expression level using gene expression
347 data (ENCODE; GSM984609). The results of the analyses may be slightly different
348 depending on gene expression data. H1-hESC (ES) and iPSC showed a strong tendency
349 of decrease of median expression levels of transcriptional target genes between
350 enhancers and promoters.

351 The gene symbols of transcription factors were sometimes different among
352 databases, because more than one gene symbol are assigned to some transcription
353 factors and some gene symbols are spelled in several different ways. These differences
354 need to be identified with manual curations. This analysis will be required to predict
355 transcriptional cascades by associating transcription factors with transcriptional target
356 genes consisting of transcription factors. In the analyses of transcriptional cascades, to
357 reduce false positive predictions of enhancer-promoter associations from open
358 chromatin regions, the identification of DNase peaks will be modified using a new tool
359 such as HINT [16].

360 In this study, I focused on three types of immune cells and stem cells such as
361 H1-hESC and iPSC to examine transcriptional target genes in a genome scale, since in
362 my previous study, I examined transcriptional cascades involved in the differentiation of

363 immune cells as introduced in Background [5]. Furthermore, I confirmed the features of
364 functional enrichments of putative transcriptional target genes are commonly found in
365 other four types of normal and disease cells (HUVEC, IMR90, MCF-7 and HMEF).

366 It is difficult to predict enhancer-promoter associations using a single parameter,
367 so that machine learning methods to combine several parameters have been proposed
368 [17-19]. These methods showed high accuracy in predicting enhancer-promoter
369 associations (I tried to use some of the tools, but they did not work properly. I am
370 waiting for the authors to update the tools). However, molecular mechanisms of
371 enhancer-promoter interactions are not clearly understood. CTCF has been found to
372 bind at chromatin interaction anchors and form chromatin interactions [12]. About
373 20-40% of chromatin interaction anchors included DNA binding sequences of CTCF,
374 when I examined public Hi-C experimental data [20, 21]. Among 33,939 RefSeq
375 transcripts, 7,202 (21%), 4,404 (13%), and 6,921 (20%) (p-value $< 10^{-5}$ in the search for
376 CTCF-binding motifs using FIMO) to 9,608 (28%), 5,806 (17%), and 9,137 (27%)
377 (p-value $< 10^{-4}$) of transcripts had forward–reverse orientation of CTCF-binding sites
378 within 1 Mb from transcriptional start sites in the three immune cell types, respectively.
379 These analyses implied that other factors might be involved in chromatin interactions.
380 ZNF143 has been reported to locate at promoter regions of chromatin interaction
381 anchors [22]. To predict the other factors and molecular mechanisms, the analyses in
382 this study would be useful to examine further the criteria in predicting
383 enhancer-promoter associations. Machine learning methods need the information what
384 parameters should be used for prediction, so it would be better to choose parameters

385 involved in predicting enhancer-promoter associations. To improve the prediction and
386 understand the molecular mechanisms of enhancer-promoter interactions, I am
387 promoting the analyses of chromatin interaction anchors, and the results of the analyses
388 will be reported elsewhere.
389

390 **Conclusion**

391 In this study, human transcriptional target genes were predicted using open
392 chromatin regions, ChIP-seq data, and DNA binding sequences of transcription factors
393 in databases. Human putative transcriptional target genes showed significant functional
394 enrichments. Journal papers showed that frequent functional enrichments were related
395 to the cellular functions. The normalized number of functional enrichments was the
396 highest in native putative transcriptional target genes, compared with target genes
397 partially replaced with randomly selected genes. The normalized number of functional
398 enrichments of human putative transcriptional target genes changed according to the
399 criteria of enhancer-promoter assignments and correlated with the median expression
400 level of the target genes. The normalized numbers of functional enrichments of
401 transcriptional target genes did not show the highest number in the criterion of
402 enhancer-promoter assignments covering the longest distance from transcriptional start
403 site among four criteria. This suggested that there is a criterion of enhancer-promoter
404 assignments that shows the highest normalized number of functional enrichments. The
405 median expression level of transcriptional target genes was 'decreased' significantly in
406 transcriptional target genes predicted using enhancers, compared with those predicted

407 using promoters in H1-hESC and iPSC, and the median expression level was 'increased'
408 significantly in target genes predicted using enhancers, compared with those predicted
409 using promoters in immune cells. These results implied that transcription factors bound
410 in enhancers act as repressors in H1-hESC (ES) and iPSC, but those act as activators in
411 immune cells. These analyses and characters of human putative transcriptional target
412 genes would be useful to examine the criteria of enhancer-promoter assignments and to
413 predict the novel mechanisms and factors such as DNA binding proteins and DNA
414 sequences of enhancer-promoter interactions.
415

416 ## Methods

417 **Searches for transcription factor binding sequences from open chromatin regions**

418 To examine transcriptional regulatory target genes, bed files of hg19 narrow
419 peaks of ENCODE DNase-DGF and DNase data for Monocytes-CD14^{+}_RO01746
420 (GSM1024791; UCSC Accession: wgEncodeEH001196), CD4^{+}_Naive_Wb11970640
421 (GSM1014537; UCSC Accession: wgEncodeEH003156), CD20^{+}_RO01778
422 (GSM1014525; UCSC Accession: wgEncodeEH002442), H1-hESC (GSM816632;
423 UCSC Accession: wgEncodeEH000556), iPSC (GSM816642; UCSC Accession:
424 wgEncodeEH001110), HUVEC (GSM1014528; UCSC Accession:
425 wgEncodeEH002460), IMR90 (GSM1008586; UCSC Accession: wgEncodeEH003482),
426 MCF-7 (GSM816627; UCSC Accession: wgEncodeEH000579), and HMEC
427 (GSM816669; UCSC Accession: wgEncodeEH001101) from the ENCODE website
428 (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwDgf/)

429 were used. For comparison with transcriptional target genes predicted using ChIP-seq
430 data, bed files of hg19 narrow peaks of ENCODE ChIP-seq data for 19 transcription
431 factors (TF) (BACH1, BRCA1, C/EBPbeta, CHD2, c-JUN, c-MYC, GTF2I, JUND,
432 MAFK, MAX, MXI1, NRF1, RAD21, RFX5, SIN3A, SUZ12, TBP, USF2, ZNF143) in
433 H1-hESC from the ENCODE website
434 (https://genome.ucsc.edu/cgi-bin/hgFileUi?db=hg19&g=wgEncodeAwgTfbsUniform)
435 were utilized.

436 To identify transcription factor binding sites (TFBS) from the DNase-DGF data,
437 TRANSFAC (2013.2), JASPAR (2010), UniPROBE, BEEML-PBM, high-throughput
438 SELEX, Human Protein-DNA Interactome, and transcription factor binding sequences
439 of ENCODE ChIP-seq data were used [23-29]. Position weight matrices of transcription
440 factor binding sequences were transformed into TRANSFAC matrices and then into
441 MEME matrices using in-house Perl scripts and transfac2meme in MEME suite [30].
442 Transcription factor binding sequences of transcription factors derived from vertebrates
443 were used for further analyses. Searches were conducted for transcription factor binding
444 sequences from the central 50-bp regions of each narrow peak using FIMO with *p*-value
445 threshold of 10^{-5} [31]. Transcription factors corresponding to transcription factor
446 binding sequences were searched computationally by comparing their names and gene
447 symbols of HGNC (HUGO Gene Nomenclature Committee) -approved gene
448 nomenclature and 31,848 UCSC known canonical transcripts
449 (http://hgdownload.cse.ucsc.edu/goldenpath/hg19/database/knownCanonical.txt.gz), as
450 transcription factor binding sequences were not linked to transcript IDs such as UCSC,

451 RefSeq, and Ensembl transcripts.

452

453 **Prediction of transcriptional target genes**

454 Target genes of a transcription factor were assigned when its TFBS was found in
455 DNase-DGF narrow peaks in promoter or extended regions for enhancer-promoter
456 association of genes (EPA). Promoter and extended regions were defined as follows:
457 promoter regions were those that were within distances of ±5 kb from transcriptional
458 start sites (TSS). Promoter and extended regions were defined as per the following four
459 association rules, which are similar or same as those defined in a previous study [11]:
460 (1) the basal plus extension association rule assigns a basal regulatory domain to each
461 gene regardless of other nearby genes. The domain is then extended to the basal
462 regulatory domain of the nearest upstream and downstream genes, and includes a 5 kb +
463 5 kb basal region and an extension up to 300 kb or the midpoint between the TSS of the
464 gene and that of the nearest gene upstream and downstream; (2) 5 kb + 1 kb basal
465 region and an extension up to 1 Mb; (3) the two nearest genes association rule, which
466 extends the regulatory domain to the TSS of the nearest upstream and downstream
467 genes without the limitation of extension length; and (4) the single nearest gene
468 association rule, which extends the regulatory domain to the midpoint between the TSS
469 of the gene and that of the nearest gene upstream and downstream without the limitation
470 of extension length. Association rule (1) was used in our previous study [5]. Association
471 rule (2), (3), and (4) were the same as those in Figure 3A of the previous study [11],
472 however, association rules (3) and (4) did not have the limitation of extension length in

473 this study. The genomic positions of genes were identified using 'knownGene.txt.gz'
474 file in UCSC bioinformatics sites [32]. The file 'knownCanonical.txt.gz' was also
475 utilized for choosing representative transcripts among various alternate forms for
476 assigning promoter and extended regions for enhancer-promoter association of the
477 genes. From the list of transcription factor binding sequences and transcriptional target
478 genes, redundant transcription factor binding sequences were removed by comparing
479 the target genes of a transcription factor binding sequence and its corresponding
480 transcription factor; if identical, one of the transcription factor binding sequences was
481 used. When the number of transcriptional target genes predicted from a transcription
482 factor binding sequence was less than five, the transcription factor binding sequence
483 was omitted.
484
485 **Gene expression analyses**
486 For gene expression data, RNA-seq reads mapped onto human hg19 genome
487 sequences were obtained, including ENCODE long RNA-seq reads with poly-A of
488 monocytes $CD14^+$ cells, $CD20^+$ B cells, H1-hESC, iPSC, HUVEC, IMR90, MCF-7, and
489 HMEC (GSM984609, GSM981256, GSE26284, GSM958733, GSM2344099,
490 GSM2344100, GSM958734, GSM2400222, GSM765388, and GSM758571), and
491 UCSF-UBC human reference epigenome mapping project RNA-seq reads with poly-A
492 of naive $CD4^+$ T cells (GSM669617). Two replicates were present for monocytes
493 $CD14^+$ cells, $CD20^+$ B cells, H1-hESC, iPSC, HUVEC, IMR90, MCF-7, and HMEC
494 and a single one for $CD4^+$ T cells. RPKMs of the RNA-seq data were calculated using

495 RSeQC [33]. For monocytes, Blueprint RNA-seq RPKM data (GSE58310,
496 GSE58310_GeneExpression.csv.gz, Monocytes_Day0_RPMI) was also used [34].
497 Based on RPKM, UCSC transcripts with expression levels among top 30% of all the
498 transcripts were selected in each cell type.
499
500 **Functional enrichment analyses**
501 The functional enrichments of target genes of a TFBS and its corresponding
502 transcription factor were examined using GO-Elite v1.2.5 with *p*-value threshold at 1,
503 and after GO-Elite analyses a false discovery rate (FDR) test was performed with
504 *q*-value threshold at 10^{-3} to correct for multiple comparisons of thousands of groups of
505 transcriptional target genes in each cell type and condition [35]. For examining
506 functional enrichments of high or low expressed genes independent of transcriptional
507 target genes, the *p*-value threshold was set to 0.01 or 0.05 to confirm that the results
508 were not significantly changed. UCSC gene IDs were transformed into RefSeq IDs prior
509 to GO-Elite analyses. GO-Elite uses 10 databases for identifying functional
510 enrichments: (1) Gene Ontology, (2) Disease Ontology, (3) Pathway Commons, (4) GO
511 Slim, (5) WikiPathways, (6) KEGG, (7) Transcription factor to target genes, (8)
512 microRNA to target genes, (9) InterPro and UniProt functional regions (Domains), and
513 (10) Cellular biomarkers (BioMarkers). To calculate the normalized numbers of
514 functional enrichments of target genes, the numbers of functional enrichments were
515 divided by the total number of target genes in each cell type and condition, and were
516 multiplied by 10^5. In tables showing the numbers of functional enrichments in 10

517 databases, heat maps were plotted according to *Z*-scores calculated from the numbers of
518 functional enrichments of each database using in-house Excel VBA scripts. In the
519 comparisons of the normalized numbers of functional enrichments of target genes in
520 cell types and conditions, if the number of a functional annotation in a cell type or
521 condition was two times larger than that in the other cell type or condition, the
522 functional annotation was recognized as more enriched than the other cell type or
523 condition.

524 To investigate whether the normalized numbers of functional enrichments of
525 transcriptional target genes correlate with the prediction of target genes, a part of target
526 genes were changed with randomly selected genes with high expression level (top 30%
527 expression level), and functional enrichments of the target genes were examined. First,
528 5%, 10%, 20%, 40%, and 60% of target genes were changed with randomly selected
529 genes with high expression level in monocytes, $CD4^+$ T cells, and $CD20^+$ B cells.
530 Second, as another randomization of target genes, the same number of 5%, 10%, 20%,
531 40%, and 60% of target genes were selected randomly from highly expressed genes,
532 then added them to the original target genes, and functional enrichments of the target
533 genes were examined. All analyses were repeated three times to estimate standard errors
534 (Figure 2A and B, Figure S1 and S2, and Table S1). The same analysis was performed
535 using DNase-DGF data and ChIP-seq data of 19 TF in H1-hESC. Transcriptional target
536 genes were predicted from promoter (Tables S2).

537

538 **CTCF-binding sites**

539 CTCF ChIP-seq data for monocytes $CD14^+$ cells
540 (GSM1003508_hg19_wgEncodeBroadHistoneMonocd14ro1746CtcfPk.broadPeak.gz),
541 $CD4^+$ T cells (SRR001460.bam), $CD20^+$ B cells
542 (GSM1003474_hg19_wgEncodeBroadHistoneCd20CtcfPk.broadPeak.gz), H1-hESC
543 (wgEncodeAwgTfbsUtaH1hescCtcfUniPk.narrowPeak.gz), iPSC (GSE96477), HUVEC
544 (wgEncodeAwgTfbsUwHuvecCtcfUniPk.narrowPeak.gz), IMR90
545 (wgEncodeAwgTfbsSydhImr90CtcfbIggrabUniPktfbsf.narrowPeak.gz), MCF-7
546 (wgEncodeAwgTfbsUwMcf7CtcfUniPktfbsf.narrowPeak.gz), and HMEC
547 (wgEncodeAwgTfbsUwHmecCtcfUniPktfbsf.narrowPeak.gz) were used.
548 SRR001460.bam was sorted and indexed by SAMtools and transformed into a bed file
549 using bamToBed of BEDTools [36, 37]. ChIP-seq peaks were predicted by SICER-rb.sh
550 of SICER with optional parameters 'hg19 1 200 150 0.74 200 100' [38]. Extended
551 regions for enhancer-promoter association (association rule 4) were shortened at the
552 genomic locations of CTCF-binding sites that were the closest to a transcriptional start
553 site, and transcriptional target genes were predicted from the shortened enhancer
554 regions using TFBS. Furthermore, promoter and extended regions for
555 enhancer-promoter association (association rule 4) were shortened at the genomic
556 locations of forward–reverse orientation of CTCF-binding sites. When forward or
557 reverse orientation of CTCF-binding sites were continuously located in genome
558 sequences several times, the most external forward–reverse orientation of
559 CTCF-binding sites were selected.
560

561 ## Abbreviations

562 TF: transcription factors

563 TFBS: transcription factor binding sites

564 TSS: transcriptional start sites

565 ENCODE: encyclopedia of DNA elements

566 ChIP-seq: ChIP-sequencing, chromatin immunoprecipitation followed by massively

567 parallel DNA sequencing

568 RNA-seq: RNA-sequencing

569 EPA: promoter and extended regions for enhancer-promoter association

570 FR: forward-reverse

571 RF: reverse-forward

572 FF: forward-forward

573 RR: reverse-reverse

574

575 ## Declarations

576 **Ethics approval and consent to participate**

577 I used experimental data in public databases, so this is not applicable.

578

579 **Consent for publication**

580 Not applicable

581

582 **Availability of data and material**

583 I used experimental data in public and commercial databases, so this is not applicable.

584 TRANSFAC, which is a database of transcription factors, is commercial database.

585

586 **Competing interests**

587 The authors declare that they have no competing interests

588

589 **Funding**

590 Publication charges for this article have been funded by JSPS KAKENHI Grant

591 Number 16K00387. This research was partially supported by the Platform Project for

592 Supporting in Drug Discovery and Life Science Research（Platform for Dynamic

593 Approaches to Living System） from Japan Agency for Medical Research and

594 Development (AMED). This research was partially supported by Development of

595 Fundamental Technologies for Diagnosis and Therapy Based upon Epigenome Analysis

596 from Japan Agency for Medical Research and Development (AMED). This work was

597 partially supported by JST CREST Grant Number JPMJCR15G1, Japan.

598

599 **Authors' contributions**

600 NO designed and performed the research

601

602 **Acknowledgements**

603 The supercomputing resource was provided by Human Genome Center of the Institute

604 of Medical Science at the University of Tokyo. Computations were partially performed

605 on the NIG supercomputer at ROIS National Institute of Genetics.

606

607 **Authors' information**

608

609 **References used in Table 4.**

610 Monocytes: IFN-gamma pathway[39], IL3-mediated signaling events[40],
611 GMCSF-mediated signaling events[41], PDGF receptor signaling network[42], VEGF
612 and VEGFR signaling network[43]

613 CD4+ T cell: Integrin family cell surface interactions[44], IFN-gamma pathway[45],
614 LKB1 signaling events[46], TCR signaling in naive CD4+ T cells[47], IL3-mediated
615 signaling events[48]

616 CD20+ B cell: Integrin family cell surface interactions[49], IL5-mediated signaling
617 events[50], Insulin Pathway[51], mTOR signaling pathway[52], Sphingosine
618 1-phosphate (S1P) pathway[53]

619

620 **References used in Table 6.**

621 Monocytes: CXCR4-mediated signaling events[54], GPVI-mediated activation
622 cascade[55], Signaling events mediated by TCPTP[56], p38 MAPK signaling
623 pathway[57], IL4-mediated signaling events[58]

624 CD4+ T cell: CXCR4-mediated signaling events[59], Fc-epsilon receptor I signaling in
625 mast cells[60], JNK signaling in the CD4+ TCR pathway[61], Reelin signaling
626 pathway[62], Semaphorin interactions[63]

627 CD20+ B cell: Syndecan-4-mediated signaling events[64], CXCR4-mediated signaling
628 events[65], EPO signaling pathway[66], Stabilization and expansion of the E-cadherin
629 adherens junction[67], Interleukin-1 signaling[68]
630

631 ## Endnotes

632

633 ## References

634 **Uncategorized References**

635 1. Valledor AF, Borras FE, Cullell-Young M, Celada A: **Transcription factors**
636 **that regulate monocyte/macrophage differentiation**. *Journal of leukocyte*
637 *biology* 1998, **63**(4):405-417.
638 2. Nagamura-Inoue T, Tamura T, Ozato K: **Transcription factors that regulate**
639 **growth and differentiation of myeloid cells**. *International reviews of*
640 *immunology* 2001, **20**(1):83-105.
641 3. Ho IC, Tai TS, Pai SY: **GATA3 and the T-cell lineage: essential functions**
642 **before and after T-helper-2-cell differentiation**. *Nature reviews Immunology*
643 2009, **9**(2):125-135.
644 4. Rothenberg EV: **B cell specification from the genome up**. *Nature immunology*
645 2010, **11**(7):572-574.
646 5. Kurotaki D, Osato N, Nishiyama A, Yamamoto M, Ban T, Sato H, Nakabayashi
647 J, Umehara M, Miyake N, Matsumoto N *et al*: **Essential role of the**
648 **IRF8-KLF4 transcription factor cascade in murine monocyte**
649 **differentiation**. *Blood* 2013, **121**(10):1839-1849.
650 6. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H,
651 Koh JL, Toufighi K, Mostafavi S *et al*: **The genetic landscape of a cell**. *Science*
652 *(New York, NY)* 2010, **327**(5964):425-431.
653 7. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz
654 M, Grouios C, Kazi F, Lopes CT *et al*: **The GeneMANIA prediction server:**

655 **biological network integration for gene prioritization and predicting gene**
656 **function**. *Nucleic acids research* 2010, **38**(Web Server issue):W214-220.
657 8. Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, Wang
658 W, Usaj M, Hanchard J, Lee SD *et al*: **A global genetic interaction network**
659 **maps a wiring diagram of cellular function**. *Science (New York, N Y)* 2016,
660 **353**(6306).
661 9. Marbach D, Costello JC, Kuffner R, Vega NM, Prill RJ, Camacho DM, Allison
662 KR, Kellis M, Collins JJ, Stolovitzky G: **Wisdom of crowds for robust gene**
663 **network inference**. *Nature methods* 2012, **9**(8):796-804.
664 10. Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan
665 ET, Metzler G, Vedenko A, Chen X *et al*: **Diversity and complexity in DNA**
666 **recognition by transcription factors**. *Science (New York, NY)* 2009,
667 **324**(5935):1720-1723.
668 11. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM,
669 Bejerano G: **GREAT improves functional interpretation of cis-regulatory**
670 **regions**. *Nature biotechnology* 2010, **28**(5):495-501.
671 12. Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S,
672 Nagae G, Ishihara K, Mishiro T *et al*: **Cohesin mediates transcriptional**
673 **insulation by CCCTC-binding factor**. *Nature* 2008, **451**(7180):796-801.
674 13. Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin DU, Jung I, Wu H, Zhai Y, Tang Y
675 *et al*: **CRISPR Inversion of CTCF Sites Alters Genome Topology and**
676 **Enhancer/Promoter Function**. *Cell* 2015, **162**(4):900-910.
677 14. Gusmao EG, Allhoff M, Zenke M, Costa IG: **Analysis of computational**
678 **footprinting methods for DNase sequencing experiments**. *Nature methods*
679 2016, **13**(4):303-309.
680 15. Thomas R, Thomas S, Holloway AK, Pollard KS: **Features that define the best**
681 **ChIP-seq peak calling algorithms**. *Brief Bioinform* 2016.
682 16. Gusmao EG, Dieterich C, Zenke M, Costa IG: **Detection of active**
683 **transcription factor binding sites with the combination of DNase**
684 **hypersensitivity and histone modifications**. *Bioinformatics (Oxford, England)*
685 2014, **30**(22):3143-3151.
686 17. He B, Chen C, Teng L, Tan K: **Global view of enhancer-promoter**
687 **interactome in human cells**. *Proc Natl Acad Sci U S A* 2014,

688 **111**(21):E2191-2199.
689 18. Wang D, Rendon A, Wernisch L: **Transcription factor and chromatin features**
690 **predict genes associated with eQTLs**. *Nucleic acids research* 2013,
691 **41**(3):1450-1463.
692 19. Whalen S, Truty RM, Pollard KS: **Enhancer-promoter interactions are**
693 **encoded by complex genomic signatures on looping chromatin**. *Nature*
694 *genetics* 2016, **48**(5):488-496.
695 20. Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L,
696 Wingett SW, Andrews S, Grey W, Ewels PA *et al*: **Mapping long-range**
697 **promoter contacts in human cells with high-resolution capture Hi-C**. *Nature*
698 *genetics* 2015, **47**(6):598-606.
699 21. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, Cairns J,
700 Wingett SW, Varnai C, Thiecke MJ *et al*: **Lineage-Specific Genome**
701 **Architecture Links Enhancers and Non-coding Disease Variants to Target**
702 **Gene Promoters**. *Cell* 2016, **167**(5):1369-1384 e1319.
703 22. Bailey SD, Zhang X, Desai K, Aid M, Corradin O, Cowper-Sal Lari R,
704 Akhtar-Zaidi B, Scacheri PC, Haibe-Kains B, Lupien M: **ZNF143 provides**
705 **sequence specificity to secure chromatin interactions at gene promoters**. *Nat*
706 *Commun* 2015, **2**:6186.
707 23. Wingender E, Dietze P, Karas H, Knuppel R: **TRANSFAC: a database on**
708 **transcription factors and their DNA binding sites**. *Nucleic acids research*
709 1996, **24**(1):238-241.
710 24. Portales-Casamar E, Thongjuea S, Kwon AT, Arenillas D, Zhao X, Valen E,
711 Yusuf D, Lenhard B, Wasserman WW, Sandelin A: **JASPAR 2010: the greatly**
712 **expanded open-access database of transcription factor binding profiles**.
713 *Nucleic acids research* 2010, **38**(Database issue):D105-110.
714 25. Newburger DE, Bulyk ML: **UniPROBE: an online database of protein**
715 **binding microarray data on protein-DNA interactions**. *Nucleic acids*
716 *research* 2009, **37**(Database issue):D77-82.
717 26. Zhao Y, Stormo GD: **Quantitative analysis demonstrates most transcription**
718 **factors require only simple models of specificity**. *Nature biotechnology* 2011,
719 **29**(6):480-483.
720 27. Jolma A, Yan J, Whitington T, Toivonen J, Nitta KR, Rastas P, Morgunova E,

721 Enge M, Taipale M, Wei G *et al*: **DNA-binding specificities of human**
722 **transcription factors**. *Cell* 2013, **152**(1-2):327-339.
723 28. Xie Z, Hu S, Blackshaw S, Zhu H, Qian J: **hPDI: a database of experimental**
724 **human protein-DNA interactions**. *Bioinformatics (Oxford, England)* 2010,
725 **26**(2):287-289.
726 29. Kheradpour P, Kellis M: **Systematic discovery and characterization of**
727 **regulatory motifs in ENCODE TF binding experiments**. *Nucleic acids*
728 *research* 2014, **42**(5):2976-2987.
729 30. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW,
730 Noble WS: **MEME SUITE: tools for motif discovery and searching**. *Nucleic*
731 *acids research* 2009, **37**(Web Server issue):W202-208.
732 31. Grant CE, Bailey TL, Noble WS: **FIMO: scanning for occurrences of a given**
733 **motif**. *Bioinformatics (Oxford, England)* 2011, **27**(7):1017-1018.
734 32. Karolchik D, Barber GP, Casper J, Clawson H, Cline MS, Diekhans M, Dreszer
735 TR, Fujita PA, Guruvadoo L, Haeussler M *et al*: **The UCSC Genome Browser**
736 **database: 2014 update**. *Nucleic acids research* 2014, **42**(Database
737 issue):D764-770.
738 33. Wang L, Wang S, Li W: **RSeQC: quality control of RNA-seq experiments**.
739 *Bioinformatics (Oxford, England)* 2012, **28**(16):2184-2185.
740 34. Saeed S, Quintin J, Kerstens HH, Rao NA, Aghajanirefah A, Matarese F, Cheng
741 SC, Ratter J, Berentsen K, van der Ent MA *et al*: **Epigenetic programming of**
742 **monocyte-to-macrophage differentiation and trained innate immunity**.
743 *Science (New York, NY)* 2014, **345**(6204):1251086.
744 35. Zambon AC, Gaj S, Ho I, Hanspers K, Vranizan K, Evelo CT, Conklin BR, Pico
745 AR, Salomonis N: **GO-Elite: a flexible solution for pathway and ontology**
746 **over-representation**. *Bioinformatics (Oxford, England)* 2012,
747 **28**(16):2209-2210.
748 36. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis
749 G, Durbin R, Genome Project Data Processing S: **The Sequence**
750 **Alignment/Map format and SAMtools**. *Bioinformatics (Oxford, England)*
751 2009, **25**(16):2078-2079.
752 37. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing**
753 **genomic features**. *Bioinformatics (Oxford, England)* 2010, **26**(6):841-842.

754 38. Zang C, Schones DE, Zeng C, Cui K, Zhao K, Peng W: **A clustering approach**
755 **for identification of enriched domains from histone modification ChIP-Seq**
756 **data**. *Bioinformatics (Oxford, England)* 2009, **25**(15):1952-1958.
757 39. Kraaij MD, Vereyken EJ, Leenen PJ, van den Bosch TP, Rezaee F, Betjes MG,
758 Baan CC, Rowshani AT: **Human monocytes produce interferon-gamma upon**
759 **stimulation with LPS**. *Cytokine* 2014, **67**(1):7-12.
760 40. Ebner S, Hofer S, Nguyen VA, Furhapter C, Herold M, Fritsch P, Heufler C,
761 Romani N: **A novel role for IL-3: human monocytes cultured in the presence**
762 **of IL-3 and IL-4 differentiate into dendritic cells that produce less IL-12**
763 **and shift Th cell responses toward a Th2 cytokine pattern**. *J Immunol* 2002,
764 **168**(12):6199-6207.
765 41. Francisco-Cruz A, Aguilar-Santelises M, Ramos-Espinosa O, Mata-Espinosa D,
766 Marquina-Castillo B, Barrios-Payan J, Hernandez-Pando R:
767 **Granulocyte-macrophage colony-stimulating factor: not just another**
768 **haematopoietic growth factor**. *Med Oncol* 2014, **31**(1):774.
769 42. Shaw RJ, Doherty DE, Ritter AG, Benedict SH, Clark RA:
770 **Adherence-dependent increase in human monocyte PDGF(B) mRNA is**
771 **associated with increases in c-fos, c-jun, and EGR2 mRNA**. *J Cell Biol* 1990,
772 **111**(5 Pt 1):2139-2148.
773 43. Heil M, Clauss M, Suzuki K, Buschmann IR, Willuweit A, Fischer S, Schaper
774 W: **Vascular endothelial growth factor (VEGF) stimulates monocyte**
775 **migration through endothelial monolayers via increased integrin expression**.
776 *Eur J Cell Biol* 2000, **79**(11):850-857.
777 44. Hogg N, Laschinger M, Giles K, McDowall A: **T-cell integrins: more than just**
778 **sticking points**. *J Cell Sci* 2003, **116**(Pt 23):4695-4705.
779 45. Ngai P, McCormick S, Small C, Zhang X, Zganiacz A, Aoki N, Xing Z: **Gamma**
780 **interferon responses of CD4 and CD8 T-cell subsets are quantitatively**
781 **different and independent of each other during pulmonary Mycobacterium**
782 **bovis BCG infection**. *Infect Immun* 2007, **75**(5):2244-2252.
783 46. MacIver NJ, Blagih J, Saucillo DC, Tonelli L, Griss T, Rathmell JC, Jones RG:
784 **The liver kinase B1 is a central regulator of T cell development, activation,**
785 **and metabolism**. *J Immunol* 2011, **187**(8):4187-4198.
786 47. Gallegos AM, Xiong H, Leiner IM, Susac B, Glickman MS, Pamer EG, van

787 Heijst JW: **Control of T cell antigen reactivity via programmed TCR**
788 **downregulation**. *Nature immunology* 2016, **17**(4):379-386.
789 48. Kronin V, Hochrein H, Shortman K, Kelso A: **Regulation of T cell cytokine**
790 **production by dendritic cells**. *Immunol Cell Biol* 2000, **78**(3):214-223.
791 49. Lu TT, Cyster JG: **Integrin-mediated long-term B cell retention in the splenic**
792 **marginal zone**. *Science (New York, NY)* 2002, **297**(5580):409-412.
793 50. Baumann MA, Paul CC: **Interleukin-5 and human B lymphocytes**. *Methods*
794 1997, **11**(1):88-97.
795 51. Winer DA, Winer S, Shen L, Wadia PP, Yantha J, Paltser G, Tsui H, Wu P,
796 Davidson MG, Alonso MN *et al*: **B cells promote insulin resistance through**
797 **modulation of T cells and production of pathogenic IgG antibodies**. *Nat Med*
798 2011, **17**(5):610-617.
799 52. Limon JJ, Fruman DA: **Akt and mTOR in B Cell Activation and**
800 **Differentiation**. *Front Immunol* 2012, **3**:228.
801 53. Sorrentino R, Bertolino A, Terlizzi M, Iacono VM, Maiolino P, Cirino G,
802 Roviezzo F, Pinto A: **B cell depletion increases**
803 **sphingosine-1-phosphate-dependent airway inflammation in mice**. *Am J*
804 *Respir Cell Mol Biol* 2015, **52**(5):571-583.
805 54. Caulfield J, Fernandez M, Snetkov V, Lee T, Hawrylowicz C: **CXCR4**
806 **expression on monocytes is up-regulated by dexamethasone and is**
807 **modulated by autologous CD3+ T cells**. *Immunology* 2002, **105**(2):155-162.
808 55. Kral JB, Schrottmaier WC, Salzmann M, Assinger A: **Platelet Interaction with**
809 **Innate Immune Cells**. *Transfus Med Hemother* 2016, **43**(2):78-88.
810 56. Heinonen KM, Bourdeau A, Doody KM, Tremblay ML: **Protein tyrosine**
811 **phosphatases PTP-1B and TC-PTP play nonredundant roles in macrophage**
812 **development and IFN-gamma signaling**. *Proc Natl Acad Sci U S A* 2009,
813 **106**(23):9368-9372.
814 57. Liu HS, Pan CE, Liu QG, Yang W, Liu XM: **Effect of NF-kappaB and p38**
815 **MAPK in activated monocytes/macrophages on pro-inflammatory**
816 **cytokines of rats with acute pancreatitis**. *World J Gastroenterol* 2003,
817 **9**(11):2513-2518.
818 58. Bonder CS, Finlay-Jones JJ, Hart PH: **Interleukin-4 regulation of human**
819 **monocyte and macrophage interleukin-10 and interleukin-12 production.**

820 **Role of a functional interleukin-2 receptor gamma-chain**. *Immunology* 1999,
821 **96**(4):529-536.
822 59. Murakami T, Nakajima T, Koyanagi Y, Tachibana K, Fujii N, Tamamura H,
823 Yoshida N, Waki M, Matsumoto A, Yoshie O *et al*: **A small molecule CXCR4**
824 **inhibitor that blocks T cell line-tropic HIV-1 infection**. *J Exp Med* 1997,
825 **186**(8):1389-1393.
826 60. Koyasu S, D'Adamio L, Arulanandam AR, Abraham S, Clayton LK, Reinherz
827 EL: **T cell receptor complexes containing Fc epsilon RI gamma homodimers**
828 **in lieu of CD3 zeta and CD3 eta components: a novel isoform expressed on**
829 **large granular lymphocytes**. *J Exp Med* 1992, **175**(1):203-209.
830 61. Dong C, Yang DD, Wysk M, Whitmarsh AJ, Davis RJ, Flavell RA: **Defective T**
831 **cell differentiation in the absence of Jnk1**. *Science (New York, NY)* 1998,
832 **282**(5396):2092-2095.
833 62. Huang Y, Clarke F, Karimi M, Roy NH, Williamson EK, Okumura M,
834 Mochizuki K, Chen EJ, Park TJ, Debes GF *et al*: **CRK proteins selectively**
835 **regulate T cell migration into inflamed tissues**. *J Clin Invest* 2015,
836 **125**(3):1019-1032.
837 63. Kumanogoh A, Marukawa S, Suzuki K, Takegahara N, Watanabe C, Ch'ng E,
838 Ishida I, Fujimura H, Sakoda S, Yoshida K *et al*: **Class IV semaphorin Sema4A**
839 **enhances T-cell activation and interacts with Tim-2**. *Nature* 2002,
840 **419**(6907):629-633.
841 64. Endo T, Ito K, Morimoto J, Kanayama M, Ota D, Ikesue M, Kon S, Takahashi D,
842 Onodera T, Iwasaki N *et al*: **Syndecan 4 Regulation of the Development of**
843 **Autoimmune Arthritis in Mice by Modulating B Cell Migration and**
844 **Germinal Center Formation**. *Arthritis Rheumatol* 2015, **67**(9):2512-2522.
845 65. Beider K, Ribakovsky E, Abraham M, Wald H, Weiss L, Rosenberg E, Galun E,
846 Avigdor A, Eizenberg O, Peled A *et al*: **Targeting the CD20 and CXCR4**
847 **pathways in non-hodgkin lymphoma with rituximab and high-affinity**
848 **CXCR4 antagonist BKT140**. *Clin Cancer Res* 2013, **19**(13):3495-3507.
849 66. Kimata H, Yoshida A, Ishioka C, Masuda S, Sasaki R, Mikawa H: **Human**
850 **recombinant erythropoietin directly stimulates B cell immunoglobulin**
851 **production and proliferation in serum-free medium**. *Clin Exp Immunol* 1991,
852 **85**(1):151-156.

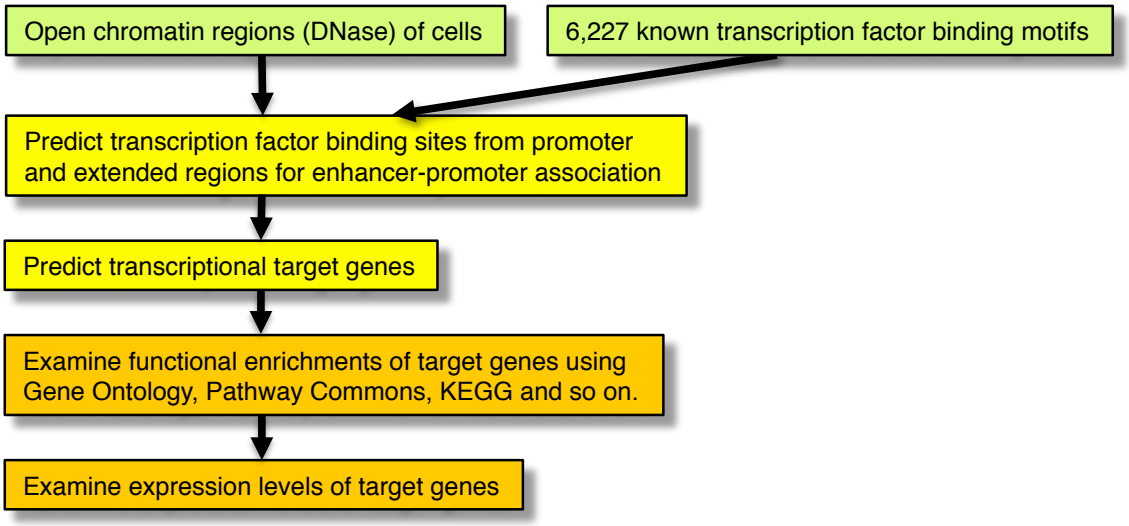
853 67. Koizumi M, Hiasa Y, Kumagi T, Yamanishi H, Azemoto N, Kobata T, Matsuura
854 B, Abe M, Onji M: **Increased B cell-activating factor promotes tumor**
855 **invasion and metastasis in human pancreatic cancer**. *PLoS One* 2013,
856 **8**(8):e71367.
857 68. Rosser EC, Oleinika K, Tonon S, Doyle R, Bosma A, Carter NA, Harris KA,
858 Jones SA, Klein N, Mauri C: **Regulatory B cells are induced by gut**
859 **microbiota-driven interleukin-1beta and interleukin-6 production**. *Nat Med*
860 2014, **20**(11):1334-1339.
861 69. de Wit E, Vos ES, Holwerda SJ, Valdes-Quezada C, Verstegen MJ, Teunissen H,
862 Splinter E, Wijchers PJ, Krijger PH, de Laat W: **CTCF Binding Polarity**
863 **Determines Chromatin Looping**. *Molecular cell* 2015, **60**(4):676-684.

864

865

866 **Figure**

867



868

869

870 **Figure 1. Analyses of functional enrichments of putative transcriptional target**

871 **genes.** Transcriptional target genes were predicted using open chromatin regions

872 (DNase-DGF) and known transcription factor binding sequences. Functional

873 enrichments of target genes were analyzed using 10 annotation databases, and were

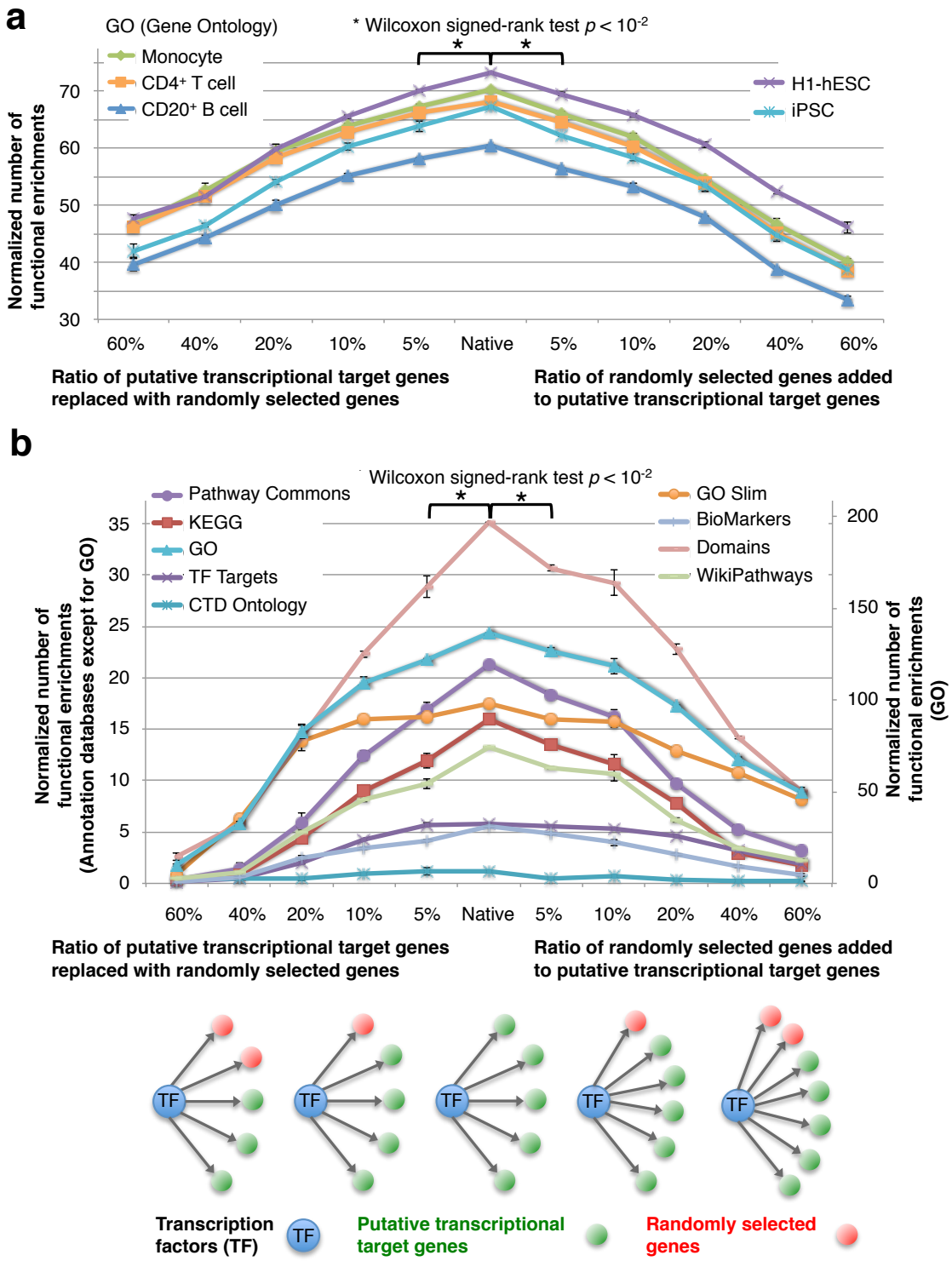
874 changed based on the criteria of promoter and extended regions for enhancer-promoter

875 association (EPA). To compare with the tendency of the normalized numbers of

876 functional enrichments, the median expression levels of target genes were examined

877 using promoter and EPA.

878

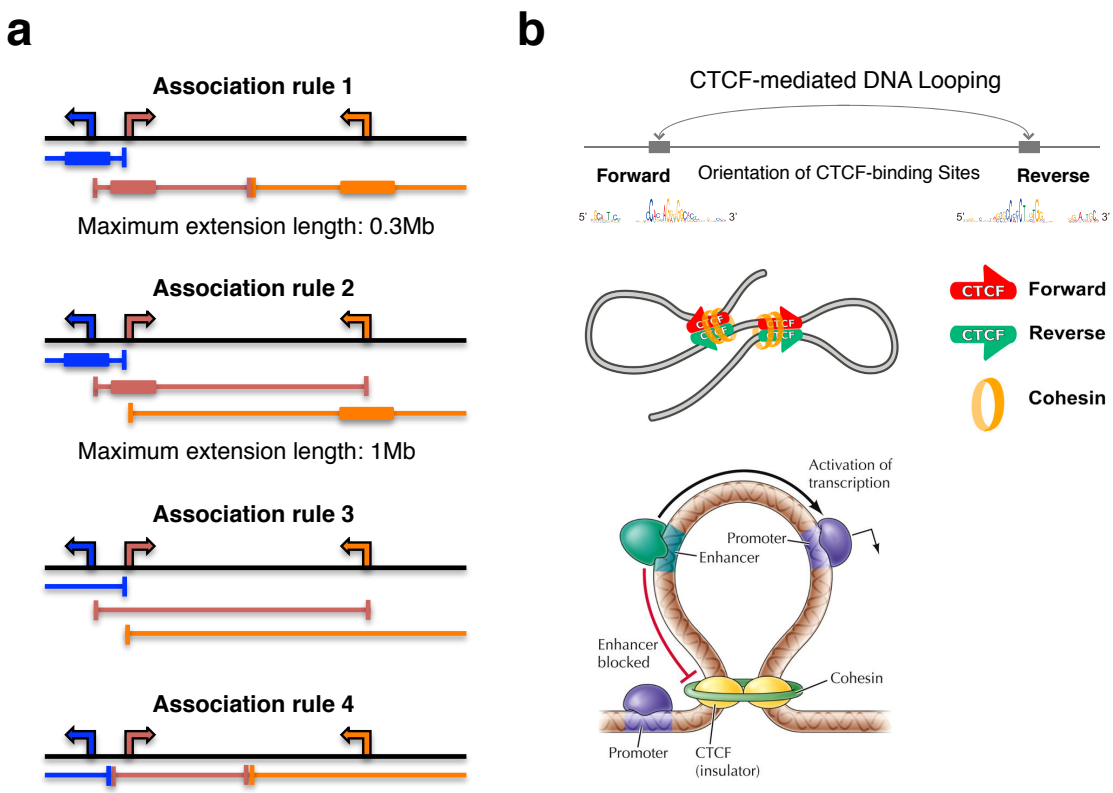

a
GO (Gene Ontology)
* Wilcoxon signed-rank test p < 10-2
Monocyte
CD4+ T cell
CD20+ B cell
H1-hESC
iPSC
Normalized number of functional enrichments
70
60
50
40
30
60%
40%
20%
10%
5%
Native
5%
10%
20%
40%
60%
Ratio of putative transcriptional target genes replaced with randomly selected genes
Ratio of randomly selected genes added to putative transcriptional target genes
b
Wilcoxon signed-rank test p < 10-2
Pathway Commons
KEGG
GO
TF Targets
CTD Ontology
GO Slim
BioMarkers
Domains
WikiPathways
Normalized number of functional enrichments (Annotation databases except for GO)
35
30
25
20
15
10
5
0
200
150
100
50
0
Normalized number of functional enrichments (GO)
60%
40%
20%
10%
5%
Native
5%
10%
20%
40%
60%
Ratio of putative transcriptional target genes replaced with randomly selected genes
Ratio of randomly selected genes added to putative transcriptional target genes
TF
Transcription factors (TF)
Putative transcriptional target genes
Randomly selected genes

879

880

881

882 **Figure 2. Effect of randomly selected genes on functional enrichments.** (**a**) Effect of
883 randomly selected genes on functional enrichments using DNase-DGF data.
884 Transcriptional target genes were predicted using DNase-DGF data in human
885 monocytes, $CD4^+$ T, $CD20^+$ B, other four somatic and two stem cells (H1-hESC and
886 iPSC) (see also Figure S1). To test whether slight changes of transcriptional target genes
887 were reflected in the normalized number of their functional enrichments, the ratio of
888 randomly selected genes in the target genes of each TF was changed between 5% and
889 60%. In the left part of the graphs, randomly selected genes were replaced with the
890 target genes where the total number of target genes was unchanged. In the right part of
891 the graphs, randomly selected genes were added to the target genes where the total
892 number of target genes was increased. The result of Gene Ontology annotation was
893 shown. The results of Pathway Commons and KEGG were shown in Figure S1. Native
894 target genes showed the most functional enrichments in most cell types. (**b**) Effect of
895 randomly selected genes on functional enrichments using ChIP-seq data. Transcriptional
896 target genes were predicted using ChIP-seq data of 19 TF in H1-hESC. The results of
897 nine functional annotation databases were shown and the result of target genes of
898 microRNAs was shown in Figure S2. Native target genes showed the most functional
899 enrichments using most annotation databases except for low frequent functional
900 annotations. Putative transcriptional target genes tend to include similar function of
901 genes.
902

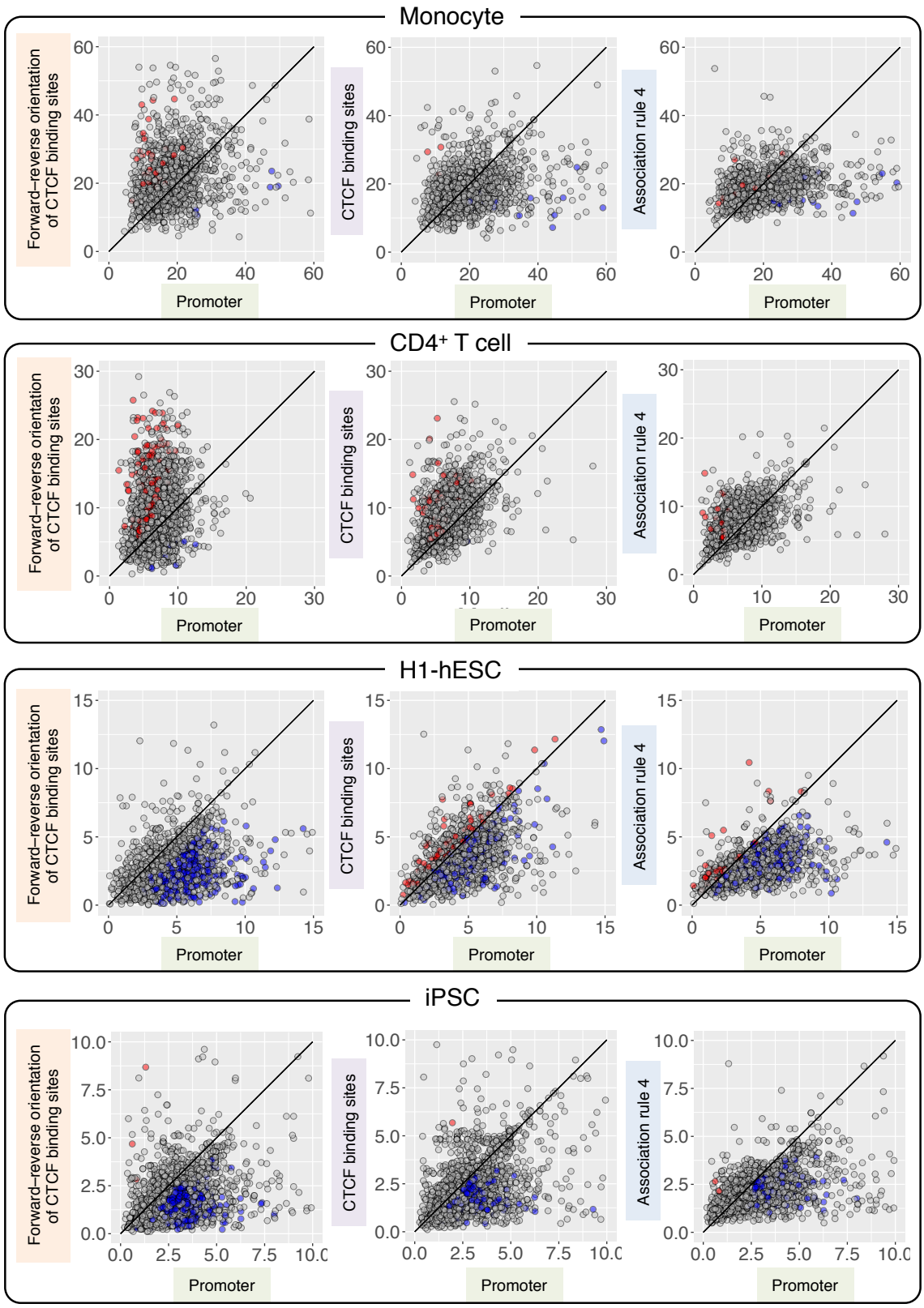


903

904

905 **Figure 3. Criteria of promoter and extended regions for enhancer-promoter**
906 **association and features of chromatin interactions.** (**a**) Computationally-defined
907 regulatory domains [11]. The transcription start site (TSS) of each gene is indicated as
908 an arrow. The corresponding regulatory domain for each gene is shown in a matching
909 color as a bracketed line. The *basal plus extension* association rule assigns a basal
910 regulatory domain to each gene regardless of genes nearby (thick line, Association rule
911 1 and 2) (see Methods). The domain is then extended to the basal regulatory domain of
912 the nearest upstream and downstream genes. The *two nearest genes* association rule
913 extends the regulatory domain to the TSS of the nearest upstream and downstream
914 genes (Association rule 3). The *single nearest gene* association rule extends the

915 regulatory domain to the midpoint between this gene's TSS and the nearest gene's TSS
916 both upstream and downstream (Association rule 4). (**b**) Forward–reverse orientation of
917 CTCF-binding sites are frequently found in chromatin interactions. CTCF can block the
918 interaction between enhancers and promoters limiting the activity of enhancers to
919 certain functional domains. Figures adapted from [13, 69] and the bottom figure in
920 panel b was in the book 'THE CELL: A molecular approach 6e, Figure 7.21, 2013
921 Sinauer Associates, Inc' and Image 68 insulators in
922 'http://slideplayer.com/slide/3836520'.
923


Monocyte
Forward–reverse orientation of CTCF binding sites
CTCF binding sites
Association rule 4
Promoter
CD4+ T cell
H1-hESC
iPSC

926 **Figure 4. Comparison of the median expression levels of transcriptional target**
927 **genes.** The median expression levels of the target genes of the same transcription factor
928 binding sequences were compared between promoters and three types of promoter and
929 extended regions for enhancer-promoter association (EPA). Red and blue dots show
930 statistically significant difference of the distribution of expression levels of target genes
931 between promoters and EPA. Red dots show the median expression level of target genes
932 was higher in EPA than promoters, and blue dots show the median expression level of
933 target genes was lower in EPA than promoters. The median expression levels of putative
934 transcriptional target genes were significantly lower using TF bound in enhancers than
935 using TF bound in promoters in stem cells such as H1-hESC and iPSC, and higher by
936 TF bound in enhancers in immune cells. These results implied that enhancers
937 significantly affect the expression of target genes.
938

Tables

Tables are attached in separate PDF files (Table 1 – 6) in Attachment folder, because the tables are colored to understand them easily and statistically different numbers are shown with lines and asterisks.

Table 1.

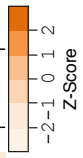
Table 1. Number of functional enrichments and unique functional enrichemnts of putative transcriptional target genes.

Number of functional enrichments of putative trancriptional target genes

	KEGG	TF Targets	CTD Ontology	GO Slim	GO	Pathway Commons	BioMarkers	MicroRNA	Domains	WikiPathways
Monocyte	349	107	209	114	2,902	1,005	42	451	1,202	242
CD4⁺ T cell	317	135	278	77	4,077	1,806	47	754	1,401	405
CD20⁺ B cell	323	103	170	88	2,778	821	39	948	950	288

Number of unique functional enrichments of gene expression information alone and putative transcriptional target genes.
*** Wilcoxon signed-rank test $p < 0.01$ (except for MicroRNA and Domains)**

	KEGG	TF Targets	CTD Ontology	GO Slim	GO	Pathway Commons	BioMarkers	MicroRNA	Domains	WikiPathways
Gene expression information alone										
Monocyte	43	0	35	11	237	101	7	314	404	58
CD4⁺ T cell	47	0	19	9	301	165	9	136	397	81
CD20⁺ B cell	42	0	27	12	239	247	6	370	409	65
Putative transcriptional target genes										
Monocyte	95	16	127	12	1,271	242	17	97	303	105
CD4⁺ T cell	105	26	146	23	1,654	415	24	224	585	133
CD20⁺ B cell	93	23	96	23	1,192	329	16	231	397	106



950 Table 2

Table 2. Functional enrichments of putative transcriptional target genes using Pathway Commons

Monocyte - Pathway Commons	No. of TFs
Proteoglycan syndecan-mediated signaling events	18
Regulation of CDC42 activity	15
LKB1 signaling events	14
Glypican pathway	13
Interferon Signaling	13
Sphingosine 1-phosphate (S1P) pathway	13
IL5-mediated signaling events	12
Syndecan-1-mediated signaling events	12
IL3-mediated signaling events	11
Mitotic Prophase	11
Golgi Cisternae Pericentriolar Stack Reorganization	11
Interferon alpha/beta signaling	11
IFN-gamma pathway	10
Signaling events mediated by Hepatocyte Growth Factor Receptor (c-Met)	10
Recruitment of mitotic centrosome proteins and complexes	10

$CD4^+$ T cell - Pathway Commons	No. of TFs
TCR signaling in naive $CD8^+$ T cells	36
IL12-mediated signaling events	24
Downstream signaling in naive $CD8^+$ T cells	21
TCR signaling in naive $CD4^+$ T cells	21
IL12 signaling mediated by STAT4	20
Validated transcriptional targets of AP1 family members Fra1 and Fra2	19
CXCR4-mediated signaling events	17
ATF-2 transcription factor network	15
Thrombin/protease-activated receptor (PAR) pathway	14
TCR signaling	14
PAR1-mediated thrombin signaling events	14
Downstream TCR signaling	14
Internalization of ErbB1	13
Urokinase-type plasminogen activator (uPA) and uPAR-mediated signaling	13
ErbB receptor signaling network	13

$CD20^+$ B cell - Pathway Commons	No. of TFs
Interferon alpha/beta signaling	12
Alpha6Beta4Integrin	11
Validated targets of C-MYC transcriptional activation	11
IL8- and CXCR2-mediated signaling events	10
Antigen processing-Cross presentation	9
BCR signaling pathway	9
IL6-mediated signaling events	9
Cell junction organization	9
ER-Phagosome pathway	8
Regulation of CDC42 activity	8
CXCR4-mediated signaling events	8
CDC42 signaling events	8
Syndecan-4-mediated signaling events	8
Noncanonical Wnt signaling pathway	7
Class I MHC mediated antigen processing & presentation	7

951

Table 3

Table 3. Normalized number of functional enrichments of putative transcriptional target genes using promoter and extended regions for enhancer-promoter association.
*** Wilcoxon signed-rank test $p < 0.05$**

	KEGG	CTD Ontology	GO Slim	GO	Pathway Commons	Domains	Wiki Pathways	
Monocyte								
Promoter	8.46	5.07	2.76	70.34	24.36	29.13	5.87	
Association rule 1	10.44	6.48	2.01	133.54	44.75	42.19	11.10	
Association rule 2	9.03	6.29	1.45	125.13	38.69	40.17	8.30	*
Association rule 3	8.06	5.29	1.37	106.60	24.25	38.96	7.62	
Association rule 4	11.47	8.22	2.46	164.18	71.42	47.85	12.78	*
CD4⁺ T cell								
Promoter	5.30	4.64	1.29	68.11	30.17	23.41	6.77	
Association rule 1	13.60	7.07	2.74	142.40	84.51	43.78	13.65	
Association rule 2	13.57	6.69	3.05	141.15	86.36	46.33	12.02	*
Association rule 3	12.40	5.89	2.50	115.76	68.84	41.85	10.00	
Association rule 4	16.40	7.86	4.03	177.55	108.08	53.86	16.72	*
CD20⁺ B cell								
Promoter	7.02	3.70	1.91	60.39	17.85	20.65	6.26	
Association rule 1	8.88	6.21	2.59	104.55	59.32	34.34	8.29	
Association rule 2	8.60	5.32	1.55	105.34	57.31	38.05	9.95	*
Association rule 3	9.01	5.28	1.42	88.85	26.49	35.17	8.26	
Association rule 4	9.95	6.62	3.07	134.46	90.99	41.30	10.67	*
HUVEC								
Promoter	7.50	5.74	1.34	79.71	22.21	22.40	4.59	
Association rule 1	11.62	8.95	3.74	160.52	37.11	65.15	8.88	
Association rule 2	11.87	8.77	4.76	176.26	36.86	74.71	8.66	*
Association rule 3	9.84	7.70	4.55	154.06	34.29	73.98	7.74	
Association rule 4	13.46	8.99	5.08	191.73	59.10	82.65	10.13	*
IMR90								
Promoter	6.63	5.21	2.43	68.31	11.64	25.50	5.47	
Association rule 1	9.63	6.78	2.85	122.38	24.89	65.39	8.03	
Association rule 2	9.32	6.56	1.43	115.31	16.36	60.36	7.47	*
Association rule 3	9.59	5.63	1.55	99.10	15.39	54.82	5.82	
Association rule 4	10.55	7.68	3.76	142.10	30.05	74.86	8.78	*
MCF-7								
Promoter	6.97	5.43	1.69	77.88	24.71	25.99	8.77	
Association rule 1	9.45	6.26	2.03	121.84	22.38	55.19	9.86	
Association rule 2	9.44	6.18	0.89	111.77	21.59	53.90	7.97	*
Association rule 3	9.83	5.13	1.12	100.35	18.96	55.30	7.37	
Association rule 4	10.80	7.38	2.77	149.64	31.20	62.40	11.07	*
HMEC								
Promoter	8.76	5.03	1.53	65.80	14.07	26.52	6.70	
Association rule 1	10.88	7.04	1.96	138.27	18.28	72.55	8.08	
Association rule 2	9.22	6.31	1.63	141.57	18.84	73.09	7.99	*
Association rule 3	9.30	6.50	1.68	123.46	17.27	64.67	7.75	
Association rule 4	11.78	7.53	3.48	166.71	23.02	90.11	10.10	*
H1-hESC								
Promoter	8.45	3.47	2.49	73.33	14.12	24.35	4.84	
Association rule 1	9.54	4.82	1.97	113.37	21.86	37.66	6.30	
Association rule 2	10.76	5.82	1.69	123.45	21.27	38.26	5.35	*
Association rule 3	8.00	4.84	1.43	101.12	18.00	33.59	5.62	
Association rule 4	10.56	5.37	3.27	141.38	35.12	46.16	6.73	*
iPSC								
Promoter	8.64	2.05	1.58	67.32	22.75	12.36	6.42	
Association rule 1	10.25	5.64	1.66	107.50	18.35	26.77	8.59	
Association rule 2	11.70	6.20	1.01	101.76	17.00	28.13	8.16	*
Association rule 3	11.69	6.89	1.01	90.85	16.60	24.89	7.76	
Association rule 4	11.39	7.21	2.53	136.46	23.60	31.49	9.27	*

Z-Score: −2, −1, 0, 1, 2

954 Table 4

Table 4. Differences in functional enrichments using Pathway Commons. Immune cell-related functional annotations were enriched more in 'promoter and extended regions for enhancer-promoter association (EPA)' than promoters. Five annotations are shown in each cell type. These functions are confirmed to be related to the cellular functions by reference journal papers.

Cell type	Term of functional annotation	No. of TF EPA (Association rule 4)	No. of TF Promoter	*P*-value	Experimental support
Monocytes	IFN-gamma pathway	51	10	6.89×10^{-4}	Ref. 39
	IL3-mediated signaling events	46	11	4.69×10^{-3}	Ref. 40
	GMCSF-mediated signaling events	43	10	5.25×10^{-3}	Ref. 41
	PDGF receptor signaling network	43	8	1.34×10^{-2}	Ref. 42
	VEGF and VEGFR signaling network	44	9	2.10×10^{-3}	Ref. 43
$CD4^+$ T cell	Integrin family cell surface interactions	112	13	2.59×10^{-12}	Ref. 44
	IFN-gamma pathway	107	12	5.26×10^{-12}	Ref. 45
	LKB1 signaling events	107	11	1.96×10^{-12}	Ref. 46
	TCR signaling in naive $CD4^+$ T cells	103	21	3.98×10^{-8}	Ref. 47
	IL3-mediated signaling events	100	8	9.54×10^{-13}	Ref. 48
$CD20^+$ B cell	Integrin family cell surface interactions	69	2	5.17×10^{-11}	Ref. 49
	IL5-mediated signaling events	64	0	2.21×10^{-11}	Ref. 50
	Insulin Pathway	63	0	3.15×10^{-11}	Ref. 51
	mTOR signaling pathway	63	0	3.15×10^{-11}	Ref. 52
	Sphingosine 1-phosphate (S1P) pathway	63	0	3.15×10^{-11}	Ref. 53

955 Chi-square tests were conducted using the total number of putative transcriptional target genes (Table S9).

956

Table 5

Table. 5 Normalized number of functional enrichments of putative transcriptional target genes using CTCF binding sites. * Wilcoxon signed-rank test $p < 0.05$

	KEGG	CTD Ontology	GO Slim	GO	Pathway Commons	Domains	Wiki Pathways
Monocytes							
Association rule 4	11.47	8.22	2.46	164.18	71.42	47.85	12.78
CTCF (FR+RF+FF+RR)	13.19	10.39	2.74	134.26	34.37	44.46	12.96
CTCF (FR)	42.92	19.53	5.66	509.86	196.58	112.14	35.11
CD4⁺ T cells							
Association rule 4	16.40	7.86	4.03	177.55	108.08	53.86	16.72
CTCF (FR+RF+FF+RR)	26.33	8.73	5.11	206.05	130.71	57.53	23.56
CTCF (FR)	69.39	14.66	24.91	560.44	220.54	133.26	46.54
CD20⁺ B cells							
Association rule 4	9.95	6.62	3.07	134.46	90.99	41.30	10.67
CTCF (FR+RF+FF+RR)	8.78	4.22	2.78	94.13	27.70	28.55	7.08
CTCF (FR)	28.86	9.72	6.61	304.01	220.77	99.89	22.68
HUVEC							
Association rule 4	13.46	8.99	5.08	191.73	59.10	82.65	10.13
CTCF (FR+RF+FF+RR)	19.35	10.59	3.85	155.11	28.36	57.87	13.58
CTCF (FR)	33.71	14.82	16.33	454.47	61.85	163.38	29.38
IMR90							
Association rule 4	10.55	7.68	3.76	142.10	30.05	74.86	8.78
CTCF (FR+RF+FF+RR)	13.88	17.55	9.22	252.24	39.92	123.18	22.45
CTCF (FR)	30.22	15.89	11.34	402.40	53.56	180.71	37.42
MCF-7							
Association rule 4	10.80	7.38	2.77	149.64	31.20	62.40	11.07
CTCF (FR+RF+FF+RR)	11.39	8.30	4.92	234.48	30.40	126.70	13.99
CTCF (FR)	22.34	11.48	6.09	397.69	51.39	145.43	37.58
HMEC							
Association rule 4	11.78	7.53	3.48	166.71	23.02	90.11	10.10
CTCF (FR+RF+FF+RR)	19.68	13.54	4.29	232.47	27.97	106.24	9.84
CTCF (FR)	29.59	14.46	6.97	390.45	43.90	124.81	30.78
H1-hESC							
Association rule 4	10.56	5.37	3.27	141.38	35.12	46.16	6.73
CTCF (FR+RF+FF+RR)	12.95	5.37	2.45	131.46	29.57	44.15	11.08
CTCF (FR)	28.21	10.55	9.02	303.41	56.95	139.44	19.72
iPSC							
Association rule 4	11.39	7.21	2.53	136.46	23.60	31.49	9.27
CTCF (FR+RF+FF+RR)	17.17	2.33	1.69	120.87	25.60	21.06	6.42
CTCF (FR)	36.94	6.09	6.86	274.30	31.37	83.40	16.80

960 Table 6

Table 6. Differences in functional enrichments using Pathway Commons. Immune cell-related functional annotations were enriched more in EPA shortened at forward–reverse orientation of CTCF-binding sites than EPA without CTCF. Five annotations are shown in each cell type. These functions are confirmed to be related to the cellular functions by reference journal papers.

Cell type	Term of functional annotation	No. of TF EPA (Association rule 4) shortened by FR CTCF	No. of TF EPA (Association rule 4) without CTCF	*P*-value	Experimental support
Monocytes	CXCR4-mediated signaling events	33	14	7.81×10^{-12}	Ref. 54
	GPVI-mediated activation cascade	26	3	8.97×10^{-15}	Ref. 55
	Signaling events mediated by TCPTP	23	10	1.46×10^{-8}	Ref. 56
	p38 MAPK signaling pathway	16	8	6.97×10^{-6}	Ref. 57
	IL4-mediated signaling events	11	3	8.80×10^{-6}	Ref. 58
$CD4^+$ T cell	CXCR4-mediated signaling events	181	82	$< 2.20 \times 10^{-16}$	Ref. 59
	Fc-epsilon receptor I signaling in mast cells	85	25	$< 2.20 \times 10^{-16}$	Ref. 60
	JNK signaling in the CD4+ TCR pathway	65	18	$< 2.20 \times 10^{-16}$	Ref. 61
	Reelin signaling pathway	62	26	$< 2.20 \times 10^{-16}$	Ref. 62
	Semaphorin interactions	51	14	$< 2.20 \times 10^{-16}$	Ref. 63
$CD20^+$ B cell	Syndecan-4-mediated signaling events	85	40	$< 2.20 \times 10^{-16}$	Ref. 64
	CXCR4-mediated signaling events	77	15	$< 2.20 \times 10^{-16}$	Ref. 65
	EPO signaling pathway	46	19	4.50×10^{-15}	Ref. 66
	Stabilization and expansion of the E-cadherin adherens junction	32	16	1.06×10^{-9}	Ref. 67
	Interleukin-1 signaling	13	1	3.98×10^{-8}	Ref. 68

961 Chi-square tests were conducted using the total number of putative transcriptional target genes (Table S12).

962