Excessive fetal growth affects HSC quiescence maintenance through epigenetic programming of EGR1 transcriptional network

Running title: Early epigenetic influence on HSC quiescence maintenance

Alexandre Pelletier[1,3,*], Arnaud Carrier[1,3,*], Yong Mei Zhao[2], Mickaël Canouil[1,3], Mehdi Derhourhi[1], Emmanuelle Durand[1], Lionel Berberian-Ferrato[1], John Greally[4], Francine Hughes[5], Philippe Froguel[1,3,6], Amélie Bonnefond[1,3,6], Fabien Delahaye[1,3].

1) Inserm U1283, CNRS UMR 8199, European Genomic Institute for Diabetes, Institut Pasteur de Lille, Lille, France.
2) Department of Obstetrics & Gynecology and Women's Health, Albert Einstein College of Medicine, 1300 Morris Park Ave, Bronx, NY, 10461, USA.
3) Lille University Hospital, University of Lille, Lille, France.
4) Department of Genetics, Albert Einstein College of Medicine, 1301 Morris Park Avenue, Price Building, Room 322, Bronx, New York 10461, USA.
5) Obstetrics & Gynecology and Women's Health, Division of Maternal-Fetal Medicine, Montefiore Medical Center/Albert Einstein College of Medicine, Bronx, NY 10461, USA.
6) Department of Metabolism, Digestion and Reproduction, Imperial College London, London, U.K.

* These authors contributed equally to this work

Corresponding authors:

Fabien Delahaye; Institut Pasteur de Lille-UMR1283, 1 rue du Prof Calmette, 59000 Lille, France ; fabien.delahaye@pasteur-lille.fr; (+33) 320877258

Philippe Froguel; EGID-UMR1283 1 place de Verdun 59045 Lille, France ; philippe.froguel@cnrs.fr; (+33) 374008101

Summary

Fetal development is a critical period to shape stem cell identity and functions. Detrimental environments during this period are associated with epigenetics alteration of hematopoietic stem and progenitor cells (HSPC) with unknown functional impacts. We implemented a single-cell resolution integrative analysis combining epigenomics, transcriptomics, and functional data to elucidate the epigenetic influence associated with excessive fetal growth on HSPCs. We showed that hematopoietic stem cells (HSC) from large for gestational age neonates present a coordinated DNA hypermethylation and decrease expression for genes of the EGR1 transcriptional network including *SOCS3, KLF2*, and *JUNB* known to sustain stem cell quiescence and pluripotency. Furthermore, these changes were associated with a decreased ability for HSCs to stay undifferentiated and a decreased ability to expand in response to stimulation. Taken together, these results show that fetal overgrowth affects hematopoietic stem cells quiescence maintenance program through an epigenetic programming of the EGR1 related transcriptional network.

47 ## Introduction

48 The hematopoietic stem cells (HSC) are involved in essential processes such as
49 angiogenesis, cardiovascular repair, and immunity throughout the entire life.[1, 2] Thus,
50 alterations in HSC ability to self-renew and to adequately produce differentiated
51 progeny have been suggested to contribute to the onset and progression of age-
52 related diseases such as cancer and cardiovascular diseases.[3, 4] Systemic alterations
53 or action of various stressors like aging[5, 6] can result in alteration of HSC destiny, and
54 ultimately HSC's functions. Although HSCs have been extensively studied, the early
55 mechanisms that control their long-term functions are not well understood, in part due
56 to the extensive heterogeneity in phenotypes and behaviors present in HSCs.[7]

57 The perinatal period, a time of rapid growth and cell differentiation, represents a critical
58 window for tissue development. Detrimental environmental exposures during this
59 period can lead to long-term consequences as exemplified by the concept of fetal
60 programming of adult chronic diseases.[8, 9] Extremes of fetal growth, such as large for
61 gestation age (LGA), have been associated with increased susceptibility to obesity,
62 hypertension, and hypertriglyceridemia later in life.[10-14] Even if the number of infants
63 born LGA has increased in the recent years, from less than 1% to 14.9% in developing
64 countries[15], the long-term impact of being born LGA is not completely understood.

65 Fetal growth impacts the number of circulating CD34+ HSCs[16, 17] in human. In mice, it
66 has been shown that maternal high-fat diet limits fetal hematopoietic stem and
67 progenitor cells (HSPC) expansion and repopulation ability while inducing myeloid-
68 biased differentiation.[18] Only few molecular studies have addressed the impact of early
69 deleterious environment on the HSPCs development and function in human. In a
70 previous study, we found a global increase of DNA methylation in cord blood derived

71 CD34+ HSPCs from LGA infants[19], suggesting that epigenetic modifications may play
72 a role in the association between early-life exposures and induce life-long changes
73 within the hematopoietic system. Still, the molecular and functional impacts of these
74 early epigenetic alterations on human HSPCs remained to be elucidated.

75 To identify signalling pathways under epigenetic influences and characterize the
76 impact of early environment on HSPCs functions, we conducted a combined
77 epigenomic, transcriptomic, and functional analysis on human cord blood derived
78 CD34+ HSPCs from either appropriately grown (CTRL) or LGA neonates in both
79 steady state and in response to extrinsic stimuli (stimulated state). We significantly
80 increased our initial DNA methylation dataset[19] and developed novel analytical
81 approach to improve integration of the epigenomic and transcriptomic data. CD34+
82 HSPCs are a relatively heterogeneous cell population[20] and the interaction between
83 DNA methylation and gene expression depends on cell specific genomic context.[21, 22]
84 Therefore, we performed single-cell RNA sequencing analysis. Finally, we challenged
85 HSPCs *in vitro* to measure the impact of early epigenetic programming on stem cell
86 capacities to differentiate and proliferate. We found that the epigenetic programming
87 associated with extreme foetal growth affects gene expression across EGR1
88 associated transcriptional networks impairing HSC maintenance and expansion
89 capacity.

90 Results

91 Extreme fetal growth is associated with hypermethylation of key genes
92 regulating HSPCs proliferation and differentiation

93 In this study, we added 32 new cord-blood derived human CD34+ HSPC samples to
94 the DNA methylation analysis, in addition to our initial DNA methylation dataset
95 (*n*=40).[19] First, we independently retrieved in this new set of data the global DNA
96 hypermethylation initially found in LGA compared to controls (**Supplemental Figure**
97 **1**). Thus, we pooled both data and detected a total of 4815 differentially methylated
98 CpGs (DMC) with 4787 CpGs hypermethylated and 28 CpGs hypomethylated in LGA
99 compared to CTRL (*p*-value<0.001 and |methylation change|>25; **Figure 2A,**
100 **Supplemental Table 1**).

101 To determine which pathways could be affected by these epigenetic changes, we used
102 gene set enrichment analysis (GSEA). However, a major issue in performing DNA
103 methylation gene set testing is how to assign DMCs to specific genes. Thus, we
104 developed a novel methylation gene-score that better reflects the influence of DNA
105 methylation changes on gene expression than considering only CpG methylation
106 change or significance of this change (see Supplemental Method section**,**
107 **Supplemental Figure 2**). For this methylation gene-score, we defined gene-CpG
108 association based on both distance and cell-type specific regulatory information for
109 CD34+ HSPCs and summarized this information at gene level. We obtained for each
110 gene (*n*=24857, **Supplemental Table 2**) a unique methylation gene-score
111 recapitulating any change in DNA methylation associated with LGA condition.

112 This methylation gene-score was used to rank genes and perform GSEA. Using the
113 gene ontology (GO) reference database, we found a significant enrichment for
114 signaling regulating fetal development as well as for key stem cell pathways such as
115 Wnt signaling, cell fate specification, and cell fate commitment pathways (adjusted *p*-
116 value<0.01, **Figure 2B**). Furthermore, considering as a reference the putative

117 susceptibility gene for each clinical traits included in the GWAS catalog[23], we found
118 significant enrichment of the differentially methylated genes (adjusted *p*-value<0.01,
119 **Figure 2C**) for blood related traits (*e.g.* Red blood cell count, Lymphocyte counts,
120 Monocyte count) and for several metabolic parameters (*e.g.* Triglyceride levels, Birth
121 weight, Obesity-related traits) indicating that DNA methylation changes target genes
122 influencing hematopoiesis and metabolism.

123 Finally, to identify regulatory factors likely impacted by these DNA methylation
124 changes, we performed transcription factor (TF) motif analysis using HOMER.
125 Considering proximal regions surrounding each DMCs (±20bp), we found significant
126 enrichment for 23 TF motifs (adjusted *p*-value<0.05, **Figure 2D**). Among them, we
127 found several members of the Kruppel-like factors and specificity protein (KLF/SP)
128 family: KLF14, KLF5, KLF1, KLF6, SP1, SP2, and SP5 that are known to bind to GC-
129 rich regions as well as known to be involved in the regulation of hematopoietic stem
130 cells pluripotency and self-renewal.[24-27] We also found a significant enrichment for
131 EGR1 motif, another key TF controlling proliferation and activation of HSCs[28]
132 suggesting that EGR1 regulatory network may be impacted by early epigenetic
133 programming in LGA neonates.

134 **Epigenetic programming impacts quiescence associated genes in**
135 **hematopoietic stem cell subpopulation**

136 To get further biological insight into the epigenetic reprogramming observed in LGA
137 HSPCs, we performed single-cell transcriptomic analysis in two conditions steady
138 state and stimulated state to challenge quiescence and proliferation. We first created
139 a hematopoietic reference map (*i.e.* hematomap) by integrating CD34+ cells
140 (*n*=18520) from 7 control neonates (**Figure 3A**). Based on cluster specific gene

141 expression, we identified 18 distinct clusters representative of major lineages (Long-
142 Term HSC, HSC, Multi-Potent progenitor, Lymphoid, Myeloid, and Erythroid) of the
143 hematopoietic compartment (**Figure 3B**, **Supplemental Figure 3**). This hematomap
144 was then used as reference to annotate cells though our different conditions to enable
145 lineage specific transcriptomic analysis.

146 We first analyzed differentially expressed genes (DEGs) at steady state. We
147 integrated single-cell transcriptomic data from 6 LGA (*n*=16791 cells) to the 7 CTRL
148 samples describe above. We only detected few DEGs among the lineages (only one
149 DEGs for Erythroid and Myeloid lineages, but none for the other lineages, adjusted *p*-
150 value<0.05 and $|\log_2$Fold Change (FC)$|>0.5$) indicating no major gene expression
151 changes between CTRL and LGA at steady state (**Supplemental Figure 4**).

152 We then compared gene expression between steady state cells and stimulated state
153 cells to identify transcriptomic differences between CTRL and LGA upon stimulation
154 of the CD34+ HSPCs. First, we validated the cell stimulation considering steady state
155 cells and stimulated state cells from the same CTRL samples (*n*=3 samples
156 representing a total of 6776 steady state cells and 1749 stimulated state cells,
157 **Supplemental Figure 5A, Supplemental Figure 5B**). We identified a consistent
158 response to stimulation with 1518 differentially expressed genes associated to the
159 stimulated state (stimulation signature gene set, adjusted *p*-value<0.05 and
160 $|\log_2 FC|>0.5$; **Supplemental Table 3**). Using GO Biological Process reference, we
161 demonstrated that the upregulated genes from this signature (*n*=1075) were enriched
162 in key pathways regulating HSPCs differentiation as well as response to stress and
163 extrinsic stimulation (adjusted *p*-value<0.05; **Supplemental Figure 5C,**
164 **Supplemental Table 4**). Similarly, using KEGG reference, we identified MAPK, FoxO,

165 TNF, and NF-kappa B signaling pathways known to be involved in the immediate-early
166 response process, further validating our stimulation protocol (adjusted *p*-value<0.05;
167 **Supplemental Figure 5C, Supplemental Table 4**). Stimulation was further validated
168 looking at cell cycle with increase in cell proportion in S and G2M phases
169 (**Supplemental Figure 5D)**. We then performed differential expression analysis
170 across lineages comparing steady state and stimulated state HSPCs and we found a
171 significant enrichment in DEGs in the HSC population compared to the other lineages
172 (*p*-value<2.2×10^{-16}, Chi-Squared Test; **Supplemental Figure 5E**). These findings are
173 consistent with a response to stimulation associated to a transition from quiescence
174 to proliferation in HSCs.

175 Then, we extended our analysis by integrating single-cell transcriptomic data from 6
176 LGA and 8 CTRL stimulated HSPC samples representing 6861 and 5823 cells,
177 respectively, and conducted a lineage specific analysis. Focusing on the 1518 DEGs
178 previously identified in response to stimulation, we observed a global similar pattern
179 of change in expression in both LGA and CTRL comparing steady state and stimulated
180 state within each group especially in HSCs (rho=0.7 and *p*-value<2.2×10^{-16},
181 Spearman's rank correlation; **Figure 4A; Supplemental Figure 6**). However, when
182 comparing stimulated LGA HSC to stimulated CTRL HSC, we observed in stimulated
183 LGA a shift toward downregulated genes (*n*=285 downregulated genes over 373
184 DEGs, adjusted *p*-value<0.05 and $\log_2$FC<(-0.5), **Figure 4A; Supplemental Table 5**),
185 especially for key genes of the immediate early response promoting quiescence and
186 self-renewal such as *SOCS3*, *EGR1*, *KLF2*, *DUSP2, JUNB*, *ID1*, *PLK3,* and *ZFP36*
187 (**Figure 4B).** This shift toward downregulation of expression was also observed at
188 pathway level with MAPK signaling pathway and TNF signaling pathway being now
189 enriched in downregulated genes (adjusted *p*-value<0.05; **Supplemental Figure 7**).

190 Taken together, these observations suggest that the LGA condition is associated with
191 an alteration of the stem cell response to stimulation at molecular level especially in
192 HSC lineage.

193 We then analyzed the correlation between DNA methylation and gene expression
194 changes using the methylation gene-score to compare DNA methylation and
195 expression at similar gene-level resolution. For expression data, we considered gene
196 expression fold change in stimulated state (stimulated LGA vs stimulated CTRL) as
197 no differences were found in steady state. DEGs also found in the stimulation
198 signature have a higher methylation gene-score (p-value<0.001) compared to other
199 DEGs or non-DEG genes (**Figure 4C**). More specifically, the downregulated genes
200 involved in the regulation of proliferation and differentiation of HSPCs such as *SOCS3*,
201 *PLK3*, *EGR1*, *KLF2, ZFP36*, *DUSP2,* and *JUNB* were hypermethylated according to
202 their methylation gene-score (**Figure 4D**) indicating that early epigenetic programming
203 observed in LGA was able to modulate expression of target genes regulating stem cell
204 activation, proliferation, and differentiation processes.

205 **Epigenetic programming affects key regulons essential to HSCs quiescence**
206 **and differentiation**

207 A challenge when modeling epigenetic influences on gene expression is the
208 identification of upstream regulators. Transcription factors (TF) have been often
209 suggested as such, with key role in modulating gene expression as well as
210 demonstrated susceptibility to DNA methylation.[29] TFs are also essential in regulating
211 HSCs function and differentiation with well-established lineage specific TFs.[30] We
212 therefore performed coregulatory network analysis to identify co-regulated genes, *i.e.*
213 regulons, and performed analysis of cis-regulatory motif to assign TF to each regulon

214 (SCENIC). We observed a lineage specific pattern of regulons associated with key
215 hematopoietic TFs such as SPI1, GATA1, GATA2, GATA3, MEIS1, TAL1, TCF3,
216 EGR1, CEBPB, HOXB4, and STAT1/3 (**Supplemental Figure 8, Supplemental**
217 **Table 6**).

218 To identify master regulators associated with the changes in gene expression
219 observed in stimulated LGA HSCs, we used the AUCell algorithm proposed by
220 SCENIC to score the activity of the entire regulon in each cell. We observed six
221 regulons with a significant decrease in activity in the stimulated LGA HSC population
222 (adjusted *p*-value<0.001 and |activity score difference|>0.04, **Supplemental Table 7**).
223 These regulons were associated to ARID5A, EGR1, KLF2, KLF4, FOSB, and JUN
224 consistent with the hypothesis that regulatory mechanisms underlying proliferation and
225 differentiation of HSCs were affected (**Figure 5A**).

226 We then assessed if the correlation observed between DNA methylation and
227 expression was preserved at the regulon level. Thus, we performed GSEA using our
228 list of regulons as reference, to first identify regulons enriched for differentially
229 methylated genes. We ranked genes by their methylation gene-score comparing LGA
230 vs CTRL samples. Among the enriched regulons (adjusted *p*-value<0.01 and
231 normalized enrichment score >1.6, *n*=33), we found the regulons associated with
232 ARID5A, EGR1, FOSB, JUN, KLF2, and KLF4. We also found enrichment for key
233 HSPC specific regulons such as ATF3 and ATF4 regulating HSC stress response;[31,
234 32] CEBPA and SPI1 promoting myeloid differentiation;[33] and HOX family (HOXA9,
235 HOXA10, HOXB4) promoting HSC expansion.[34-36] Interestingly, we identified master
236 functional modules (*i.e.*, network of co-regulated regulons) considering the top
237 enriched regulons for hypermethylated genes (*n*=33; **Figure 5B**). The principal module

238 included ARID5A, EGR1, FOSB, JUN, JUNB, KLF2, KLF4, and KLF10 highlighting
239 how these key regulons are co-regulated and altered in LGA. We then identified
240 another module based on HOXA9, HOXA10, and HOXB4. The last module including
241 FOXJ3, FOXO1, FOXP1, IRF8 and XBP1 is more involved in proliferation and lineage
242 specific differentiation process.

243 Furthermore, considering only the regulons enriched for differentially methylated
244 genes (*n*=33), we looked for enrichment on DEGs. We observed that our top affected
245 regulons ARID5A, JUN, FOSB, KLF4, KLF2 and EGR1 were enriched in DMCs and
246 in downregulated genes (adjusted *p*-value<0.01, **Figure 5C**). Looking at a gene
247 regulatory network constructed based on the top affected regulons (**Figure 5D**), we
248 confirmed that *SOCS3*, *ID1*, *DUSP2* and *ZFP36,* genes previously identified as
249 downregulated in stimulated LGA (**Figure 4B**), were targets of these affected
250 regulons. Interestingly, SPI1 and XPB1 regulons associated with HSPC expansion
251 were enriched in DMCs and in upregulated genes (adjusted *p*-value<0.01, **Figure 5C**).
252 Together, these results confirm at regulon level, the correlation between change in
253 expression and change in DNA methylation observed at gene level. They also
254 demonstrate an alteration of signaling pathways involved in the balance between stem
255 cell quiescence and differentiation, notably through programming of the early
256 immediate response regulatory network, in LGA.

257 **Epigenetic programming is associated with a HSC shift toward more**
258 **differentiated cells**

259 HSPCs represent a population of cells from progenitors to progressively restricted
260 cells of the erythroid, myeloid, or lymphoid lineages. To follow cell distribution through
261 these levels of differentiation and assess the influence of the LGA environment we

262 used Monocle.[37] Monocle generates a pseudotime, *i.e.* a measure that reflects how
263 far an individual cell is in the differentiation process. Collecting these pseudotimes
264 across our different cell populations, we observed a positive correlation between
265 pseudotime and lineage differentiation as expected (r=0.99, Pearson correlation,
266 **Figure 6A**, **Supplemental Figure 9A**). We then compared the distribution of the
267 pseudotime between LGA and CTRL using long-term HSCs as roots. At steady state,
268 no major differences were observed (**Supplemental Figure 9B)**. However, upon
269 stimulation, we observed at population level an increase in pseudotime in stimulated
270 LGA (p-value<0.001), suggesting that in response to stimulation LGA cells fail to
271 maintain their pool of progenitor cells. Indeed, we observed a decrease in number of
272 cells presenting pseudotime associated with the HSC state in stimulated LGA (p-
273 value<0.05) and a shift toward cells presenting elevated pseudotime suggesting that
274 stimulated LGA HSCs leave quiescence but failed to properly self-renew compared to
275 stimulated CTRL HSCs (**Figure 6B**).

276 **Epigenetic programming is associated with alteration of the hematopoietic**
277 **compartment function and subpopulation distribution**

278 To further assess the influence of LGA exposure on stem compartment homeostasis,
279 we monitored cell populations distribution across conditions at molecular resolution
280 using our single-cell expression dataset. At steady state, no significant changes in
281 lineage distribution were observed (**Supplemental Figure 10**). After stimulation, we
282 observed a decrease in HSC cells (p-value=0.015) and a trend toward increased MPP
283 cells (p-value=0.13, **Figure 6C**). Such a shift in cell populations is likely to reflect the
284 altered balance between quiescence and differentiation in LGA HSCs.

285 Finally, we tested *in vitro* differentiation and proliferation capacities of HSPCs using
286 colony forming unit (CFU) assays on 4 CTRL and 4 LGA samples. After 14 days of
287 expansion, colonies were classified in 3 categories: common myeloid progenitors
288 (CFU_GEMM), erythroid progenitors (BFU-E), and granulocyte-macrophage
289 progenitors (CFU_GM) based on the morphology of each colony. We observed a
290 significant decrease in the number of common myeloid progenitors in LGA samples
291 (p-value<0.05; **Figure 6D**) as well as striking differences in shape and size of more
292 differentiated colonies (**Figure 6E**) validating that HSPCs from LGA were less able to
293 expand and confirming the functional impact of the early epigenetic programming.

294 Discussion

295 Hematopoietic stem cells differentiation and self-renew rely on a synergic interplay
296 between genetically encoded signaling as well as cell-intrinsic and cell-extrinsic
297 factors. The integration of these inputs toward a coherent lineage specific
298 differentiation program is thought to rely on epigenetic modifiers.[38] In our previous
299 study[19], we considered bulk DNA methylation and we suggested an alteration of this
300 program in response to extreme fetal growth. Here, our integrative analysis combining
301 single-cell expression, DNA methylation, and *in vitro* data allowed us to build, at cell
302 subpopulation resolution, a model around EGR1 associated networks in which early
303 epigenetic modifications are linked to defective HSC self-renewal and differentiation
304 capacities in neonates exposed to excessive fetal growth (**Figure 7**). This model,
305 based on human samples, demonstrated how early epigenetic programming
306 influences HSC response to further extrinsic stimulation with consequences toward
307 quiescence maintenance. It presents a possible mechanism to explain the association
308 found between LGA and increased susceptibility to adult chronic diseases. Indeed,

309 fine-tuning of HSC quiescence mechanisms is of crucial relevance for correct
310 hematopoiesis. While, not responsive dormant HSCs would lead to hematopoietic
311 failure due to a lack of differentiated blood cells. Highly responsive HSCs, as identified
312 in LGA, would get to exhaustion of the population and lack of long-term maintenance
313 of the hematopoietic system.[39]

314 We demonstrated correlated increase of DNA methylation and decrease of gene
315 expression associated with extreme fetal growth, targeting interconnected regulons
316 under the influence of key transcription factors known to regulate HSC self-renew and
317 differentiation ARID5A, EGR1, FOSB, JUN, JUNB, KLF2, and KLF4. ARID5A, a
318 quiescence associated transcription factor[40] interacts with STAT3 and IL6[41, 42], two
319 key regulators of HSC self-renewal and proliferation, respectively.[43, 44] EGR1, FOSB,
320 and JUN are member of the immediate-early response transcription factor family with
321 essential roles in stress response and in differentiation[45]. EGR1 promotes quiescence
322 in HSCs.[28] JUNB regulates myelopoiesis[46], HSC proliferation and differentiation.[47] A
323 recent study in human has demonstrated that downregulation of *JUN* promotes HSC
324 expansion.[48] FOS members are thought to be gatekeepers to HSC mitotic entry[49] with
325 *FOSB* being downregulated in highly proliferative HSCs.[50] Finally, KLF family is
326 implicated in key stem cell functions. KLF4 is the most well-known factor of this family
327 due to its role in reprogramming somatic cells into induced pluripotent stem cells
328 suggesting a function in preservation of stemness.[51] Thus, our approach supports a
329 model in which the influence of early exposure to deleterious environment (LGA) on
330 HSPCs is mediated through an epigenetically programmed down regulation of key
331 signaling pathways involved in HSC quiescence maintenance.

332 Interestingly, ARID5A, EGR1, and KLFs are known to interact with epigenetic
333 regulators. ARID5A is a subunit of the SWI/SNF family of chromatin-remodeling
334 complexes implicated in the control of cell differentiation and lineage specification.
335 EGR1 forms a complex with the DNA methyltransferase 3 responsible for targeted *de*
336 *novo* methylation.[52] While it was reported that TET2 binds to KLF4 through protein-to-
337 protein interaction to drive locus-specific demethylation during reprogramming of B
338 cells into iPSCs.[53] These observations further support the interaction between
339 epigenetic modifications and regulatory activities of the transcription factors founds in
340 our study.

341 We then challenged differentiation and proliferation of HSPCs *in vitro* and
342 demonstrated that the LGA associated epigenetic programming was linked to an
343 alteration of HSC preservation and expansion. Even if we did not directly measure
344 HSC quiescence, the study of the cell populations at molecular level reveals a
345 decrease in HSC population associated with an increase of multipotent progenitor
346 population suggesting that upon stimulation, HSCs are prone to leave quiescence
347 toward differentiation failing to properly maintain the initial pool of HSCs. Alterations
348 at progenitor level were further confirm using lineage specific cytokine exposure (CFU
349 assays). The decreased in common myeloid progenitors (CMPs) colonies observed in
350 LGA may be associated with a decrease in the initial HSC cells proportion or with an
351 alteration of their capacity to respond to stimulation leading to decreased proliferation
352 as CMPs originate from HSCs. These findings corroborate previous studies on
353 developmental programming of the hematopoietic system.[16, 17] A reduction in self-
354 renewal of HSPCs and increase differentiation in both lymphoid and myeloid lineages
355 have been observed in a mouse model of maternal obesity.[18] These effects may drive
356 long-term consequences in human health as illustrated by the study performed by

357 Kotowski et *al.* in which the integrity of the hematopoietic system in neonates was
358 associated with susceptibility to onset of hematopoietic pathologies, that can be further
359 associated to age-related diseases.[54]

360 Interestingly, the LGA associated epigenetic influence at the transcriptomic and
361 functional levels was exacerbated upon stimulation of the HSC compartment. Indeed,
362 in the absence of environmental stimuli (steady state), no major difference was
363 detected at transcriptomic level as HSCs linger in a quiescent state. However, under
364 extrinsic stimulation (stimulated state), HSCs adaptative response was altered leading
365 to a decrease in early progenitors' pool likely to impact long-term system function. This
366 observation fits with the concept of developmental programming which relies not only
367 on early impairment of organ development but most importantly on a decreased
368 adaptability to further environmental challenges that will enhance disease
369 susceptibility and highlights that an interaction between epigenetic programming and
370 further environmental factors is needed to trigger disease.

371 In summary, we provided a comprehensive model recapitulating the epigenetic early
372 programming of the hematopoietic system and its influence on the hematopoietic stem
373 cell fitness in response to stimulation. This study by identifying epigenetic
374 modifications associated with gene expression and functional alterations contribute to
375 elucidate how early exposure can affect long-term tissue maintenance and possibly
376 increase risk to chronic diseases.

377 Acknowledgments

378 The authors thank the UMR 8199 LIGAN-PM Genomics platform (Lille, France) which
379 belongs to the 'Federation de Recherche' 3508 Labex EGID (European Genomics

380 Institute for Diabetes; ANR-10-LABX-46) and was supported by the ANR Equipex
381 2010 session (ANR-10-EQPX-07-01; 'LIGAN-PM'). The LIGAN-PM Genomics
382 platform (Lille, France) is also supported by the FEDER and the Region Nord-Pas-de-
383 Calais-Picardie. This project is cofunded in the frame of CPER CTRL program by the
384 European Union - European Regional Development Fund (ERDF), Hauts de France
385 Region (contract n°17003781), Métropole Européenne de Lille (contract
386 n°2016_ESR_05), and French State (contract n°2017-R3-CTRL-Phase 1). The
387 present work was also supported by the National Center for Precision Diabetic
388 Medicine – PreciDIAB, which is jointly supported by the French National Agency for
389 Research (ANR-18-IBHU-0001), by the European Union (FEDER), by the Hauts-de-
390 France Regional Council and by the European Metropolis of Lille (MEL) and by the
391 European Research Council (ERC Reg-Seq – 715575).
392 Support for this project was provided by the Roadmap Epigenomics Program, R01
393 HD063791 (Einstein/Greally). Support was also provided by Einstein's Center for
394 Epigenomics, including the Epigenomics Shared Facility and Computational
395 Epigenomics Group.

396 Author contributions

397 AP, AC, YMZ, ED, LBF, and FD were responsible for conducting research and
398 analyzing data. MD and MC provided feedback on the data analysis. AP, AC, FD, AB,
399 and PF contributed to writing the manuscript. FD, FH, and JG were responsible for
400 designing the study.

401 Competing Interest

402 The authors declare no competing financial interests in relation to the work
403 described.

404

405

406 ## Figure Legend

407 **Figure 1: Overview of the experimental design.**

408

409 **Figure 2: LGA is associated with DNA hypermethylation targeting key stem cell**
410 **signaling pathways.** (A) Volcano plot of DNA methylation score differences for LGA
411 compared to CTRL based on 754931 loci throughout the genome. Differentially
412 methylated loci with *p*-value <0.001 and |methylation difference| > 25 are shown in
413 red. (B) Network representation of GO pathways. Significantly enriched pathways
414 were identify using GSEA based on the methylation gene-score. Edges represent
415 interactions between pathways. (C) Gene-set enrichment analysis for susceptibility
416 genes associated to pathophysiological traits based on GWAS catalog. Dot plot
417 representing enrichment for selected relevant traits. Gene ratio corresponds to the
418 ratio between the total number of genes associated to the trait (using GWAS catalog)
419 and the number of differentially methylated genes. (D) Dot plot representing
420 enrichment for TF binding motifs using HOMER considering a ±20bp region around
421 DMCs. Dots are color coded based on significance of the enrichment and y-axis
422 represent the number of regions with binding motif among DMCs.

423

424 **Figure 3: HSPCs lineage annotation using single-cell RNAseq.** (A) Hematomap,
425 UMAP representation of distinct HSPC lineages. (B) Dot plot representing key markers

426 used to annotate cell populations. LT-HSC, long-term hematopoietic stem cell; HSC,
427 hematopoietic stem cell; MPP, multipotent progenitor; LMPP, lymphoid-primed
428 multipotent progenitors; Erythro-Mas, erythroid and mast precursor; Mk/Er,
429 megakaryocyte and erythrocyte; DC, dendritic cell.
430
431 **Figure 4: Response to stimulation is altered in LGA.** (A) Heatmap representing
432 change in expression for stimulation associated DEG (*n*=1518) in HSC population. (B)
433 Volcano plot representing DEG comparing LGA vs CTRL in stimulated state.
434 Differentially expressed genes with adjusted *p*-value <0.05 and $|\log_2 FC|$ >0.5 are
435 shown in red if also part of the stimulation associated DEGs, otherwise in pink. (C)
436 Boxplots representing methylation gene-score distribution associated to non DEG,
437 DEG and stimulation associated DEG (*p*-value< 0.001, Wilcoxon test). (D) Scatter plot
438 representing the correlation between change in expression and change in methylation.
439 x-axis represents the $\log_2 FC$ of expression in stimulated HSCs and y-axis represents
440 the percentile of methylation gene-score comparing LGA vs CTRL. Differentially
441 expressed gene with adjusted *p*-value <0.05 and $|\log_2 FC|$ >0.5 are shown in red if also
442 part of the stimulation associated DEGs, otherwise in pink.
443
444 **Figure 5: Epigenetic programming of HSC specific regulons altered in LGA**
445 **neonates.** Regulons and TF target information were obtained through the SCENIC
446 workflow. (A) Boxplots representing regulon activity score in stimulated CTRL and
447 stimulated LGA HSC lineage. Barplot representing the change in regulon activity and
448 significance comparing stimulated LGA vs stimulated CTRL. Only significantly affected
449 regulons are represented (adjusted *p*-value <0.001 and |activity score difference|
450 >0.04). (B) Network analysis representing regulons interactions considering regulons

451 significantly enriched for differentially methylated genes (adjusted *p*-value<0.01,
452 GSEA). Edges represent overlap between regulons (a minimum of 20% overlap is
453 required to connect regulons). (C) Dot plot representing enrichment for DEGs in
454 regulons enriched for differentially methylated genes. (D) Network representing
455 interactions between target genes and transcription factors considering our top
456 affected regulons ARID5A, EGR1, KLF2, KLF4, FOSB and JUN. Each dot represents
457 a gene within the network, triangle represents transcription factor, arrow represents
458 interaction between transcription factor and target genes, shapes are color coded to
459 reflect the change in methylation gene-score, and DEGs are labeled in red. Size of the
460 shape represents the number of interactions. Only genes with 2 or more interactions
461 are represented.
462
463 **Figure 6: LGA is associated with a HSC shift toward more differentiated cells**
464 **and decrease expansion capacity.** (A) UMAPs representing pseudotimes across
465 lineages. (B) Box plots representing the cumulative percentage of cells per
466 pseudotime in stimulated CTRL and stimulated LGA. Boxplots in the vignette
467 represent overall pseudotime distribution in stimulated CTRL and stimulated LGA.
468 Density plots correspond to cell populations distribution across pseudotimes. (C) Dot
469 plot representing change in cell distribution comparing stimulated LGA vs stimulated
470 CTRL. Dot sizes represent for each lineage the percent distribution of total distribution.
471 Boxplots representing the cell distribution across hematopoietic main lineages in
472 stimulated CTRL and stimulated LGA. (D) Bar plot representing colonies distribution
473 after CFU assays. (E) Representative capture of colonies morphological differences
474 found in CTRL and LGA. LT-HSC, long-term hematopoietic stem cell; HSC,
475 hematopoietic stem cell; MPP, multipotent progenitor; LMPP, lymphoid-primed

476 multipotent progenitors; Erythro-Mas, erythroid and mast precursor; Mk/Er,
477 megakaryocyte and erythrocyte; DC, dendritic cell; CFU-GEMM, common myeloid
478 progenitors; BFU-E, erythroid progenitors; CFU-GM, granulocyte-macrophage
479 progenitors.
480
481 **Figure 7: Model recapitulating the epigenomic, transcriptomic, and functional**
482 **influence of extreme fetal growth on HSCs.** HSC, hematopoietic stem cell; MPP,
483 multipotent progenitor.
484

485 ## Methods

486 See the Supplemental Methods for additional information. Overview of the study
487 design is presented in Figure 1.

488 **Ethics approval.** This study was approved by the Institutional Review Board of the
489 Montefiore Medical Center and the Committee on Clinical Investigation at the Albert
490 Einstein College of Medicine and is in accordance with Health Insurance Portability
491 and Accountability Act regulations. Written informed consent was obtained from all
492 subjects before participation.

493 **Clinical sample collection.** Cord blood samples were obtained from CTRL and LGA
494 neonates. LGA were defined by birth weight and ponderal index values greater than
495 the 90th percentile adjusted for gestational age and sex. Control infants had normal
496 parameters (between 10th and 90th percentiles) for both birth weight and ponderal
497 index. Maternal and infant characteristics are shown in Supplemental Table 8.

498 **Isolation of CD34+ HSPCs.** Mononuclear cells were separated using PrepaCyte-
499 WBC following which CD34+ cells were obtained by positive immunomagnetic bead
500 selection, using the AutoMACS Separator (Miltenyi Biotech). Cells were
501 cryopreserved in 10% dimethyl sulfoxide using controlled rate freezing upon analysis.

502 **Genome-Wide DNA methylation assay.** DNA methylation levels for >1.7M CpGs
503 were obtained using the HELP-tagging assay as previously described.[55]

504 **Single-cell RNA sequencing libraries preparation.** After cell count and viability
505 check, cell suspension was loaded into the Chromium 10x Genomics controller and
506 library was generated using the chromium single-cell v3 chemistry following
507 manufacturer recommendations. Gene expression library was sequenced using
508 100bp paired-end reads on the Illumina NovaSeq 6000 system.

509 **HSPCs extrinsic stimulation.** After cell counting and viability check and prior to cell
510 suspension loading on the Chromium 10x Genomics controller, cell hashtag (HTO)
511 staining was used following the cell-hashing protocol.[56] This protocol resulting in
512 extended cell manipulation was used to mimic response to extrinsic stimulation.
513 Response to stimulation was validated at transcriptomic level.

514 **Colony Forming Unit Assay.** To assess clonogenic progenitor frequencies, 3×10^4
515 CD34+ HSPC cells were plated in methylcellulose containing SCF, GM-CSF, IL-3, and
516 EPO (H4434; STEMCELL Technologies). Colonies were scored 14 days later.

517 **Data processing and statistical analysis.** For DNA methylation analysis, low quality
518 CpGs were filtered out based on detection rate and confidence score. 754931 out of
519 1709224 CpGs were conserved for further analysis. Linear regression and statistical
520 modeling using LIMMA R package[57] were used to identify differentially methylated

521 CpGs (DMC). To link gene to CpGs and performed gene set enrichment analysis
522 (GSEA), we generated a methylation gene-score; first by scoring association between
523 individual CpG and unique gene, based on evidence for association between gene
524 and CpG (distance, eQTL) and on belonging in regulatory regions; and then, by
525 concatenating CpG scores across each gene. We assessed enrichment for biological
526 pathways performing GSEA using the ClusterProfiler package.[58] We performed
527 transcription factor (TF) motif enrichment analysis using HOMER tool[59] considering a
528 20bp region around the DMCs.

529 For single-cell RNAseq (sc-RNAseq) analysis, data were preprocessed using the
530 CellRanger count pipeline. Data filtering, normalization and integration as well as
531 cluster identifications were performed using Seurat (v4) pipeline. Pseudo-bulk
532 differential expression analysis between LGA and CTRL cells within each
533 hematopoietic lineage was performed using DESeq2 R package.[60] Over
534 representation test was performed on differentially expressed genes (DEGs) using
535 enrichGO and enrichKEGG of the ClusterProfiler Package. The SCENIC workflow[61]
536 was used to identify co-regulated modules associated to a TF (regulons) and to
537 generate cell-specific activity score for each regulon. Differentiation trajectory analysis
538 and pseudotime attribution were conducted with Monocle.[37] The code to perform the
539 analyses in this manuscript is available at
540 https://github.com/umr1283/LGA_HSPC_PAPER.git.

541 **Data sharing statement.** The DNA methylation and gene expression data will be
542 made available upon request to A.P., P.F., or F.D.

543

544

545 ## References

546 1. Doulatov S, Notta F, Laurenti E, Dick JE. Hematopoiesis: a human perspective. *Cell*
547 *Stem Cell*. 2012; **10**(2): p. 120-36.
548 2. Eaves CJ. Hematopoietic stem cells: concepts, definitions, and the new reality. *Blood*.
549 2015; **125**(17): p. 2605-13.
550 3. Balistreri CR, Garagnani P, Madonna R, Vaiserman A, Melino G. Developmental
551 programming of adult haematopoiesis system. *Ageing Res Rev*. 2019; **54**: p. 100918.
552 4. Pardali E, Dimmeler S, Zeiher AM, Rieger MA. Clonal hematopoiesis, aging, and
553 cardiovascular diseases. *Exp Hematol*. 2020; **83**: p. 95-104.
554 5. Lee CC, Fletcher MD, Tarantal AF. Effect of age on the frequency, cell cycle, and lineage
555 maturation of rhesus monkey (Macaca mulatta) CD34+ and hematopoietic progenitor
556 cells. *Pediatr Res*. 2005; **58**(2): p. 315-22.
557 6. Chandel NS, Jasper H, Ho TT, Passegue E. Metabolic regulation of stem cell function in
558 tissue homeostasis and organismal ageing. *Nat Cell Biol*. 2016; **18**(8): p. 823-32.
559 7. Guenechea G, Gan OI, Dorrell C, Dick JE. Distinct classes of human stem cells that differ
560 in proliferative and self-renewal potential. *Nat Immunol*. 2001; **2**(1): p. 75-82.
561 8. Barker DJ, Osmond C. Infant mortality, childhood nutrition, and ischaemic heart
562 disease in England and Wales. *Lancet*. 1986; **1**(8489): p. 1077-81.
563 9. Barker DJ, Osmond C, Kajantie E, Eriksson JG. Growth and chronic disease: findings in
564 the Helsinki Birth Cohort. *Ann Hum Biol*. 2009; **36**(5): p. 445-58.
565 10. Cnattingius S, Villamor E, Lagerros YT, Wikstrom AK, Granath F. High birth weight and
566 obesity--a vicious circle across generations. *Int J Obes (Lond)*. 2012; **36**(10): p. 1320-4.
567 11. Wang Y, Gao E, Wu J, et al. Fetal macrosomia and adolescence obesity: results from a
568 longitudinal cohort study. *Int J Obes (Lond)*. 2009; **33**(8): p. 923-8.
569 12. Eriksson J, Forsen T, Tuomilehto J, Osmond C, Barker D. Size at birth, childhood growth
570 and obesity in adult life. *Int J Obes Relat Metab Disord*. 2001; **25**(5): p. 735-40.
571 13. Harder T, Rodekamp E, Schellong K, Dudenhausen JW, Plagemann A. Birth weight and
572 subsequent risk of type 2 diabetes: a meta-analysis. *Am J Epidemiol*. 2007; **165**(8): p.
573 849-57.
574 14. Pocobelli G, Dublin S, Enquobahrie DA, Mueller BA. Birth Weight and Birth Weight for
575 Gestational Age in Relation to Risk of Hospitalization with Primary Hypertension in
576 Children and Young Adults. *Matern Child Health J*. 2016; **20**(7): p. 1415-23.
577 15. Koyanagi A, Zhang J, Dagvadorj A, et al. Macrosomia in 23 developing countries: an
578 analysis of a multicountry, facility-based, cross-sectional survey. *Lancet*. 2013;
579 **381**(9865): p. 476-83.
580 16. Al-Sweedan SA, Musalam L, Obeidat B. Factors predicting the hematopoietic stem
581 cells content of the umbilical cord blood. *Transfus Apher Sci*. 2013; **48**(2): p. 247-52.
582 17. Chandra T, Afreen S, Kumar A, Singh U, Gupta A. Does umbilical cord blood-derived
583 CD34+ cell concentration depend on the weight and sex of a full-term infant? *J Pediatr*
584 *Hematol Oncol*. 2012; **34**(3): p. 184-7.
585 18. Kamimae-Lanning AN, Krasnow SM, Goloviznina NA, et al. Maternal high-fat diet and
586 obesity compromise fetal hematopoiesis. *Mol Metab*. 2015; **4**(1): p. 25-38.
587 19. Delahaye F, Wijetunga NA, Heo HJ, et al. Sexual dimorphism in epigenomic responses
588 of stem cells to extreme fetal growth. *Nat Commun*. 2014; **5**: p. 5187.
589 20. Stella CC, Cazzola M, De Fabritiis P, et al. CD34-positive cells: biology and clinical
590 relevance. *Haematologica*. 1995; **80**(4): p. 367-87.

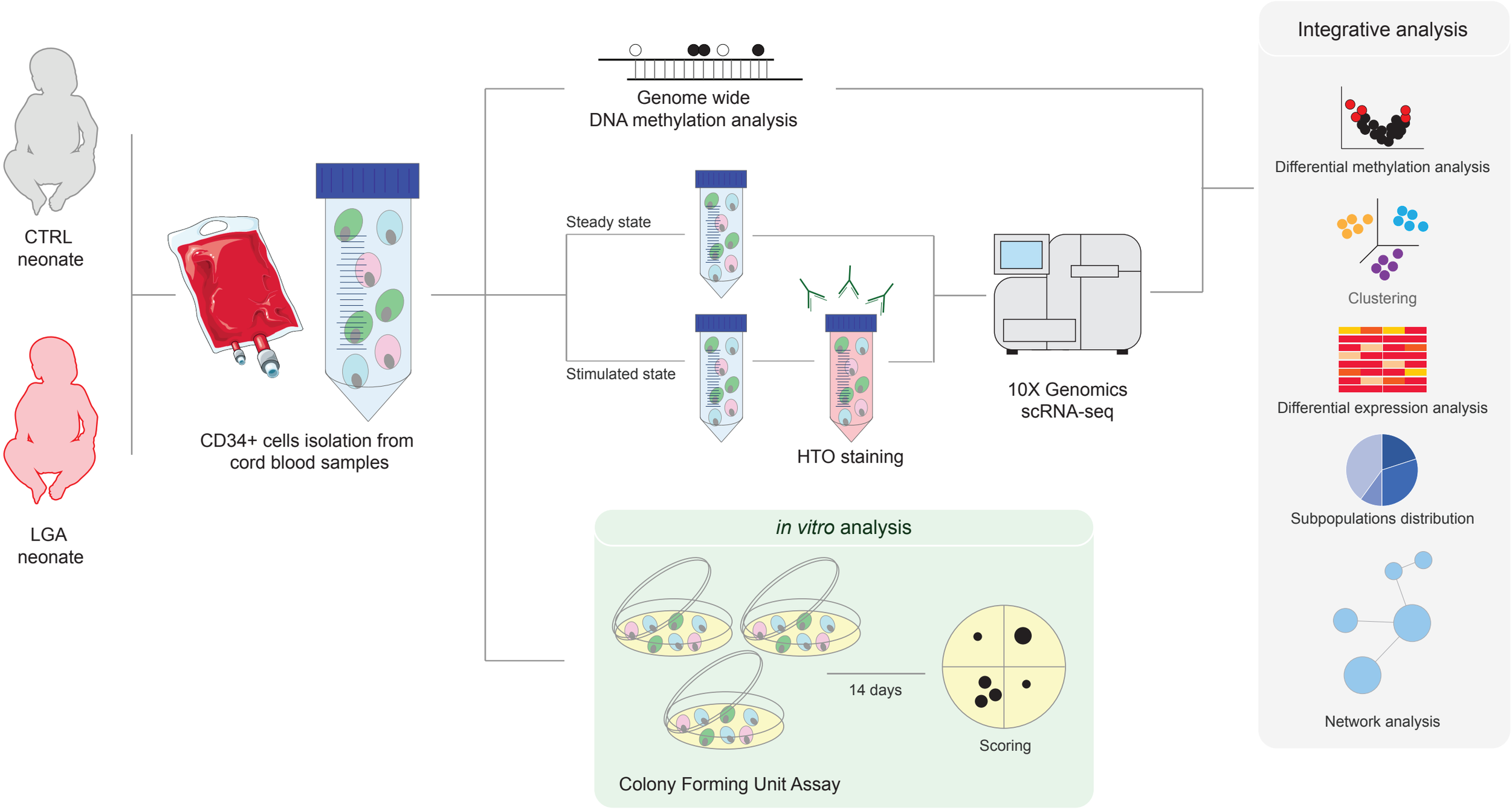
591 21. Thurman RE, Rynes E, Humbert R, et al. The accessible chromatin landscape of the
592 human genome. *Nature*. 2012; **489**(7414): p. 75-82.
593 22. Delahaye F, Do C, Kong Y, et al. Genetic variants influence on the placenta regulatory
594 landscape. *PLoS Genet*. 2018; **14**(11): p. e1007785.
595 23. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published
596 genome-wide association studies, targeted arrays and summary statistics 2019.
597 *Nucleic Acids Res*. 2019; **47**(D1): p. D1005-D1012.
598 24. Jiang J, Chan YS, Loh YH, et al. A core Klf circuitry regulates self-renewal of embryonic
599 stem cells. *Nat Cell Biol*. 2008; **10**(3): p. 353-60.
600 25. Zhao T, Liu C, Chen L. Roles of Klf5 Acetylation in the Self-Renewal and the
601 Differentiation of Mouse Embryonic Stem Cells. *PLoS One*. 2015; **10**(9): p. e0138168.
602 26. Lecuyer E, Herblot S, Saint-Denis M, et al. The SCL complex regulates c-kit expression
603 in hematopoietic cells through functional interaction with Sp1. *Blood*. 2002; **100**(7): p.
604 2430-40.
605 27. Gilmour J, O'Connor L, Middleton CP, et al. Robust hematopoietic specification
606 requires the ubiquitous Sp1 and Sp3 transcription factors. *Epigenetics Chromatin*.
607 2019; **12**(1): p. 33.
608 28. Min IM, Pietramaggiori G, Kim FS, Passegue E, Stevenson KE, Wagers AJ. The
609 transcription factor EGR1 controls both the proliferation and localization of
610 hematopoietic stem cells. *Cell Stem Cell*. 2008; **2**(4): p. 380-91.
611 29. Zhu H, Wang G, Qian J. Transcription factors as readers and effectors of DNA
612 methylation. *Nat Rev Genet*. 2016; **17**(9): p. 551-65.
613 30. Zhu J, Emerson SG. Hematopoietic cytokines, transcription factors and lineage
614 commitment. *Oncogene*. 2002; **21**(21): p. 3295-313.
615 31. van Galen P, Mbong N, Kreso A, et al. Integrated Stress Response Activity Marks Stem
616 Cells in Normal Hematopoiesis and Leukemia. *Cell Rep*. 2018; **25**(5): p. 1109-1117 e5.
617 32. Liu Y, Chen Y, Deng X, Zhou J. ATF3 Prevents Stress-Induced Hematopoietic Stem Cell
618 Exhaustion. *Front Cell Dev Biol*. 2020; **8**: p. 585771.
619 33. Tenen DG, Hromas R, Licht JD, Zhang DE. Transcription factors, normal myeloid
620 development, and leukemia. *Blood*. 1997; **90**(2): p. 489-519.
621 34. Antonchuk J, Sauvageau G, Humphries RK. HOXB4-induced expansion of adult
622 hematopoietic stem cells ex vivo. *Cell*. 2002; **109**(1): p. 39-45.
623 35. Sun Y, Zhou B, Mao F, et al. HOXA9 Reprograms the Enhancer Landscape to Promote
624 Leukemogenesis. *Cancer Cell*. 2018; **34**(4): p. 643-658 e5.
625 36. Magnusson M, Brun AC, Miyake N, et al. HOXA10 is a critical regulator for
626 hematopoietic stem cells and erythroid/megakaryocyte development. *Blood*. 2007;
627 **109**(9): p. 3687-96.
628 37. Qiu X, Hill A, Packer J, Lin D, Ma YA, Trapnell C. Single-cell mRNA quantification and
629 differential analysis with Census. *Nat Methods*. 2017; **14**(3): p. 309-315.
630 38. Yu VWC, Yusuf RZ, Oki T, et al. Epigenetic Memory Underlies Cell-Autonomous
631 Heterogeneous Behavior of Hematopoietic Stem Cells. *Cell*. 2017; **168**(5): p. 944-945.
632 39. Wilson A, Laurenti E, Trumpp A. Balancing dormant and self-renewing hematopoietic
633 stem cells. *Curr Opin Genet Dev*. 2009; **19**(5): p. 461-8.
634 40. Anwar T, Sen B, Aggarwal S, et al. Differentially regulated gene expression in
635 quiescence versus senescence and identification of ARID5A as a quiescence associated
636 marker. *J Cell Physiol*. 2018; **233**(5): p. 3695-3712.

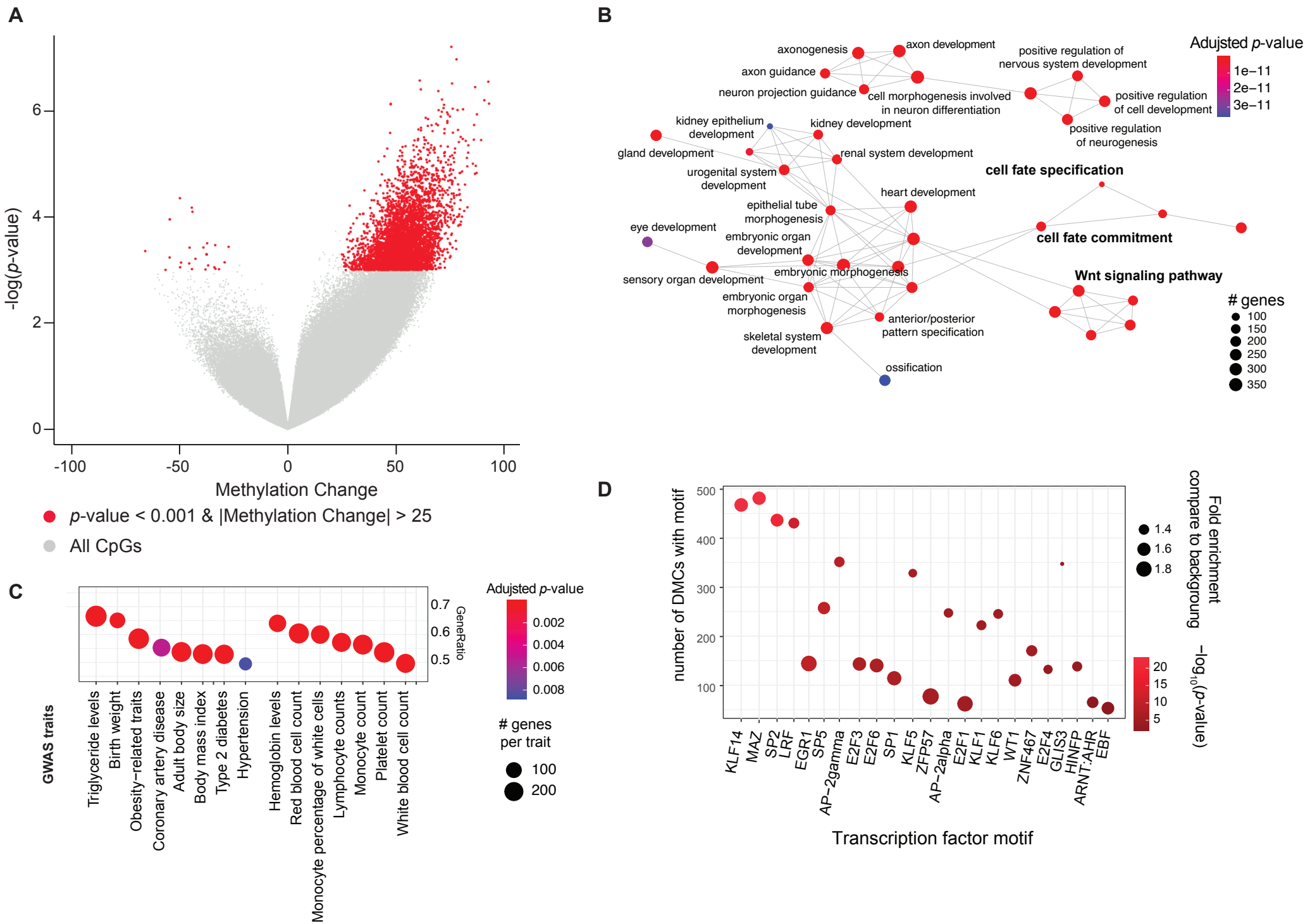
637 41. Masuda K, Ripley B, Nishimura R, et al. Arid5a controls IL-6 mRNA stability, which
638 contributes to elevation of IL-6 level in vivo. *Proc Natl Acad Sci U S A*. 2013; **110**(23):
639 p. 9409-14.
640 42. Nyati KK, Masuda K, Zaman MM, et al. TLR4-induced NF-kappaB and MAPK signaling
641 regulate the IL-6 mRNA stabilizing protein Arid5a. *Nucleic Acids Res*. 2017; **45**(5): p.
642 2687-2703.
643 43. Goodsell DS. The molecular perspective: estrogen sulfotransferase. *Oncologist*. 2006;
644 **11**(4): p. 418-9.
645 44. Tie R, Li H, Cai S, et al. Interleukin-6 signaling regulates hematopoietic stem cell
646 emergence. *Exp Mol Med*. 2019; **51**(10): p. 1-12.
647 45. Bahrami S, Drablos F. Gene regulation in the immediate-early response process. *Adv*
648 *Biol Regul*. 2016; **62**: p. 37-49.
649 46. Passegue E, Wagner EF, Weissman IL. JunB deficiency leads to a myeloproliferative
650 disorder arising from hematopoietic stem cells. *Cell*. 2004; **119**(3): p. 431-43.
651 47. Santaguida M, Schepers K, King B, et al. JunB protects against myeloid malignancies
652 by limiting hematopoietic stem cell proliferation and differentiation without affecting
653 self-renewal. *Cancer Cell*. 2009; **15**(4): p. 341-52.
654 48. Xiao X, Lai W, Xie H, et al. Targeting JNK pathway promotes human hematopoietic
655 stem cell expansion. *Cell Discov*. 2019; **5**: p. 2.
656 49. Okada S, Fukuda T, Inada K, Tokuhisa T. Prolonged expression of c-fos suppresses cell
657 cycle entry of dormant hematopoietic stem cells. *Blood*. 1999; **93**(3): p. 816-25.
658 50. McKinney-Freeman S, Cahan P, Li H, et al. The transcriptional landscape of
659 hematopoietic stem cell ontogeny. *Cell Stem Cell*. 2012; **11**(5): p. 701-14.
660 51. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic
661 and adult fibroblast cultures by defined factors. *Cell*. 2006; **126**(4): p. 663-76.
662 52. Cartron PF, Blanquart C, Hervouet E, Gregoire M, Vallette FM. HDAC1-mSin3a-NCOR1,
663 Dnmt3b-HDAC1-Egr1 and Dnmt1-PCNA-UHRF1-G9a regulate the NY-ESO1 gene
664 expression. *Mol Oncol*. 2013; **7**(3): p. 452-63.
665 53. Sardina JL, Collombet S, Tian TV, et al. Transcription Factors Drive Tet2-Mediated
666 Enhancer Demethylation to Reprogram Cell Fate. *Cell Stem Cell*. 2018; **23**(5): p. 727-
667 741 e9.
668 54. Kotowski M, Safranow K, Kawa MP, et al. Circulating hematopoietic stem cell count is
669 a valuable predictor of prematurity complications in preterm newborns. *BMC Pediatr*.
670 2012; **12**: p. 148.
671 55. Suzuki M, Jing Q, Lia D, Pascual M, McLellan A, Greally JM. Optimized design and data
672 analysis of tag-based cytosine methylation assays. *Genome Biol*. 2010; **11**(4): p. R36.
673 56. Stoeckius M, Zheng S, Houck-Loomis B, et al. Cell Hashing with barcoded antibodies
674 enables multiplexing and doublet detection for single cell genomics. *Genome Biol*.
675 2018; **19**(1): p. 224.
676 57. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for
677 RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015; **43**(7): p. e47.
678 58. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological
679 themes among gene clusters. *OMICS*. 2012; **16**(5): p. 284-7.
680 59. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining
681 transcription factors prime cis-regulatory elements required for macrophage and B
682 cell identities. *Mol Cell*. 2010; **38**(4): p. 576-89.

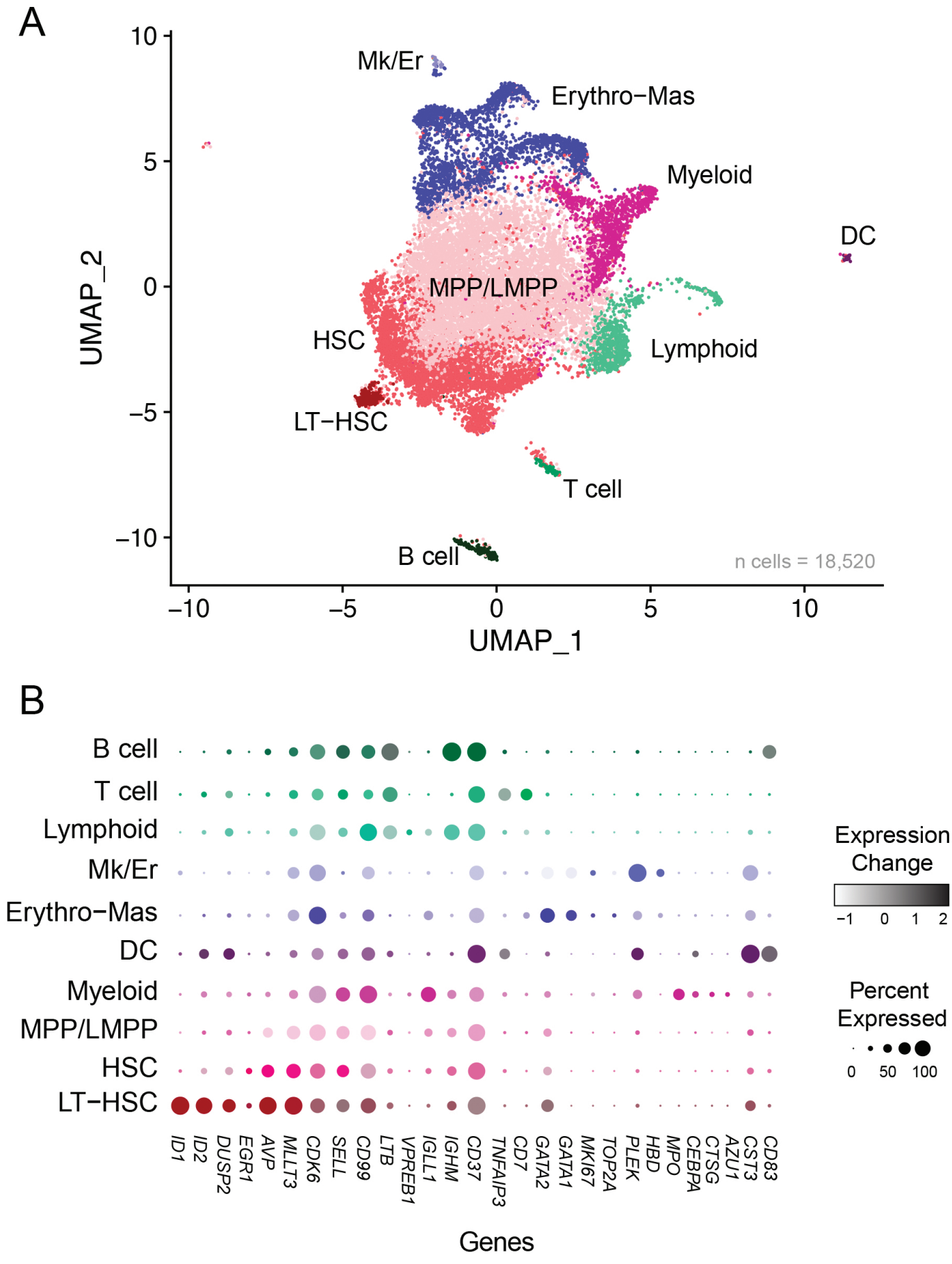
683 60. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for
684 RNA-seq data with DESeq2. *Genome Biol*. 2014; **15**(12): p. 550.
685 61. Aibar S, Gonzalez-Blas CB, Moerman T, et al. SCENIC: single-cell regulatory network
686 inference and clustering. *Nat Methods*. 2017; **14**(11): p. 1083-1086.
687

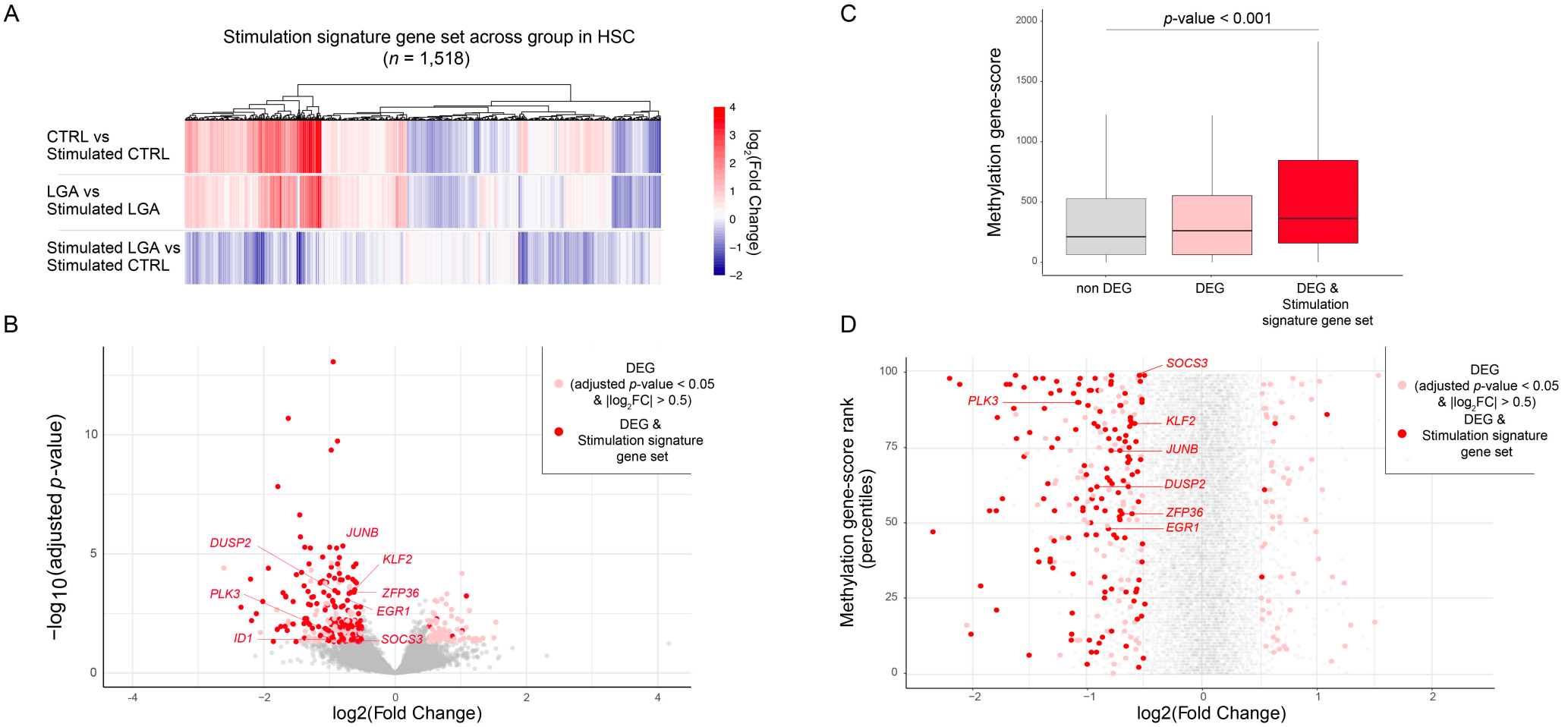
688

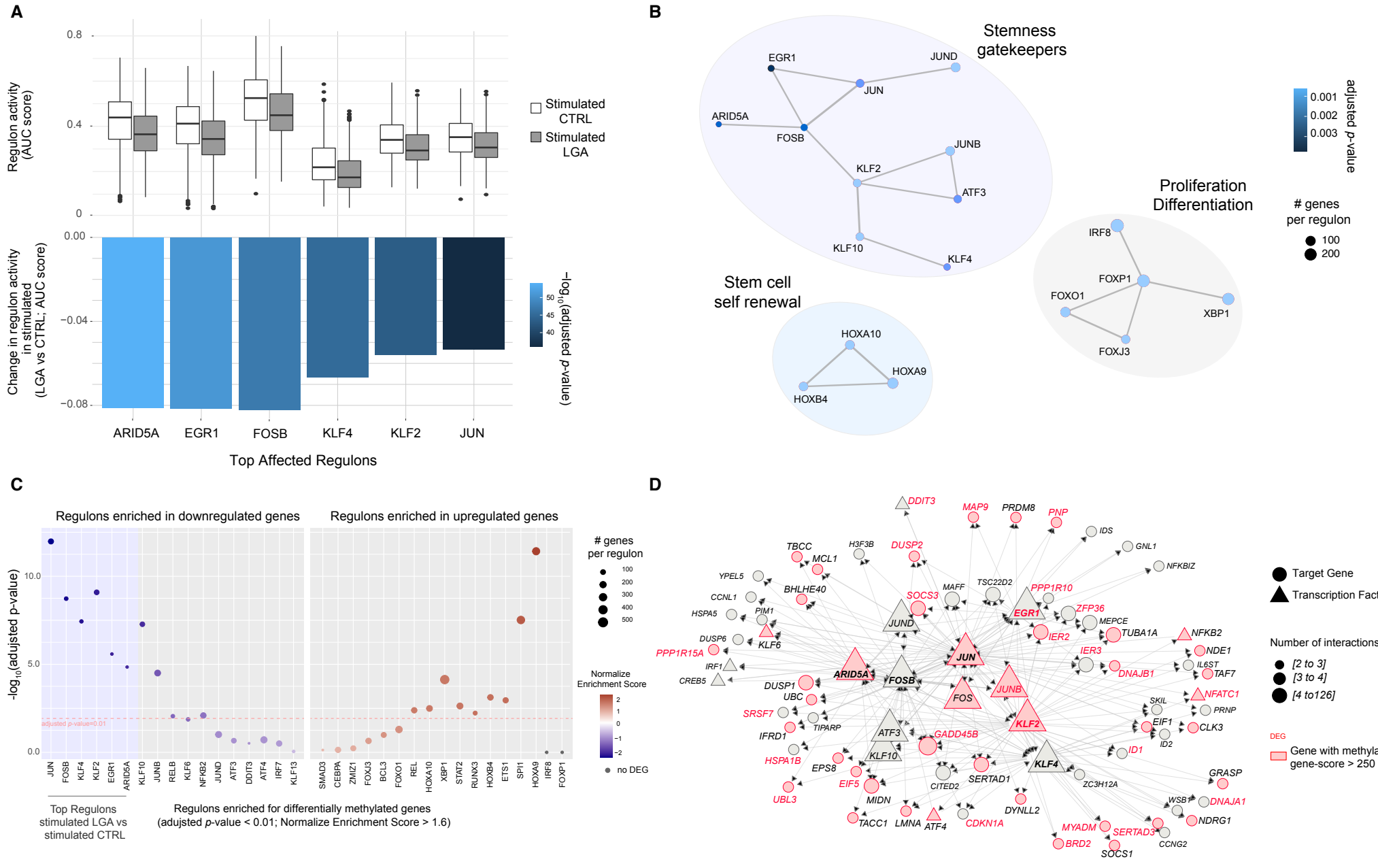
689

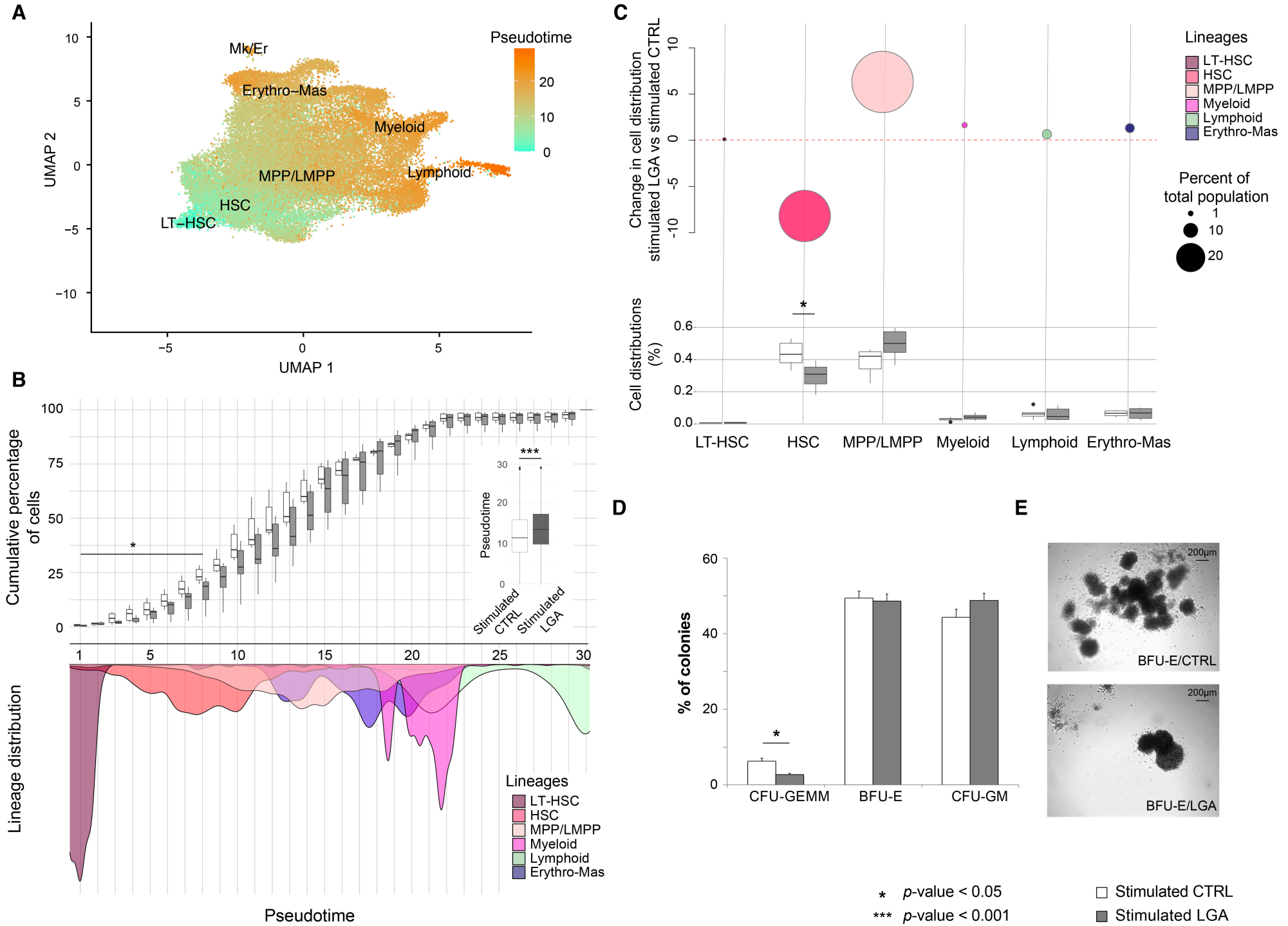

CTRL
neonate
LGA
neonate
CD34+ cells isolation from
cord blood samples
Genome wide
DNA methylation analysis
Steady state
Stimulated state
HTO staining
10X Genomics
scRNA-seq
in vitro analysis
14 days
Scoring
Colony Forming Unit Assay
Integrative analysis
Differential methylation analysis
Clustering
Differential expression analysis
Subpopulations distribution
Network analysis

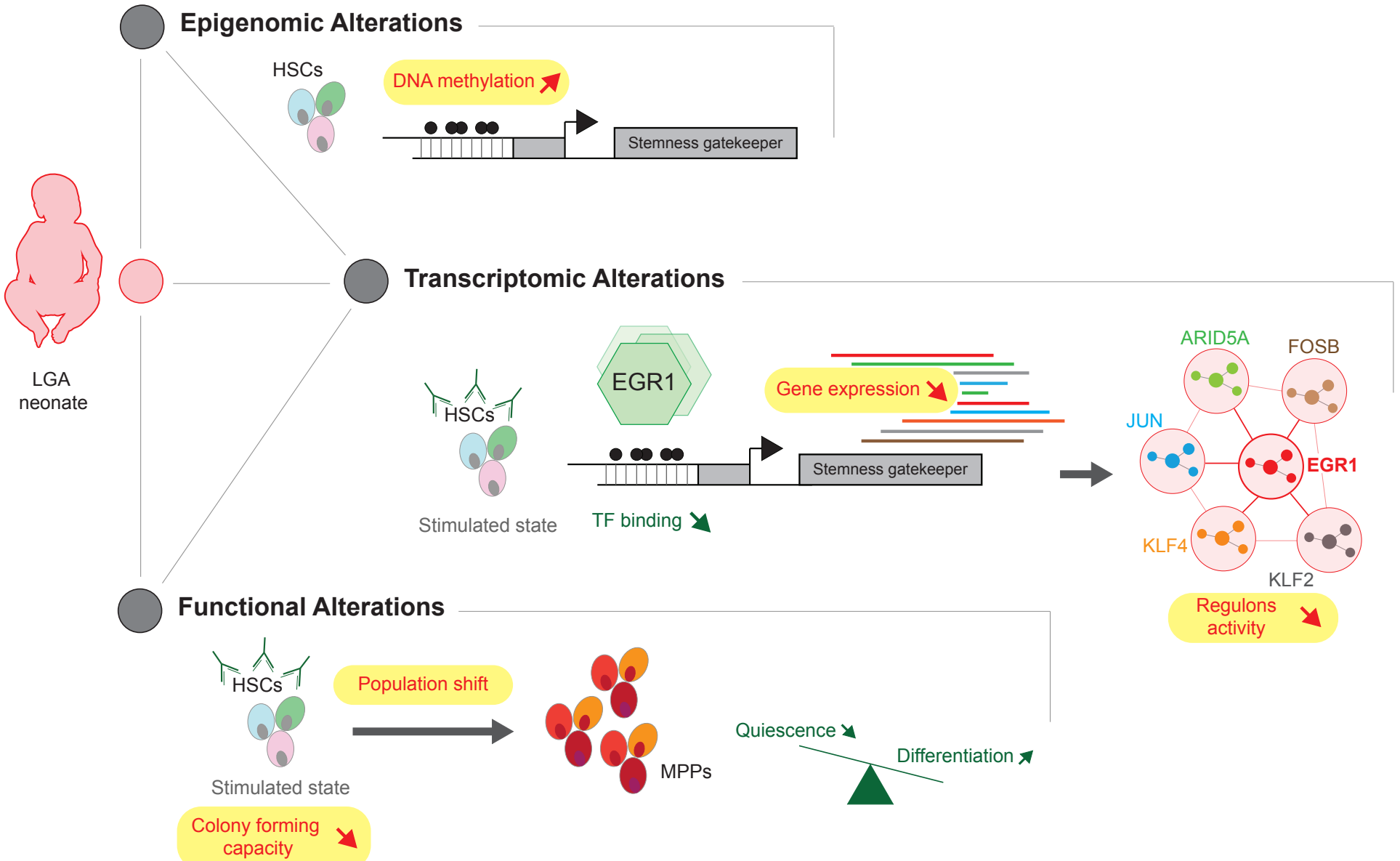

A
-log(p-value)
Methylation Change
p-value < 0.001 & |Methylation Change| > 25
All CpGs
B
axonogenesis
axon development
axon guidance
neuron projection guidance
cell morphogenesis involved in neuron differentiation
positive regulation of nervous system development
positive regulation of cell development
positive regulation of neurogenesis
kidney epithelium development
kidney development
gland development
renal system development
urogenital system development
cell fate specification
heart development
epithelial tube morphogenesis
eye development
embryonic organ development
cell fate commitment
sensory organ development
embryonic morphogenesis
Wnt signaling pathway
embryonic organ morphogenesis
anterior/posterior pattern specification
skeletal system development
ossification
Adjusted p-value
1e−11
2e−11
3e−11
genes
100
150
200
250
300
350
C
GeneRatio
GWAS traits
Triglyceride levels
Birth weight
Obesity−related traits
Coronary artery disease
Adult body size
Body mass index
Type 2 diabetes
Hypertension
Hemoglobin levels
Red blood cell count
Monocyte percentage of white cells
Lymphocyte counts
Monocyte count
Platelet count
White blood cell count
Adjusted p-value
0.002
0.004
0.006
0.008
genes per trait
100
200
D
number of DMCs with motif
Transcription factor motif
KLF14
MAZ
SP2
LRF
EGR1
SP5
AP−2gamma
E2F3
E2F6
SP1
KLF5
ZFP57
AP−2alpha
E2F1
KLF1
KLF6
WT1
ZNF467
E2F4
GLIS3
HINFP
ARNT::AHR
EBF
Fold enrichment compare to backgroung
−log10(p-value)


A
Mk/Er
Erythro−Mas
Myeloid
DC
MPP/LMPP
HSC
Lymphoid
LT−HSC
T cell
B cell
n cells = 18,520
UMAP_2
UMAP_1
10
5
0
−5
−10
B
B cell
T cell
Lymphoid
Mk/Er
Erythro−Mas
DC
Myeloid
MPP/LMPP
HSC
LT−HSC
ID1
ID2
DUSP2
EGR1
AVP
MLLT3
CDK6
SELL
CD99
LTB
VPREB1
IGLL1
IGHM
CD37
TNFAIP3
CD7
GATA2
GATA1
MKI67
TOP2A
PLEK
HBD
MPO
CEBPA
CTSG
AZU1
CST3
CD83
Genes
Expression Change
−1
0
1
2
Percent Expressed
0
50
100


A
Stimulation signature gene set across group in HSC
(n = 1,518)
CTRL vs
Stimulated CTRL
LGA vs
Stimulated LGA
Stimulated LGA vs
Stimulated CTRL
log2(Fold Change)
4
3
2
1
0
−1
−2
B
−log10(adjusted p-value)
log2(Fold Change)
DEG
(adjusted p-value < 0.05
& |log2FC| > 0.5)
DEG &
Stimulation signature
gene set
DUSP2
JUNB
KLF2
PLK3
ZFP36
EGR1
ID1
SOCS3
C
p-value < 0.001
Methylation gene-score
non DEG
DEG
DEG &
Stimulation
signature gene set
D
Methylation gene-score rank
(percentiles)
log2(Fold Change)
DEG
(adjusted p-value < 0.05
& |log2FC| > 0.5)
DEG &
Stimulation signature
gene set
SOCS3
PLK3
KLF2
JUNB
DUSP2
ZFP36
EGR1


A
Regulon activity (AUC score)
Stimulated CTRL
Stimulated LGA
Change in regulon activity in stimulated (LGA vs CTRL; AUC score)
−log10(adjusted p-value)
ARID5A
EGR1
FOSB
KLF4
KLF2
JUN
Top Affected Regulons
B
Stemness gatekeepers
Proliferation Differentiation
Stem cell self renewal
adjusted p-value
genes per regulon
C
Regulons enriched in downregulated genes
Regulons enriched in upregulated genes
-log10(adjusted p-value)
genes per regulon
Normalize Enrichment Score
no DEG
Top Regulons stimulated LGA vs stimulated CTRL
Regulons enriched for differentially methylated genes (adjusted p-value < 0.01; Normalize Enrichment Score > 1.6)
D
Target Gene
Transcription Factor
Number of interactions
[2 to 3]
[3 to 4]
[4 to126]
DEG
Gene with methylation gene-score > 250


A
Pseudotime
Mk/Er
Erythro–Mas
Myeloid
MPP/LMPP
Lymphoid
HSC
LT–HSC
UMAP 2
UMAP 1
B
Cumulative percentage of cells
Pseudotime
Stimulated CTRL
Stimulated LGA
Lineage distribution
Lineages
LT-HSC
HSC
MPP/LMPP
Myeloid
Lymphoid
Erythro-Mas
Pseudotime
C
Change in cell distribution stimulated LGA vs stimulated CTRL
Cell distributions (%)
Percent of total population
LT-HSC
HSC
MPP/LMPP
Myeloid
Lymphoid
Erythro-Mas
D
% of colonies
CFU-GEMM
BFU-E
CFU-GM
E
200µm
BFU-E/CTRL
BFU-E/LGA
* p-value < 0.05
*** p-value < 0.001
Stimulated CTRL
Stimulated LGA


Epigenomic Alterations
HSCs
DNA methylation
Stemness gatekeeper
LGA
neonate
Transcriptomic Alterations
HSCs
Stimulated state
EGR1
TF binding
Gene expression
Stemness gatekeeper
ARID5A
FOSB
JUN
EGR1
KLF4
KLF2
Regulons
activity
Functional Alterations
HSCs
Stimulated state
Population shift
MPPs
Quiescence
Differentiation
Colony forming
capacity