Compensatory mechanisms upon Gata2b haploinsufficiency causes aberrant transcriptional regulatory signatures in HSPCs resulting in myeloid and erythroid dysplasia in zebrafish

Emanuele Gioacchino[1], Wei Zhang[1, *], Cansu Koyunlar[1, *], Joke Zink[1], Hans de Looper[1,2], Kirsten J. Gussinklo[1], Remco Hoogenboezem[1], Dennis Bosch[1], Eric Bindels[1], Ivo P. Touw[1] and Emma de Pater[1,2].

Affiliations

[1] Department of Hematology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands

[2] Cancer Genome Editing Center, Erasmus MC Cancer Institute, Rotterdam, The Netherlands

[*]equal contribution

Corresponding author: e.depater@erasmusmc.nl

Abstract

The transcription factor GATA2 has pivotal roles in hematopoiesis. Germline *GATA2* mutations in patients result in GATA2 haploinsufficiency syndrome characterized by immunodeficiency, bone marrow failure, and predispositions to myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML). Clinical symptoms in GATA2 patients are diverse and mechanisms driving GATA2 related phenotypes largely unknown. To explore the impact of GATA2 haploinsufficiency on hematopoiesis, we generated a zebrafish model carrying a heterozygous mutation of *gata2b* (*gata2b*$^{+/-}$), an orthologue of *GATA2*. Morphological analysis revealed myeloid and erythroid dysplasia in *gata2b*$^{+/-}$ kidney marrow (KM). single nucleus (sn)-ATAC-seq showed that the co-accessibility between the transcription start site (TSS) and a +3.5-4.1kb enhancer was more robust in *gata2b*$^{+/-}$ zebrafish HSPCs compared to wild type, increasing *gata2b* expression. This is suggestive of an auto-regulatory feedback mechanism, where *gata2b* expression remains at sufficient levels after the loss of a single allele to maintain the HSPC pool. As a result, *gata2b*$^{+/-}$ chromatin is also more accessible in the erythroid and myeloid lineage, causing several defects. scRNA-seq data revealed a differentiation block in erythroid progenitors, hallmarked by downregulation of intrinsic

signals like cytoskeletal transcripts, aberrant proliferative signatures, and downregulation of *Gata1a*, a master regulator of erythropoiesis, likely preceding erythroid dysplasia. This shows that the cell intrinsic compensatory mechanisms for the maintenance of normal levels of Gata2b to maintain HSPC integrity result in aberrant lineage differentiation and a preleukemia syndrome.

Keywords

Gata2b, Zebrafish, MDS, Haploinsufficiency, HSCs

Introduction

The transcription factor GATA2 plays a major role in the generation and maintenance of the hematopoietic system[1-3]. In humans, heterozygous germline mutations in GATA2 often lead to a loss of function of one allele, causing GATA2 haploinsufficiency. The clinical manifestations of GATA2 haploinsufficiency are broad and include immunodeficiency, pulmonary-, vascular- and/or lymphatic dysfunctions and a strong propensity to develop myelodysplastic syndromes (MDS) or acute myeloid leukemia (AML)[4,5]. These conditions, now collectively referred to as GATA2 deficiency syndromes, were previously known as Emberger syndrome[6], Monocytopenia and Mycobacterium Avium Complex (MonoMAC) syndrome[7], dendritic cell, monocyte, B- and natural killer (NK) cell deficiency (DCML)[8], and familial forms of AML[9]. In addition to the various disease phenotypes, the risk of developing MDS/AML in GATA2 deficient patients is approximately 80% before the age of 40[4,5]. No clear correlation is found between the occurrence of GATA2 mutations and the severity of hematopoietic deficiencies, even among family members who share the same mutation[10,11]. Therefore, it is essential to gain insight into the mechanism of underlying GATA2 deficiencies in well-defined uniform experimental models.

In mice, Gata2 has an essential regulatory function in hematopoietic stem cell (HSC) generation and maintenance[1-3]. However, whereas Gata2-null mice are lethal at embryonic day (E) 10.5[3], *Gata2* heterozygous (*Gata2*$^{+/-}$) mice survive to adulthood with normal blood values. Notwithstanding this apparently normal blood phenotype, *Gata2*$^{+/-}$ mice have a diminished HSC compartment in the bone marrow (BM) which show a reduced repopulation capacity in competitive transplantation studies[12,13]. Whereas mouse models thus emerged as a useful source to identify the function of *GATA2* in HSC generation and fitness, they leave

62 the mechanisms causing the different aspects of GATA2 deficiency syndromes largely
63 undiscovered.
64 To better understand the biology of GATA2 haploinsufficiency syndromes, zebrafish
65 serves as an attractive alternative model. Zebrafish have the advantage of having 2 *GATA2*
66 orthologues; *Gata2a* and *Gata2b*. *Gata2a* is expressed predominantly in the vasculature[14]
67 and is required for programming of the hemogenic endothelium[15,16]. *Gata2b* is expressed in
68 hematopoietic stem/progenitor cells (HSPCs)[14] and homozygous deletion (*gata2b*$^{-/-}$)
69 redirects HSPC differentiation to the lymphoid lineage in expense of myeloid differentiation
70 causing a lymphoid bias with an incomplete B-cell differentiation in the kidney marrow (KM),
71 thus mimicking one of the GATA2 haploinsufficiency phenotypes found in patients.
72 Additionally, the most primitive HSC compartment was lost in *gata2b*$^{-/-}$ KM, but none of the
73 animals displayed signs of dysplasia[16,17]. Because patients carry heterozygous rather than
74 homozygous *GATA2* mutations, we hypothesize that Gata2b haploinsufficiency could be a
75 driver for leukemia onset. To investigate this, we assessed hematopoietic cell differentiation
76 in *gata2b* heterozygous zebrafish (*gata2b*$^{+/-}$) KM. We observed erythroid dysplasia in all
77 *gata2b*$^{+/-}$ but not in wild-type (WT) zebrafish and myeloid dysplasia in 25% of the *gata2b*$^{+/-}$
78 zebrafish. Single-cell RNA and ATAC sequencing analysis revealed an auto-regulatory
79 feedback mechanism of Gata2b where *gata2b* chromatin was over-accessible, underlying
80 the hematopoietic lineage differentiation defects in adults. Further characterization of
81 *gata2b*$^{+/-}$ zebrafish showed reduction of HSC proliferation, failure to initiate the "GATA2 to
82 GATA1 switch" and depletion of *Gata1a* and its co-factor *FOG1* in erythroid progenitor cells.
83 Taken together, our results reveal a concentration-dependent function of *Gata2b* and
84 hematopoietic defects in zebrafish hematopoiesis that reflect its dysfunction in
85 GATA2-deficient patients.
86
87 ***Material and methods***
88 **Generation and genotyping of Gata2b heterozygous zebrafish**
89 *Gata2b*$^{+/-}$ and wild type (WT) clutch mates were used for all analyses[16] and animals
90 were maintained under standard conditions. A knockout allele was generated by
91 introducing a 28bp out-of-frame insertion in exon 3 as previously described[16].
92 Zebrafish embryos were kept at 28.5°C on a 14h/10h light-dark cycle in HEPES-buffered
93 E3 medium. Zebrafish were anesthetized using tricaine and euthanized by ice-water. Animal

94 studies were approved by the animal Welfare/Ethics Committee in accordance to Dutch
95 legislation.
96
97 **Kidney marrow isolation and analysis**
98 Kidney marrow was dispersed mechanically using tweezers and dissociated by
99 pipetting in phosphate buffered saline (PBS)/10% fetal calf serum (FCS) to obtain single-cell
100 suspensions as previously described[16]. The KM cells were sorted in non-stick cooled micro
101 tubes (Ambion) containing 10% FCS in PBS. Proliferation was assessed by anti-Ki67 staining
102 in fixed (4% paraformaldehyde) KM cells. 7-AAD (7-amino-actinomycin D) (Stem-Kit
103 Reagents) 0.5mg/L or DAPI 1mg/L was used for live/dead discrimination. FACS sorting and
104 analysis were performed using FACS AriaIII (BD Biosciences).
105
106 **May–Grünwald-Giemsa stain of KM smears**
107 Kidney marrow smears were fixed in 100% MeOH before staining in May-Grünwald
108 solution (diluted 1:1 in phosphate buffer) and Giemsa solution (diluted 1:20 in phosphate
109 buffer) followed by a last rinsing step in tap water.
110 Morphological analysis was performed by pathologist by counting 200-500
111 hematopoietic cells of each kidney marrow smear; excluding mature erythrocytes and
112 thrombocytes. Cells were categorized as: blast, myelocyte, neutrophil, eosinophil,
113 lymphocyte or erythroblast. Furthermore, if dysplasia was observed within a specific
114 lineage, the percentage of dysplastic cells within that lineage was determined by additional
115 counting of at least 50 cells within that specific lineage.
116
117 **Single cell omics sequencing**
118 For single cell transcriptomics sequencing (scRNA-seq), kidney marrow cells were
119 isolated and $7x10^4$ viable cells were sorted from 2 pooled *Tg(CD41:GFP*[18]*; runx1:DsRed*[19]*)*
120 WT or *gata2b*$^{+/-}$ male zebrafish at 1 year of age. For additional replicates, $7x10^4$ single viable
121 cells were sorted from kidney marrows of one WT and one *gata2b*$^{+/-}$ female zebrafish
122 between 18-20 months of age. cDNA was prepared using the manufacturers protocol
123 (Chromium Single Cell 3' reagent kits kit v3, 10x Genomics) and sequenced on a Novaseq
124 6000 instrument (Illumina). After sample processing and quality control analysis, 18,147

125 cells for WT and 10,849 cells for *gata2b*$^{+/-}$ were processed for further analysis. The read
126 depth was over 20K reads for all replicates.
127 We collected the same KM cell composition as scRNA-seq for single nucleus chromatin
128 accessibility sequencing (snATAC-seq, Chromium Single Cell ATAC reagent kits, 10x
129 Genomics). After low quality data filtering, 4,190 nuclei for WT and 7,572 nuclei for
130 *gata2b*$^{+/-}$ were processed for downstream analysis.
131 Data was analyzed using the Seurat[20] and Signac[21] pipeline. Trajectory inference was
132 performed using Monocle3[22] R packages.
133
134 **Statistics**
135 Statistical analysis was carried out in GraphPad Prism 8 (GraphPad Software). Unless
136 otherwise specified, data were analyzed using unpaired, 2-tailed Student's t-test. Statistical
137 significance was defined as $p<0.05$. Graphs are means ± standard error of mean (SEM) and
138 the number of replicates is indicated in the figure legend.
139
140 ***Results***
141 *Gata2b*$^{+/-}$ *zebrafish have erythroid and myeloid dysplasia in the kidney marrow*
142 We first assessed hematopoietic cell morphology in KM smears of WT and *gata2b*$^{+/-}$
143 zebrafish with ages ranging from 9 months post fertilization (mpf) to 18 mpf. Morphological
144 analysis showed that, while WT zebrafish had normal KM cell morphology, all *gata2b*$^{+/-}$ KM
145 samples had a considerable fraction of dysplastic cells in the erythroid lineage (Figure 1A,
146 panel 1, 2, 4, 6, and 8-10). On average 0.5% of WT erythroid cells showed dysplastic features,
147 compared to 9.9% of *gata2b*$^{+/-}$ erythroid cells (Figure 1A and B), the latter representing 4.5%
148 of the total kidney marrow population of *gata2b*$^{+/-}$ zebrafish. Furthermore, we found
149 evidence of myeloid lineage dysplasia in the *gata2b*$^{+/-}$ KM in 25% of the fish (Figure 1A,
150 panel 3, 5 and 7, and B). In these samples, 30% of myeloid cells were dysplastic compared to
151 0.3% in WT. Myeloid dysplasia was mostly represented by multi-lobulated nuclei, whereas
152 the erythroid abnormalities ranged from nuclear deformities and double nuclei to irregular
153 cytoplasm or an almost complete lack of cytoplasm (Figure 1A). The remaining cell types
154 were not affected morphologically by Gata2b haploinsufficiency. These results indicate that
155 *gata2b* heterozygosity induces dysplasia, predominantly in erythroid progenitors.
156

157 To investigate if hematopoietic lineage differentiation ratios were still intact, WT and
158 *gata2b*$^{+/-}$ KM was assessed by flow cytometry (supplemental Figure 1A). *Gata2b*$^{+/-}$ zebrafish
159 KM showed no significant difference in the distribution of either mature myeloid-, erythroid-,
160 lymphoid- or progenitor- and HSPC populations compared to WT (supplemental Figure 1B-E).
161 Since GATA2 haploinsufficiency manifestations might require longer periods of time to
162 become evident[4,5], we tested the effect of *gata2b* heterozygosity in zebrafish up to 18
163 months of age. However, we found no significant differences in the ratios of different
164 lineages in *gata2b*$^{+/-}$ kidney marrow compared to WT during this period (supplemental
165 Figure 1B-E). Interestingly, the erythroid population significantly increased, indiscriminately
166 of genotype, after 12 months of age ($P<0.001$) (supplemental Figure 1D), while the
167 remaining myeloid, lymphoid and HSPC and progenitor populations did not vary as
168 dramatically (supplemental Figure 1B, C and E). To confirm the presence of a normal
169 hematopoietic lineage distribution in *gata2b*$^{+/-}$ kidney marrow we bred our zebrafish with
170 transgenic reporter zebrafish. Lineage analysis in transgenic lines specifically marking
171 neutrophils by *Tg(mpx:GFP)* (supplemental Figure 2A-B), T-cells by *Tg(lck:GFP)* (supplemental
172 Figure 2C-D), B-cells by *Tg(Igm:GFP)* and *Tg(mpeg:GFP)* (supplemental Figure 2E-F and I-J) or
173 monocytes by *Tg(mpeg:GFP)*[23,24] (supplemental Figure 2G-H) showed no significant
174 alterations in lineage distribution between WT and *gata2b*$^{+/-}$ KM [25-28]. KM smear
175 quantification by May–Grünwald-Giemsa (MGG) staining revealed that *gata2b*$^{+/-}$ KM had a
176 significant decrease in eosinophils compared to WT (supplemental Figure 2I), representing
177 less than 5% of the total KM cells. In summary, based on flow cytometry analysis and
178 transgenic reporters we conclude that the differentiation in major hematopoietic lineages is
179 not altered in *gata2b*$^{+/-}$ KM.
180
181 *Single cell RNA- and ATAC sequencing analysis reveals a cell maturation delay in gata2b*$^{+/-}$
182 *KM*
183 To investigate the molecular mechanisms underlying dysplasia in *gata2b*$^{+/-}$ KM, we
184 assessed chromatin accessibility, a measure for gene activation, and transcriptional
185 differences in dysplastic cell populations. We flow-sorted 4 cell populations based on light
186 scatter and observed dysplastic cells in the progenitor and lymphoid + HSPCs population of
187 *gata2b*$^{+/-}$ KM, indicating that dysplastic cells could be viably sorted (Figure 1A, panel 9 and
188 10). We did not identify a uniquely separated population of dysplastic cells, possibly caused
189 by a heterogeneity in their shape. Therefore, we sorted progenitor and lymphoid + HSPC

190 populations from WT and *gata2b*$^{+/-}$ KMs and performed single-cell (sc) RNA- and
191 ATAC-sequencing to identify transcriptome and epigenetic regulation signatures defining
192 dysplastic cells (Figure 2A). After filtering out low quality cells, we obtained 28,996 cells (WT
193 = 18,147, *gata2b*$^{+/-}$ = 10,849) for scRNA-seq analysis and 11,762 cells (WT = 4,190, *gata2b*$^{+/-}$
194 = 7,572) for snATAC-seq analysis (Figure 2B).

195 scRNA-sequencing resulted in 19 clusters based on a nearest neighbor algorithm using
196 the R Seurat package[20] (Figure 2B). Each cluster was classified based on differentially
197 expressed genes (supplemental Figure 3A-D) and known differentiation markers[29-34] (Figure
198 2C). We then used this labeled scRNA-seq dataset as a reference to interpret and identify 16
199 clusters in the snATAC-seq dataset based on the R Signac package[21] (Figure 2B-E). Cluster
200 proportion analysis indicated an overall similar distribution of cells within clusters between
201 *gata2b*$^{+/-}$ and WT (Figure 2F). However, a clear difference in differentiation trajectory was
202 observed between WT and *gata2b*$^{+/-}$ KM in snATAC-seq data based on the R package
203 Monocle3[22], and particularly note the difference in pseudo-time scores between WT and
204 *gata2b*$^{+/-}$ erythroid progenitors 1, suggesting a delayed differentiation in *gata2b*$^{+/-}$ KM
205 (Figure 2G). Whereas the initial pseudo-time score for cell differentiation at the HSPCs stage
206 was similar (Supplemental figure 4A), the pseudo-time scores of differentiation of erythroid,
207 myeloid, and lymphoid lineages were reduced in *gata2b*$^{+/-}$ KM (Figure 2G and supplemental
208 Figure 4A). Based on these findings, we conclude that *gata2b*$^{+/-}$ KM showed an overall
209 lineage maturation delay. As expected, a clear distinction in open chromatin patterns of
210 lineage specific transcription factors (TF) between immature and mature cells was seen in
211 WT KM (Supplemental Figure 4B, D and F, left panels). In contrast, this distinction was less
212 pronounced in *gata2b*$^{+/-}$ KM (Supplemental Figure 4C, E and G, left panels). Specifically, the
213 ETV family of transcription factors and the GATA family transcription factors were affected in
214 erythroid differentiation, SPI1, SPIB, and CEBPA family of transcription factors were affected
215 in myeloid differentiation, and TCF4 was most affected in lymphoid differentiation and
216 showed significantly different patterns of activation (supplemental Figure 4B-G). Surprisingly,
217 this did not always lead to a reduction in gene expression (Supplemental Figure 4 B-G, right
218 panels), particularly note the increased expression of *gata2b* in erythroid lineage
219 differentiation (Supplemental Figure 4C, right panel).
220

221 *gata2b heterozygosity results in over-accessible gata2b chromatin leading to self-renewal*
222 *defects in HSPCs.*

223 To further investigate the mechanism underlying the effects of haploinsufficient *gata2b*
224 expression, we assessed the accessible peak distribution of *gata2b* and its upstream region.
225 First of all, the chromatin accessibility signals of *gata2b* were strongest in the HSPCs cluster
226 (Figure 3A, top line), then gradually disappeared with cell maturation (Figure 3A, all other
227 lines), which is consistent with the GATA2 function in HSC maintenance[1-3]. However, when
228 we compared the signals between WT on the left and *gata2b*$^{+/-}$ on the right, we observed
229 that the chromatin of the *gata2b* gene body and upstream regions was more accessible in
230 *gata2b*$^{+/-}$ HSPCs. Chromatin co-accessibility analysis, which was used to predict
231 *cis*-regulatory interaction in the genome, suggested that there were two regions, +3.5-4.1kb
232 termed enhancer 1(E1) and +5.4-6.2kb, termed enhancer 2 (E2), which could be strongly
233 interacting with *gata2b* transcription start site (TSS) (Figure 3A, yellow lines in Links). These
234 two upstream regions might work as enhancers for *gata2b* transcription. When we assessed
235 the cell distribution of the two enhancers and *gata2b* TSS (Figure 3B-G), we observed that
236 the accessibility signals of +5.4-6.2kb region-E2, (Figure 3B, C), was accessible in HSPCs and
237 *gata2b*$^{+/-}$ erythroid progenitors, while a +3.5-4.1kb region-E1, (Figure 3D, E) was mainly
238 accessible in HSPCs and *gata2b*$^{+/-}$ myeloid progenitors. This result suggests that E1 and E2
239 might be responsible for different functions in hematopoiesis, and particularly in *gata2b*$^{+/-}$
240 KM, E1 affects myeloid differentiation and E2 affects erythroid differentiation. The
241 percentage of cells with *gata2b* mRNA expression did not show a significant difference as a
242 result from the up-regulated chromatin accessibility and stronger enhancer-TSS
243 co-accessibility (Supplementary Figure 5A, B), but we did see an increase in the level of
244 *gata2b* expression in the HSPC cluster as a result of the increase in enhancer-TSS
245 co-accessibility (Figure 3A, H, and I and Supplementary Figure 4C, right panel).
246
247 *The functional consequence of epigenetic changes of the gata2b locus in HSPCs is a*
248 *reduction in G2-M phase in cell cycle*
249 Because the most significant signal differences occurred in HSPCs clusters and the cell
250 differentiation velocity differences appeared at "progenitors" clusters, we compared the
251 different TF activity in progenitor-like clusters, including "HSPCs", "progenitors", and "HSPCs
252 and thrombocytes". *Gata2b*$^{+/-}$ progenitor-like clusters showed significantly up-regulated
253 chromatin accessibility of motifs of the BATF family of transcription factors. The BATF family
254 of TFs are known to function in limiting the self-renewal of HSCs in response to the
255 accumulation of DNA damage[35] (Figure 4A), suggesting a defect in self-renewal of *gata2b*$^{+/-}$

256 HSCs. To investigate this, proliferation was assessed in CD41:GFP^{low} purified HSCs (Figure 4B).
257 Although numerically unaffected by *gata2b* heterozygosity (Figure 4C), cell cycle phase
258 distribution was markedly different in *gata2b*$^{+/-}$ HSCs compared to WT (Figure 4D and E),
259 with increased G1 phase and decreased G2-M phase. As stem cells are known to have a
260 short G1 phase to reduce cellular differentiation potential, this may already be an indication
261 that these HSCs lose self-renewal capacity and differentiate, which is halted in the next
262 differentiation stages[36].
263
264 *Epigenetic dysregulation results in an erythroid differentiation block in gata2b*$^{+/-}$ *kidney*
265 *marrow*
266 Next, we sought to reveal the molecular mechanism behind the erythroid dysplasia in
267 *Gata2*$^{+/-}$ KM. Compared to WT, *Gata2b*$^{+/-}$ progenitor-like clusters had significantly reduced
268 accessibility of GATA2 and GATA1::TAL1 motifs (Figure 4A). Despite the increase in *gata2b*
269 expression, this indicated an overall reduction of Gata2b function in activation of
270 downstream targets in the *gata2b*$^{+/-}$ zebrafish genome. Activation of GATA2 motifs were
271 mainly enriched in the erythroid progenitors (Figure 5A). An important downstream target
272 of Gata2 is Gata1, which is the master regulator of erythropoiesis. We therefore investigated
273 *Gata1a* accessibility and expression in our dataset (Figure 5B-F). The *gata1a* TSS was
274 generally more accessible in *gata2b*$^{+/-}$ KM (Figure 5B), but not in erythroid progenitors 1
275 (Figure 5C). Furthermore, we found downregulation of *gata1a* mRNA expression (Figure 5D)
276 and the GATA1::TAL1 motif was less accessible (Figure 5E), further supporting the reduced
277 functionality of Gata1a in erythroid progenitors. This could be due to reduced expression of
278 *zfpm1* (also known as *Friend of GATA* (*FOG1)*), which can facilitate *GATA1* expression[37](Figure
279 5F). The different pattern of the GATA family (Increased expression of *gata2b* in HSPCs
280 (Figure 3I) and reduced expression of *gata1a* in erythroid progenitors (Figure 5D) suggested
281 that the "GATA factor switch", which is a shift from GATA2 to GATA1 occupation during
282 erythroid lineage differentiation[38-40], might be silenced, causing deficiencies in erythroid
283 lineage differentiation.
284 Considering that the origin of dysplastic cells were readily recognized as myeloid or
285 erythroid in KM smears, we hypothesized that dysplastic cells occupy a proportion of
286 erythroid or myeloid clusters. For a more precise examination of small populations along the
287 myeloid and erythroid differentiation lineage, we sub-clustered ''HSPCs'', "proliferative
288 progenitors'', "progenitors", "myeloid progenitors 1" and "erythroid progenitors 2" clusters

289 and identified 5 significantly overrepresented sub-clusters in *gata2b*$^{+/-}$ KM (Figure 5G-I,
290 supplemental Figure 6 A-H). These clusters are only present in *gata2b*$^{+/-}$ KM and could
291 therefore be the dysplastic cells. These sub-clusters represented 5.6% of the total sequenced
292 *gata2b*$^{+/-}$ cells, comparable to the proportion of dysplastic cells observed in *gata2b*$^{+/-}$ KM.
293 Differentially expressed genes (DEGs) analysis of overrepresented sub-clusters in *gata2b*$^{+/-}$
294 KM compared to the rest of that cluster showed downregulation of tubulin transcripts (not
295 shown), suggesting a loss of cytoskeletal structure, a characteristic of dysplasia. Besides, TF
296 motif score analysis with snATAC-seq indicated a general reduction for lineage-specific
297 determination and differentiation motifs in *gata2b*$^{+/-}$ KM (supplemental Figure 5B-G),
298 suggesting that *gata2b*$^{+/-}$ cell maturation could be impaired resulting in a differentiation
299 block, as this is often the case in dysplasia[41].
300 To identify the differentiation block mechanisms in *gata2b*$^{+/-}$ KM, we compared the
301 DEGs and their function in the "erythroid progenitor 2" cluster. The result showed that
302 *gata2b*$^{+/-}$ "erythroid progenitor 2" cluster cells had upregulated signatures of DNA
303 replication together with the downregulation of transcripts necessary for mitosis (Figure 5J),
304 which can explain the origin of multi-lobulated nuclei and other nuclear abnormalities
305 observed in *gata2b*$^{+/-}$ dysplastic cells[42,43]. Pseudo-time trajectory inference showed a clearly
306 diminished differentiation in *gata2b*$^{+/-}$ "erythroid progenitor 2" cluster with a lower
307 pseudo-time score (Figure 5K and L). Pseudo-time dynamic gene expression level also
308 indicated the erythropoiesis block in *gata2b*$^{+/-}$ "erythroid progenitor 2" cluster with
309 persistent high open chromatin, but lower expression of *gata1a*, which should gradually
310 decrease toward terminal erythroid maturation, and the increasing level of *mki67*, which
311 should be decreasing during maturation, indicative of incomplete differentiation (comparing
312 Figure 5M-O to Figure 5P-R). This could be due to a diminished open chromatin structure at
313 the locus leading to a reduction in Gata2b binding sites to drive proper *Gata1a* expression,
314 to support the erythroid differentiation process. Furthermore, TF motif score analysis of the
315 Erythroid progenitor 2 cluster revealed a significant depletion for GATA family motifs, but an
316 enrichment for Ets and Etv family of transcription factor motifs in *gata2b*$^{+/-}$ Erythroid
317 progenitors 2 cells (Figure 5S), suggesting an epigenetic dysregulation related to decreased
318 maturation of erythroblasts resulting from Gata2b haploinsufficiency.
319
320
321 ***Discussion***

322 In humans, a balanced *GATA2* expression is essential for proper hematopoiesis.
323 Consequently, more than 80% of *GATA2* mutation carriers progress to hematological
324 malignancy by 40 years of age[4]. While the clinical consequences of *GATA2* mutations became
325 obvious over the last decades, the regulation of *GATA2* activity and its contribution to
326 human bone marrow failure syndromes and progression to hematological malignancy
327 remain incompletely understood.

328 Here, we used transgenic reporters, morphological phenotyping, and single-cell omics
329 sequencing to analyze *gata2b* heterozygous zebrafish. We showed that, while major
330 differentiation lineages remain intact, *gata2b* heterozygosity causes dysplasia in erythroid
331 and myeloid progenitors of zebrafish KM. snATAC-seq revealed an auto-regulatory feedback
332 mechanism for Gata2b haploinsufficiency, characterized by *gata2b* chromatin
333 over-accessibility and stronger co-accessibility of *gata2b* enhancers and TSS, indicating
334 epigenetic dysregulation in hematopoiesis in the zebrafish model. scRNA-seq analysis did
335 not identify a single population of dysplastic cells. Interestingly, a single erythroid progenitor
336 cluster was identified using snATAC-seq, and could indicate a specific dysplastic erythroid
337 population. The difference in recognition of the dysplastic cluster between scRNAseq and
338 snATACseq may point to the underlying mechanism of the dysplasia where we have evidence
339 that the epigenetic compensatory mechanism of maintaining normal levels of Gata2b may
340 cause the dysplastic features during lineage differentiation, particularly in the GATA2 to
341 GATA1 switch and therefore affects erythroid lineage differentiation most. This also
342 underscores the importance of epigenetic analysis in other types of MDS and the fact that
343 many epigenetic regulators are mutated in MDS and AML[46]. In *gata2b*$^{+/-}$ erythroid
344 progenitors we found increased proliferative signatures together with decreased expression
345 of genes related to mitosis indicating an impaired cell cycle progression. Furthermore, we
346 found the depletion of GATA family factors indicating a failure of the "GATA factor switch"
347 that is indispensable for erythroid lineage differentiation[38-40]. We propose that these
348 alterations could play a role in the onset of dysplasia found in *gata2b*$^{+/-}$ zebrafish.

349 Interestingly, whereas *gata2b*$^{-/-}$ zebrafish display abrogated myeloid lineage
350 differentiation and a bias toward lymphoid differentiation [16,17], *gata2b*$^{+/-}$ did not simply
351 result in an intermediate phenotype between WT and *gata2b*$^{-/-}$. Instead, Gata2b
352 haploinsufficiency uniquely caused dysplasia, not observed in *gata2b*$^{-/-}$, potentially caused
353 by compensatory mechanisms in haploinsufficiency that are otherwise overcome in
354 *gata2b*$^{-/-}$ zebrafish[47]. This may also be the reason that missense mutations have a different

355 molecular mechanism, because these alleles are not prone to mRNA degradation and may
356 thus not be compensated by epigenetic deregulation. The differences in the homozygous
357 and heterozygous Gata2b knockout phenotype support a role for gene dosage underlying
358 the GATA2 deficiency phenotype, possibly explaining the phenotypic heterogeneity between
359 patients. Since both erythroid and myeloid dysplasia can be observed in GATA2 patients[4], we
360 propose that the presence of dysplastic cells in *gata2b*$^{+/-}$ resembles the clinical phenotypes
361 associated with GATA2 heterozygosity. In the future, the isolation of single dysplastic cells
362 could help us to further explore the effect of Gata2 haploinsufficiency in malignant
363 transformation. Nevertheless, it remains to be established how *gata2b*$^{+/-}$ HSPCs would
364 respond to secondary insults such as infections or severe bleeding.

365 In conclusion, while the major lineage differentiation remains intact, *gata2b*$^{+/-}$ zebrafish
366 possess a stressed proliferative HSPC compartment which leads to the generation of
367 erythroid and myeloid dysplastic cells. Taken together, our model provides insights into the
368 consequences of Gata2b dosage, illustrated the alteration of the microenvironment, and
369 reveal how changes in epigenetics affect the outcome in lineage differentiation after Gata2b
370 haploinsufficiency in zebrafish.
371

372 **Acknowledgements**
373 We thank Dr Monteiro (University of Birmingham) for careful reading of the manuscript. We
374 thank the Experimental Animal Facility of Erasmus MC for animal husbandry and the
375 Erasmus Optical Imaging Center for confocal microscopy services. This research is supported
376 by the European Hematology Association (junior and senior non clinical research fellowship)
377 (EdP), the Dutch Cancer Foundation KWF/Alpe d'HuZes (SK10321)(EdP), the Daniel den Hoed
378 Foundation for support of the Cancer Genome Editing Center (IT) and the Josephine Nefkens
379 Foundation for purchase of the Chromium 10x (IT).
380

381 **Authorship contributions**
382 EdP and EG conceived the study; EG, CK, HdL, JZ, DB, and EB performed experiments; EG, WZ,
383 CK, RH, KG and EdP analysed results; IT provided resources and EG, WZ, CK and EdP wrote
384 the manuscript and IT revised the manuscript.
385

386 **Disclosures**
387 The authors declare no conflicts of interests

388

389

390 **References**

391

392 1. Gao X, Johnson KD, Chang YI, et al. Gata2 cis-element is required for
393 hematopoietic stem cell generation in the mammalian embryo. *J Exp Med*. Dec 16
394 2013;210(13):2833-42. doi:jem.20130733 [pii]
395 10.1084/jem.20130733
396 2. de Pater E, Kaimakis P, Vink CS, et al. Gata2 is required for HSC generation and
397 survival. *J Exp Med*. Dec 16 2013;210(13):2843-50. doi:jem.20130751 [pii]
398 10.1084/jem.20130751
399 3. Tsai FY, Keller G, Kuo FC, et al. An early haematopoietic defect in mice lacking
400 the transcription factor GATA-2. *Nature*. Sep 15 1994;371(6494):221-6.
401 doi:10.1038/371221a0
402 4. Donadieu J, Lamant M, Fieschi C, et al. Natural history of GATA2 deficiency in a
403 survey of 79 French and Belgian patients. *Haematologica*. Aug
404 2018;103(8):1278-1287. doi:haematol.2017.181909 [pii]
405 10.3324/haematol.2017.181909
406 5. Spinner MA, Sanchez LA, Hsu AP, et al. GATA2 deficiency: a protean disorder of
407 hematopoiesis, lymphatics, and immunity. *Blood*. Feb 6 2014;123(6):809-21.
408 doi:S0006-4971(20)36040-7 [pii]
409 10.1182/blood-2013-07-515528
410 6. Ostergaard P, Simpson MA, Connell FC, et al. Mutations in GATA2 cause
411 primary lymphedema associated with a predisposition to acute myeloid leukemia
412 (Emberger syndrome). *Nat Genet*. Sep 4 2011;43(10):929-31. doi:ng.923 [pii]
413 10.1038/ng.923
414 7. Hsu AP, Sampaio EP, Khan J, et al. Mutations in GATA2 are associated with the
415 autosomal dominant and sporadic monocytopenia and mycobacterial infection
416 (MonoMAC) syndrome. *Blood*. Sep 8 2011;118(10):2653-5.
417 doi:blood-2011-05-356352 [pii]
418 10.1182/blood-2011-05-356352
419 8. Dickinson RE, Griffin H, Bigley V, et al. Exome sequencing identifies GATA-2
420 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency.

421 *Blood*. Sep 8 2011;118(10):2656-8. doi:S0006-4971(20)40687-1 [pii]
422 10.1182/blood-2011-06-360313
423 9. Hahn CN, Chong CE, Carmichael CL, et al. Heritable GATA2 mutations
424 associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat*
425 *Genet*. Sep 4 2011;43(10):1012-7. doi:ng.913 [pii]
426 10.1038/ng.913
427 10. Mutsaers PG, van de Loosdrecht AA, Tawana K, Bodor C, Fitzgibbon J, Menko
428 FH. Highly variable clinical manifestations in a large family with a novel GATA2
429 mutation. *Leukemia*. Nov 2013;27(11):2247-8. doi:leu2013105 [pii]
430 10.1038/leu.2013.105
431 11. Wang X, Muramatsu H, Okuno Y, et al. GATA2 and secondary mutations in
432 familial myelodysplastic syndromes and pediatric myeloid malignancies.
433 *Haematologica*. Oct 2015;100(10):e398-401. doi:haematol.2015.127092 [pii]
434 10.3324/haematol.2015.127092
435 12. Rodrigues NP, Janzen V, Forkert R, et al. Haploinsufficiency of GATA-2 perturbs
436 adult hematopoietic stem-cell homeostasis. *Blood*. Jul 15 2005;106(2):477-84.
437 doi:2004-08-2989 [pii]
438 10.1182/blood-2004-08-2989
439 13. Ling KW, Ottersbach K, van Hamburg JP, et al. GATA-2 plays two functionally
440 distinct roles during the ontogeny of hematopoietic stem cells. *J Exp Med*. Oct 4
441 2004;200(7):871-82. doi:jem.20031556 [pii]
442 10.1084/jem.20031556
443 14. Butko E, Distel M, Pouget C, et al. Gata2b is a restricted early regulator of
444 hemogenic endothelium in the zebrafish embryo. *Development*. Mar 15
445 2015;142(6):1050-61. doi:10.1242/dev.119180
446 15. Dobrzycki T, Mahony CB, Krecsmarik M, et al. Deletion of a conserved Gata2
447 enhancer impairs haemogenic endothelium programming and adult Zebrafish
448 haematopoiesis. *Commun Biol*. Feb 13 2020;3(1):71.
449 doi:10.1038/s42003-020-0798-3
450 10.1038/s42003-020-0798-3 [pii]
451 16. Gioacchino E, Koyunlar C, Zink J, et al. Essential role for Gata2 in modulating
452 lineage output from hematopoietic stem cells in zebrafish. *Blood Adv*. Jul 13
453 2021;5(13):2687-2700. doi:S2473-9529(21)00348-7 [pii]

454 10.1182/bloodadvances.2020002993
455 17. Avagyan S, Weber MC, Ma S, et al. Single-cell ATAC-seq reveals
456 GATA2-dependent priming defect in myeloid and a maturation bottleneck in lymphoid
457 lineages. *Blood Adv*. Jul 13 2021;5(13):2673-2686. doi:S2473-9529(21)00347-5 [pii]
458 10.1182/bloodadvances.2020002992
459 18. Ma D, Zhang J, Lin HF, Italiano J, Handin RI. The identification and
460 characterization of zebrafish hematopoietic stem cells. *Blood*. Jul 14
461 2011;118(2):289-97. doi:blood-2010-12-327403 [pii]
462 10.1182/blood-2010-12-327403
463 19. Tamplin OJ, Durand EM, Carr LA, et al. Hematopoietic stem cell arrival triggers
464 dynamic remodeling of the perivascular niche. *Cell*. Jan 15 2015;160(1-2):241-52.
465 doi:10.1016/j.cell.2014.12.032
466 20. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal
467 single-cell data. *Cell*. Jun 24 2021;184(13):3573-3587 e29.
468 doi:S0092-8674(21)00583-3 [pii]
469 10.1016/j.cell.2021.04.048
470 21. Stuart T, Srivastava A, Madad S, Lareau CA, Satija R. Single-cell chromatin state
471 analysis with Signac. *Nat Methods*. Nov 2021;18(11):1333-1341.
472 doi:10.1038/s41592-021-01282-5 [pii]
473 10.1038/s41592-021-01282-5
474 22. Cao J, Spielmann M, Qiu X, et al. The single-cell transcriptional landscape of
475 mammalian organogenesis. *Nature*. Feb 2019;566(7745):496-502.
476 doi:10.1038/s41586-019-0969-x [pii]
477 10.1038/s41586-019-0969-x
478 23. Boatman S, Barrett F, Satishchandran S, Jing L, Shestopalov I, Zon LI. Assaying
479 hematopoiesis using zebrafish. *Blood Cells Mol Dis*. Dec 2013;51(4):271-6.
480 doi:S1079-9796(13)00160-5 [pii]
481 10.1016/j.bcmd.2013.07.009
482 24. Traver D, Paw BH, Poss KD, Penberthy WT, Lin S, Zon LI. Transplantation and in
483 vivo imaging of multilineage engraftment in zebrafish bloodless mutants. *Nat Immunol*.
484 Dec 2003;4(12):1238-46. doi:10.1038/ni1007
485 ni1007 [pii]
486 25. Langenau DM, Ferrando AA, Traver D, et al. In vivo tracking of T cell

487 development, ablation, and engraftment in transgenic zebrafish. *Proc Natl Acad Sci U*
488 *S A*. May 11 2004;101(19):7369-74. doi:10.1073/pnas.0402248101
489 0402248101 [pii]
490 26. Renshaw SA, Loynes CA, Trushell DM, Elworthy S, Ingham PW, Whyte MK. A
491 transgenic zebrafish model of neutrophilic inflammation. *Blood*. Dec 15
492 2006;108(13):3976-8. doi:blood-2006-05-024075 [pii]
493 10.1182/blood-2006-05-024075
494 27. Ellett F, Pase L, Hayman JW, Andrianopoulos A, Lieschke GJ. mpeg1 promoter
495 transgenes direct macrophage-lineage expression in zebrafish. *Blood*. Jan 27
496 2011;117(4):e49-56. doi:blood-2010-10-314120 [pii]
497 10.1182/blood-2010-10-314120
498 28. Page DM, Wittamer V, Bertrand JY, et al. An evolutionarily conserved program of
499 B-cell development and activation in zebrafish. *Blood*. Aug 22 2013;122(8):e1-11.
500 doi:blood-2012-12-471029 [pii]
501 10.1182/blood-2012-12-471029
502 29. Carmona SJ, Teichmann SA, Ferreira L, et al. Single-cell transcriptome analysis
503 of fish immune cells provides insight into the evolution of vertebrate immune cell
504 types. *Genome Res*. Mar 2017;27(3):451-461. doi:10.1101/gr.207704.116
505 30. Danilova N, Bussmann J, Jekosch K, Steiner LA. The immunoglobulin
506 heavy-chain locus in zebrafish: identification and expression of a previously unknown
507 isotype, immunoglobulin Z. *Nat Immunol*. Mar 2005;6(3):295-302. doi:ni1166 [pii]
508 10.1038/ni1166
509 31. Kortum AN, Rodriguez-Nunez I, Yang J, et al. Differential expression and ligand
510 binding indicate alternative functions for zebrafish polymeric immunoglobulin receptor
511 (pIgR) and a family of pIgR-like (PIGRL) proteins. *Immunogenetics*. Apr
512 2014;66(4):267-79. doi:10.1007/s00251-014-0759-4
513 32. Macaulay IC, Svensson V, Labalette C, et al. Single-Cell RNA-Sequencing
514 Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells. *Cell Rep*.
515 Feb 2 2016;14(4):966-977. doi:S2211-1247(15)01538-7 [pii]
516 10.1016/j.celrep.2015.12.082
517 33. Tang Q, Iyer S, Lobbardi R, et al. Dissecting hematopoietic and renal cell
518 heterogeneity in adult zebrafish at single-cell resolution using RNA sequencing. *J Exp*
519 *Med*. Oct 2 2017;214(10):2875-2887. doi:10.1084/jem.20170976

520 34. Athanasiadis EI, Botthof JG, Andres H, Ferreira L, Lio P, Cvejic A. Single-cell
521 RNA-sequencing uncovers transcriptional states and fate decisions in
522 haematopoiesis. *Nat Commun*. Dec 11 2017;8(1):2045.
523 doi:10.1038/s41467-017-02305-6 [pii]
524 2305 [pii]
525 10.1038/s41467-017-02305-6
526 35. Wang J, Sun Q, Morita Y, et al. A differentiation checkpoint limits hematopoietic
527 stem cell self-renewal in response to DNA damage. *Cell*. Mar 2 2012;148(5):1001-14.
528 doi:S0092-8674(12)00145-6 [pii]
529 10.1016/j.cell.2012.01.040
530 36. Pauklin S, Vallier L. The cell-cycle state of stem cells determines cell fate
531 propensity. *Cell*. Sep 26 2013;155(1):135-47. doi:S0092-8674(13)01025-8 [pii]
532 10.1016/j.cell.2013.08.031
533 37. Cantor AB, Orkin SH. Coregulation of GATA factors by the Friend of GATA (FOG)
534 family of multitype zinc finger proteins. *Semin Cell Dev Biol*. Feb 2005;16(1):117-28.
535 doi:S1084-9521(04)00102-8 [pii]
536 10.1016/j.semcdb.2004.10.006
537 38. Grass JA, Boyer ME, Pal S, Wu J, Weiss MJ, Bresnick EH. GATA-1-dependent
538 transcriptional repression of GATA-2 via disruption of positive autoregulation and
539 domain-wide chromatin remodeling. *Proc Natl Acad Sci U S A*. Jul 22
540 2003;100(15):8811-6. doi:10.1073/pnas.1432147100
541 1432147100 [pii]
542 39. Pal S, Cantor AB, Johnson KD, et al. Coregulator-dependent facilitation of
543 chromatin occupancy by GATA-1. *Proc Natl Acad Sci U S A*. Jan 27
544 2004;101(4):980-5. doi:10.1073/pnas.0307612100
545 0307612100 [pii]
546 40. Bresnick EH, Lee HY, Fujiwara T, Johnson KD, Keles S. GATA switches as
547 developmental drivers. *J Biol Chem*. Oct 8 2010;285(41):31087-93.
548 doi:S0021-9258(19)88840-3 [pii]
549 10.1074/jbc.R110.159079
550 41. Monika Belickova M, Merkerova MD, Votavova H, et al. Up-regulation of
551 ribosomal genes is associated with a poor response to azacitidine in myelodysplasia
552 and related neoplasms. *Int J Hematol*. Nov 2016;104(5):566-573.

553 doi:10.1007/s12185-016-2058-3
554 10.1007/s12185-016-2058-3 [pii]
555 42. Nakayama Y, Yamaguchi N. Multi-lobulation of the nucleus in prolonged S phase
556 by nuclear expression of Chk tyrosine kinase. *Exp Cell Res*. Apr 1
557 2005;304(2):570-81. doi:S0014-4827(04)00703-7 [pii]
558 10.1016/j.yexcr.2004.11.027
559 43. Ullah Z, Lee CY, Depamphilis ML. Cip/Kip cyclin-dependent protein kinase
560 inhibitors and the road to polyploidy. *Cell Div*. Jun 2 2009;4:10. doi:1747-1028-4-10
561 [pii]
562 10.1186/1747-1028-4-10
563 44. Do C, Xing Z, Yu YE, Tycko B. Trans-acting epigenetic effects of chromosomal
564 aneuploidies: lessons from Down syndrome and mouse models. *Epigenomics*. Feb
565 2017;9(2):189-207. doi:10.2217/epi-2016-0138
566 45. Brault V, Duchon A, Romestaing C, et al. Opposite phenotypes of muscle
567 strength and locomotor function in mouse models of partial trisomy and monosomy
568 21 for the proximal Hspa13-App region. *PLoS Genet*. Mar 2015;11(3):e1005062.
569 doi:PGENETICS-D-14-01836 [pii]
570 10.1371/journal.pgen.1005062
571 46. Haferlach T, Nagata Y, Grossmann V, et al. Landscape of genetic lesions in 944
572 patients with myelodysplastic syndromes. *Leukemia*. Feb 2014;28(2):241-7.
573 doi:leu2013336 [pii]
574 10.1038/leu.2013.336
575 47. Rossi A, Kontarakis Z, Gerri C, et al. Genetic compensation induced by
576 deleterious mutations but not gene knockdowns. *Nature*. Aug 13
577 2015;524(7564):230-3. doi:nature14580 [pii]
578 10.1038/nature14580
579
580
581

582 **Figure legends**

583 ***Figure. 1: Gata2b$^{+/-}$ kidney marrow shows erythroid and myeloid dysplasia***

584 (A) Representative pictures of kidney marrow smears after May–Grünwald-Giemsa staining
585 of WT KM smears (panel 1 and 2) and *gata2b*$^{+/-}$ KM smears (panel 3-10). 3) Binucleated
586 erythroblast; 4) Blebbing in cytoplasm of proerythroblast; 5) Binucleated promyelocyte; 6)
587 Lobed nucleus and micronucleus in erythroblast; 7) Multinucleated promyelocyte; 8)
588 Irregular cytoplasm in erythroid precursor; 9) Lobed nucleus in erythrocytes of sorted cell
589 after cytospin; 10) Blebbing in cytoplasm of blast of sorted cell after cytospin.

590 (B) Frequency of dysplastic cells of the erythroid, myeloid and lymphoid lineage in KM
591 smears of WT (n=8) and *gata2b*$^{+/-}$ (n=8) zebrafish.

592

593 ***Figure. 2: Single cell transcriptome and chromatin accessibility analysis reveals a***
594 ***differentiation block in gata2b$^{+/-}$ KM***

595 (A) The flow cytometry-based sorting strategy for kidney marrow cells for scRNA-seq and
596 snATAC-seq analysis.

597 (B) UMAP analysis of scRNA-seq (top panels) and snATAC-seq (bottom panels) showing all
598 cell types, which was 19 cell types in scRNA-seq and 16 cell types in snATAC-seq. UMAP atlas
599 was separated between WT and *gata2b*$^{+/-}$ cells.

600 (C) Feature plot showing the scRNA-seq expression of marker gene *hemoglobin beta adult 1*
601 (*hbba1*) in the erythroid lineage, *granulin1* (*grn1*) in the myeloid lineage, *cluster of*
602 *differentiation 37* (*CD37*) in the B cell lineage, and *GATA binding protein 2b* (*gata2b*) in
603 HSPCs to show cluster annotation.

604 (D) Feature plot showing the snATAC-seq gene activity of *hbba1* in the erythroid lineage,
605 *grn1* in the myeloid lineage, *CD37* in the B lineage and *gata2b* in HSPCs to show cluster
606 annotation.
607 (E) Coverage plot illustrating the chromatin accessibility peaks of *hbba1* in the erythroid
608 lineage, *grn1* in the myeloid lineage, *CD37* in the B lineage, and *gata2b* in hematopoietic
609 stem and progenitor cells (HSPCs) to show cluster annotation.
610 (F) Cluster distribution quantification of scRNA-seq (left panel) and snATAC-seq (right panel)
611 between WT and *gata2b*$^{+/-}$ cells, with significantly differentially distributed clusters
612 represented in pink. A statistical significance threshold of FDR < 0.05 and log2 fold change >
613 1 were applied to determine the significance of the observed differences.
614 (G) UMAP plot colored by pseudotime scores representing cell developmental trajectories
615 split by WT and *gata2b*$^{+/-}$ cells. Lower pseudotime scores correspond to the root of
616 trajectories, while higher pseudotime scores represent the terminal differentiation
617 trajectories. The lines plotted on the UMAP atlas illustrate the differentiation paths of cells,
618 while the circled numbers indicate the location and order of the branch points along the cell
619 differentiation trajectories.
620
621 ***Figure. 3: Differences in chromatin accessibility of Gata2b result in upregulation of Gata2b***
622 ***expression in gata2b$^{+/-}$ HSPCs***
623 (A) Coverage plots showing significant differences of *gata2b* chromatin accessibility
624 between WT and *gata2b*$^{+/-}$ cells. The normalized peak signal range is 0 to 40. Highlighted
625 genomic regions within grey blocks indicate the location of two enhancers, enhancer 1 (E1)

626 and enhancer 2 (E2). The gene block depicts gene bodies contained within the plotted
627 genomic regions, *gata2b* is depicted in blue. The link between the blocks illustrate predicted
628 cis-regulatory interactions between the *gata2b* gene body and its upstream region, based
629 on examination of genome co-accessibility. Link colors correspond to predicted
630 co-accessibility scores, with yellow indicating stronger predicted interactions.
631 (B-G) UMAP feature plots depict the open chromatin of Enhancer 2 (E2; B and C), Enhancer
632 1 (E1; D and E), and transcription start site (TSS; F and G) of *gata2b*. UMAP atlas were
633 separated by WT (left panel) and *gata2b*$^{+/-}$ cells (right panel).
634 (H) Open chromatin of the *gata2b* locus per HSPC cluster cell comparing WT and *gata2b*$^{+/-}$
635 HSPCs.
636 (I) Violin plot displaying the expression level of *gata2b* in HSPCs in WT (pink) and *gata2b*$^{+/-}$
637 (turqois).
638
639 ***Figure 4. Functional consequences in HSPCs upon Gata2b overexpression***
640 (A) The dot plot displays the log2 fold change values of differentially activated transcription
641 factor (TF) motifs in progenitor clusters, including "HSPCs", "progenitors", and "HSPCs and
642 thrombocytes", between WT and *gata2b*$^{+/-}$ cells. The TF motifs are ranked in descending
643 order based on log2 fold change values. Dotted lines indicate the threshold for significance,
644 with absolute log2 fold change values > 0.58. The blue and red colors represent TF motifs
645 that are statistically significant (P value < 0.05) with absolute log2 fold change values > 0.58.
646 (B) Representative figure of identification of CD41:GFP^{int} population (in green) and
647 distribution in the FSC-A SSC-A kidney marrow population.

648 (C) Quantification of the percentage of CD41:GFP^{int} cells in WT (n=12) and *gata2b*$^{+/-}$ (n=12)
649 KM single live cells.
650 (D) Representative flow cytometry plots of cycle analysis by anti-Ki67 and DAPI staining in
651 WT and *gata2b*$^{+/-}$ CD41:GFP^{int} KM.
652 (E) Quantification of the percentages of WT (n=3) and *gata2b*$^{+/-}$ (n=3) CD41:GFP^{int} cells in
653 different cell cycle stages (* P value<0.5). Data represents as mean ± standard error of the
654 mean.
655
656 ***Figure. 5: Epigenetic changes resulting in the erythroid differentiation block in gata2$^{+/-}$ KM***
657 (A) UMAP feature plots depicting open chromatin of GATA2 motif sequence indicating
658 activity of GATA2. UMAP atlas were separated by WT (left panel) and *gata2b*$^{+/-}$ cells (right
659 panel). The top left corner of UMAP plots display the TF sequence logo corresponding to the
660 analyzed motif, generated by the function *MotifPlot*.
661 (B) Coverage plots showing differences of *gata1a* chromatin accessibility between WT and
662 *gata2b*$^{+/-}$ cells.
663 (C) Percentage of cells derived from WT and *gata2b*$^{+/-}$ KM with accessibility of the *gata1a*
664 locus across clusters.
665 (D) Percentage of cells derived from WT and *gata2b*$^{+/-}$ KM expressing *gata1a* across clusters.
666 (E) UMAP plots depict the activity of combined motif GATA1::TAL1. UMAP atlas were
667 separated by WT (left panel) and *gata2b*$^{+/-}$ cells (right panel). The top left corner of the
668 UMAP plots display the TF sequence logo corresponding to the analyzed motif, generated by
669 the function *MotifPlot*.

670 (F) Feature plot showing the *zfpm1* (FOG1) expression in WT and *gata2b*$^{+/-}$ KM.

671 (G-H) UMAP plots of the subclustering of the erythroid progenitors 2 cluster in scRNA-seq

672 data. UMAP atlas were separated by WT(G) and *gata2b*$^{+/-}$ cells (H)

673 (I) Quantification of distribution between WT and *gata2b*$^{+/-}$ erythroid progenitor 2

674 subclusters. Significantly differentially distributed clusters in pink. Dotted box indicates

675 overrepresented subcluster in *gata2b*$^{+/-}$ cells.

676 (J) STRING network of upregulated and downregulated transcripts in *gata2b*$^{+/-}$ "erythroid

677 progenitors 2". Only networks with more than 2 interactions were represented. Highlighted

678 in red DNA replication genes from KEGG pathways and highlighted in blue mitotic

679 cytokinesis genes from biological processes (Gene Ontology).

680 (K-L) Lineage trajectory analysis of WT (K) and *gata2b*$^{+/-}$ (L) erythroid progenitor 2

681 subclusters. Note the difference in pseudo-time scale between the WT and *gata2b*$^{+/-}$

682 trajectories.

683 (M-R) WT (M-O) and *gata2b*$^{+/-}$ (P-R) pseudo-time expression of individual genes *hemoglobin*

684 *beta adult 1* (*hbba1*) (M and P), *GATA binding protein 1 alpha (gata1a)* (N and Q), and *mki67*

685 (O and R). Fold change > 0.05 & adjusted P value < 0.05. FDR=False discovery rate. FD=Fold

686 difference.

687 (S) Dot plot displaying log2 fold change values of differentially activated TF motifs in the

688 erythroid progenitor 2 cluster between WT and *gata2b*$^{+/-}$ cells in snATAC-seq data. The TF

689 motifs are ranked in descending order based on log2 fold change values. Dotted lines

690 indicate the threshold (absolute log2 fold change < -0.58 and >0.58) for significance, The

691 blue (downregulated) and red (upregulated) colors represent TFs that are statistically

692 significant (P value < 0.05).

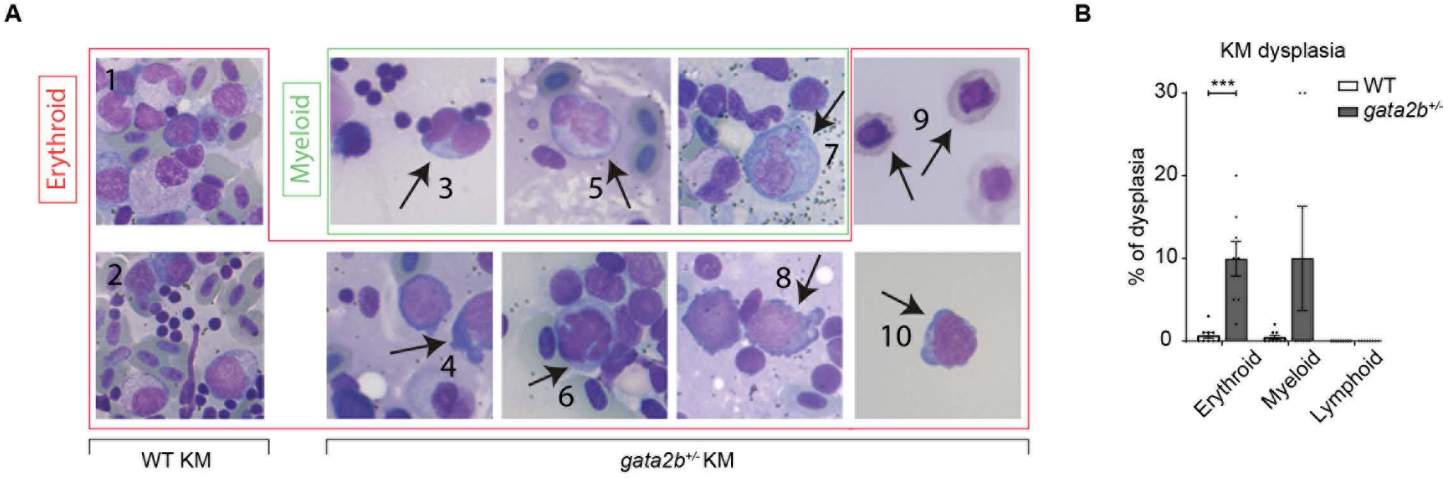

A
Erythroid
Myeloid
1
2
3
4
5
6
7
8
9
10
WT KM
gata2b+/- KM
B
KM dysplasia
WT
gata2b+/-
% of dysplasia
30
20
10
0
***
Erythroid
Myeloid
Lymphoid

Figure 1

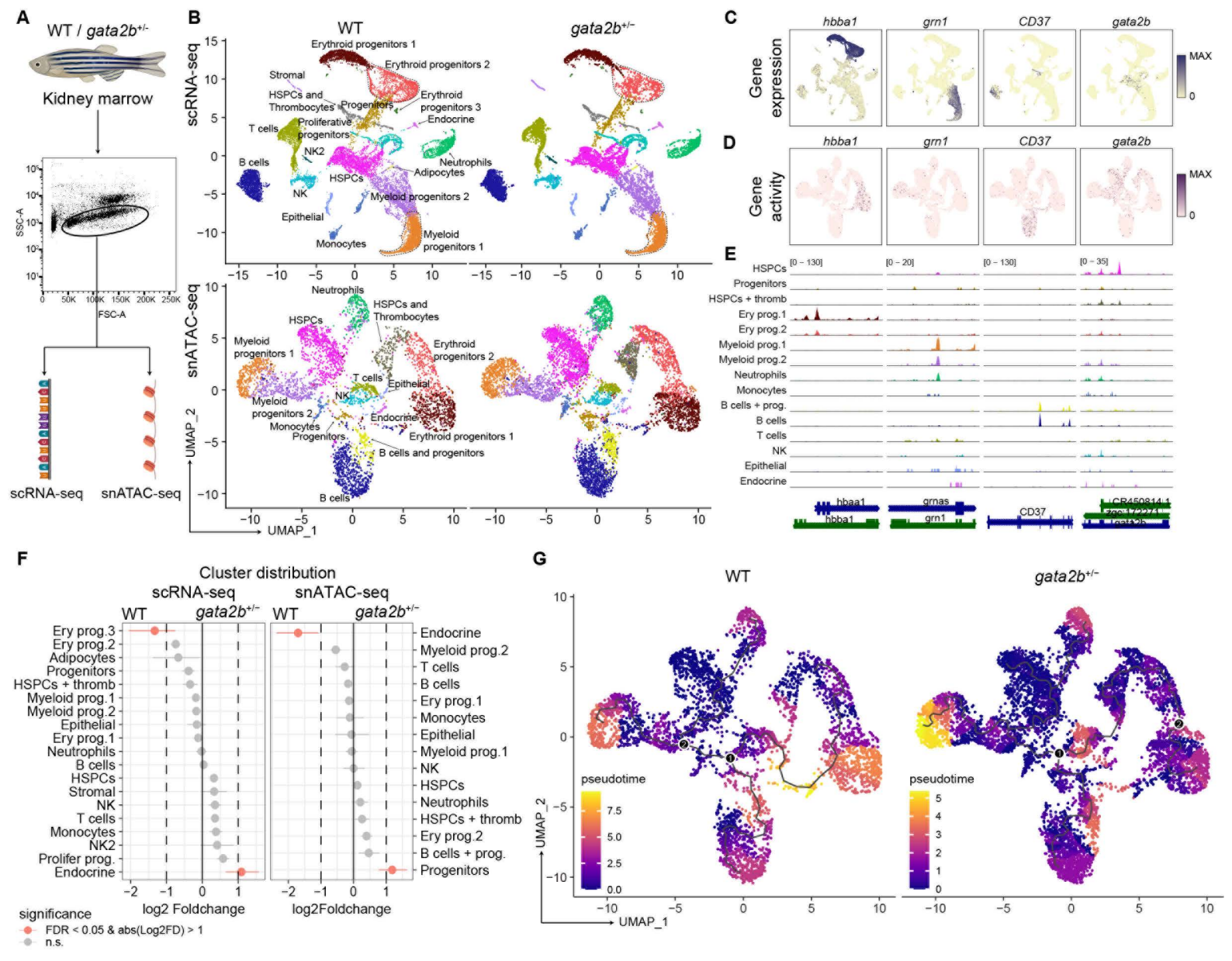

A
WT / gata2b+/-
Kidney marrow
SSC-A
FSC-A
scRNA-seq
snATAC-seq
B
WT
gata2b+/-
scRNA-seq
snATAC-seq
UMAP_1
UMAP_2
Erythroid progenitors 1
Erythroid progenitors 2
Erythroid progenitors 3
Stromal
HSPCs and Thrombocytes
Progenitors
Endocrine
T cells
Proliferative progenitors
NK2
B cells
HSPCs
Neutrophils
Adipocytes
NK
Myeloid progenitors 2
Epithelial
Monocytes
Myeloid progenitors 1
HSPCs and Thrombocytes
Erythroid progenitors 2
T cells
Epithelial
Endocrine
Erythroid progenitors 1
B cells and progenitors
C
Gene expression
hbba1
grn1
CD37
gata2b
MAX
0
D
Gene activity
E
HSPCs
Progenitors
HSPCs + thromb
Ery prog.1
Ery prog.2
Myeloid prog.1
Myeloid prog.2
Neutrophils
Monocytes
B cells + prog.
B cells
T cells
NK
Epithelial
Endocrine
[0 – 130]
[0 – 20]
[0 – 130]
[0 – 35]
hbaa1
hbba1
grnas
grn1
CD37
F
Cluster distribution
scRNA-seq
snATAC-seq
WT
gata2b+/-
Ery prog.3
Ery prog.2
Adipocytes
Progenitors
HSPCs + thromb
Myeloid prog.1
Myeloid prog.2
Epithelial
Ery prog.1
Neutrophils
B cells
HSPCs
Stromal
NK
T cells
Monocytes
NK2
Prolifer prog.
Endocrine
Myeloid prog.2
T cells
B cells
Ery prog.1
Monocytes
Epithelial
Myeloid prog.1
NK
HSPCs
Neutrophils
HSPCs + thromb
Ery prog.2
B cells + prog.
Progenitors
log2 Foldchange
significance
FDR < 0.05 & abs(Log2FD) > 1
n.s.
G
WT
gata2b+/-
pseudotime
UMAP_1
UMAP_2

Figure 2

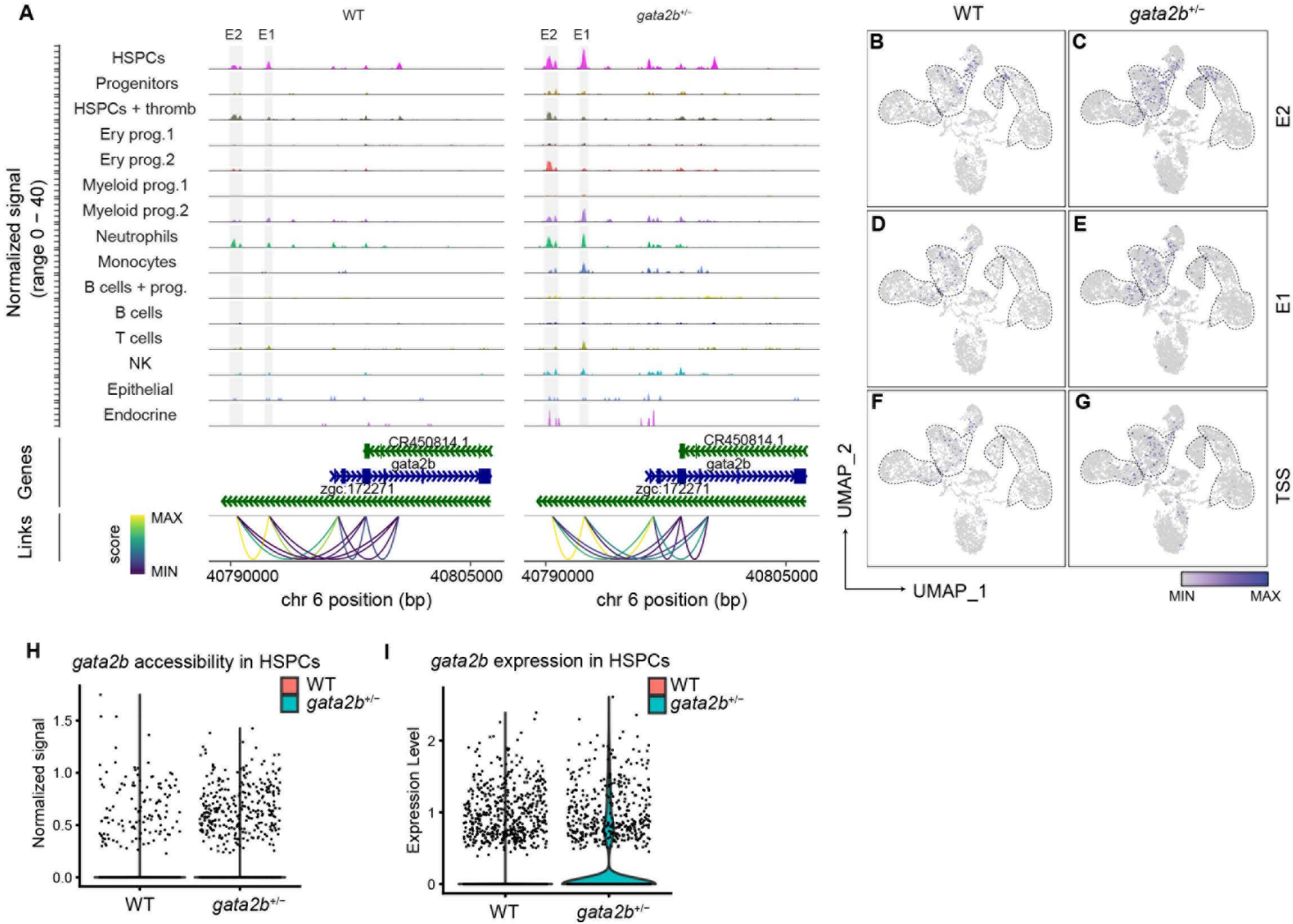

A
WT
gata2b+/-
E2
E1
HSPCs
Progenitors
HSPCs + thromb
Ery prog.1
Ery prog.2
Myeloid prog.1
Myeloid prog.2
Neutrophils
Monocytes
B cells + prog.
B cells
T cells
NK
Epithelial
Endocrine
Normalized signal (range 0 – 40)
Genes
Links
CR450814.1
gata2b
zgc:172271
score
MAX
MIN
40790000
40805000
chr 6 position (bp)
B
C
D
E
F
G
E2
E1
TSS
UMAP_2
UMAP_1
MIN
MAX
H
gata2b accessibility in HSPCs
Normalized signal
1.5
1.0
0.5
0.0
WT
gata2b+/-
I
gata2b expression in HSPCs
Expression Level
2
1
0
WT
gata2b+/-

Figure 3

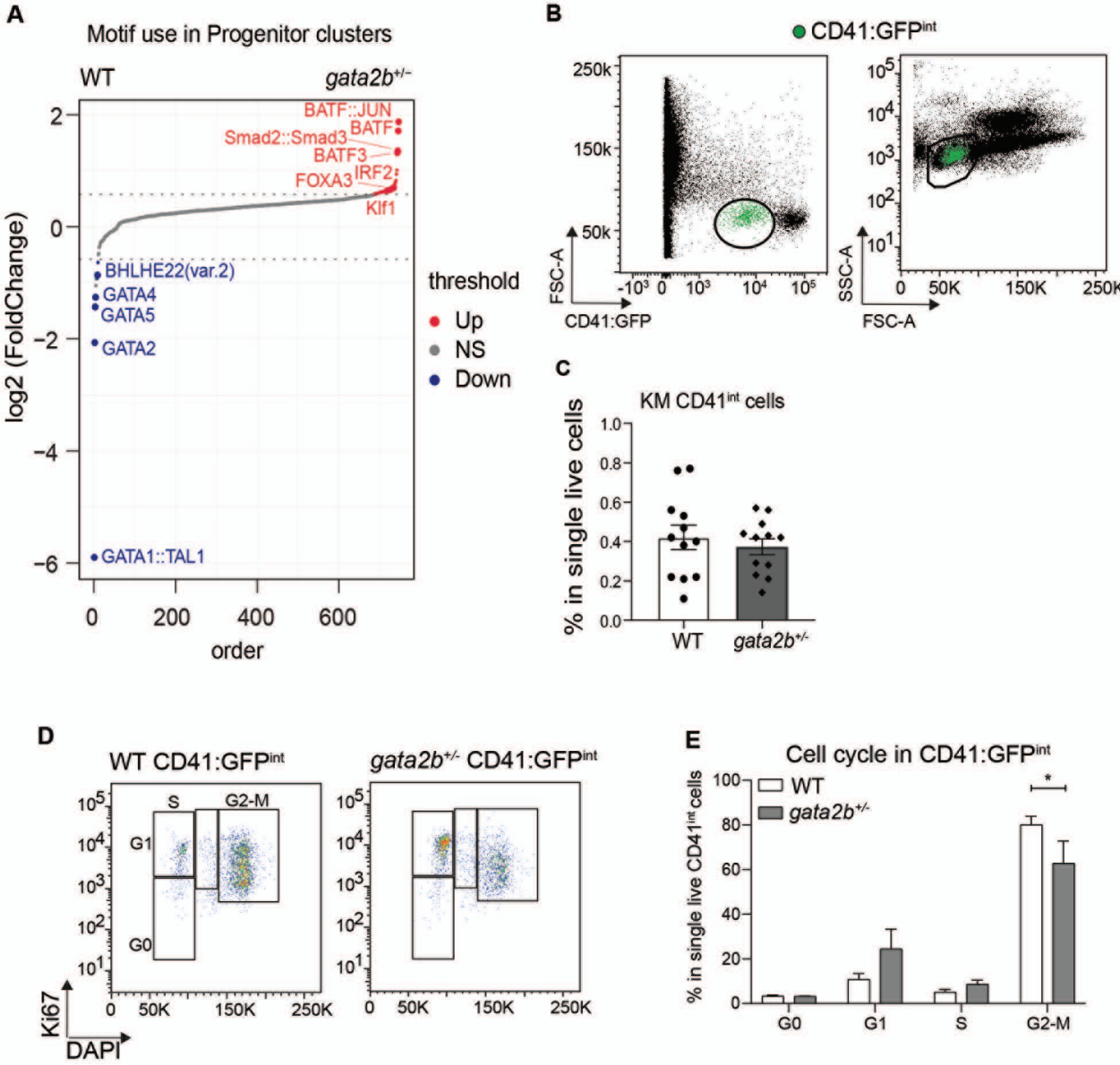

A
Motif use in Progenitor clusters
WT
gata2b+/-
BATF::JUN
BATF
Smad2::Smad3
BATF3
FOXA3
IRF2
Klf1
BHLHE22(var.2)
GATA4
GATA5
GATA2
GATA1::TAL1
log2 (FoldChange)
order
threshold
Up
NS
Down
B
CD41:GFPint
FSC-A
CD41:GFP
SSC-A
FSC-A
C
KM CD41int cells
% in single live cells
WT
gata2b+/-
D
WT CD41:GFPint
gata2b+/- CD41:GFPint
S
G2-M
G1
G0
Ki67
DAPI
E
Cell cycle in CD41:GFPint
WT
gata2b+/-
% in single live CD41int cells
G0
G1
S
G2-M

Figure 4

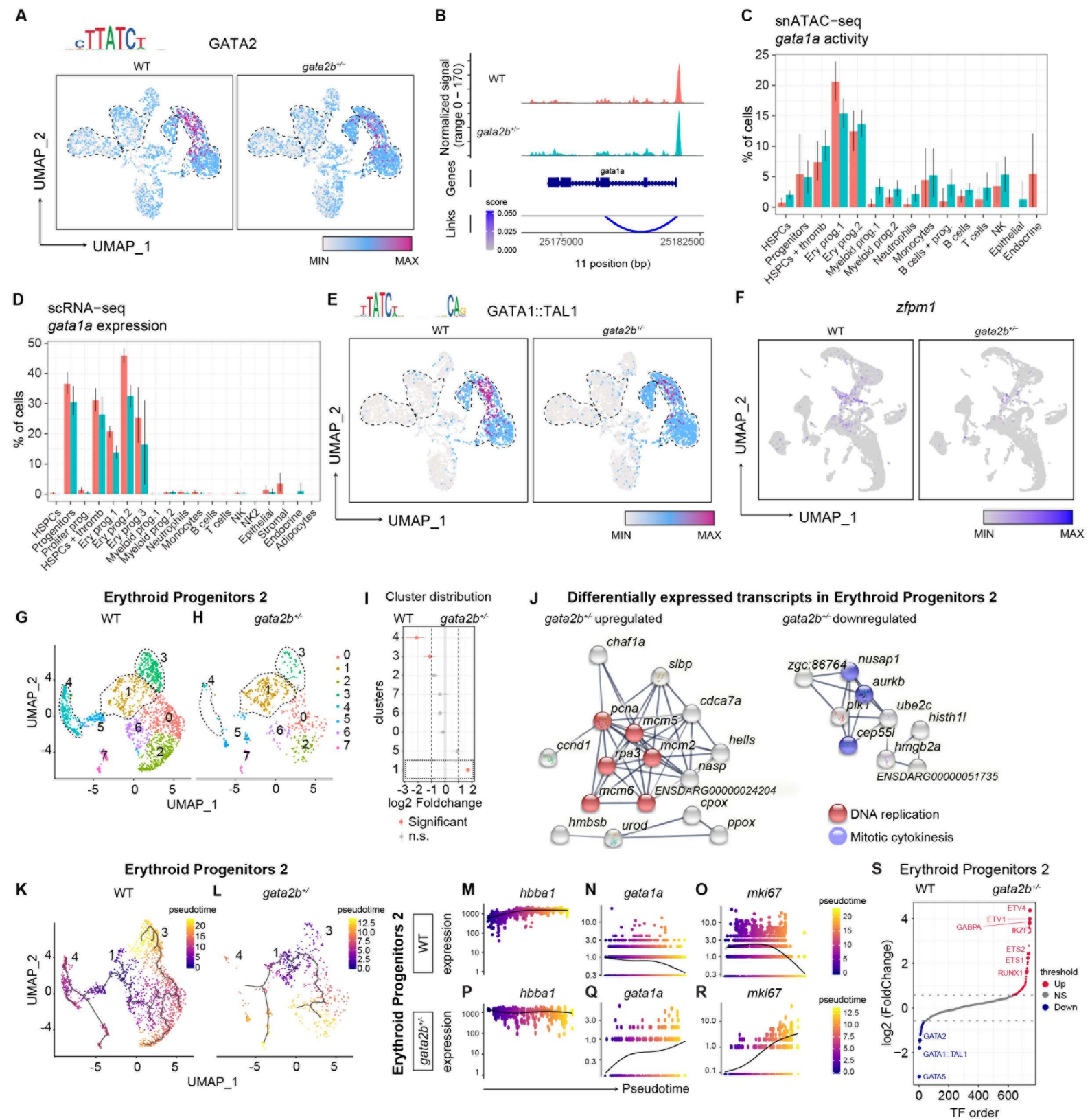
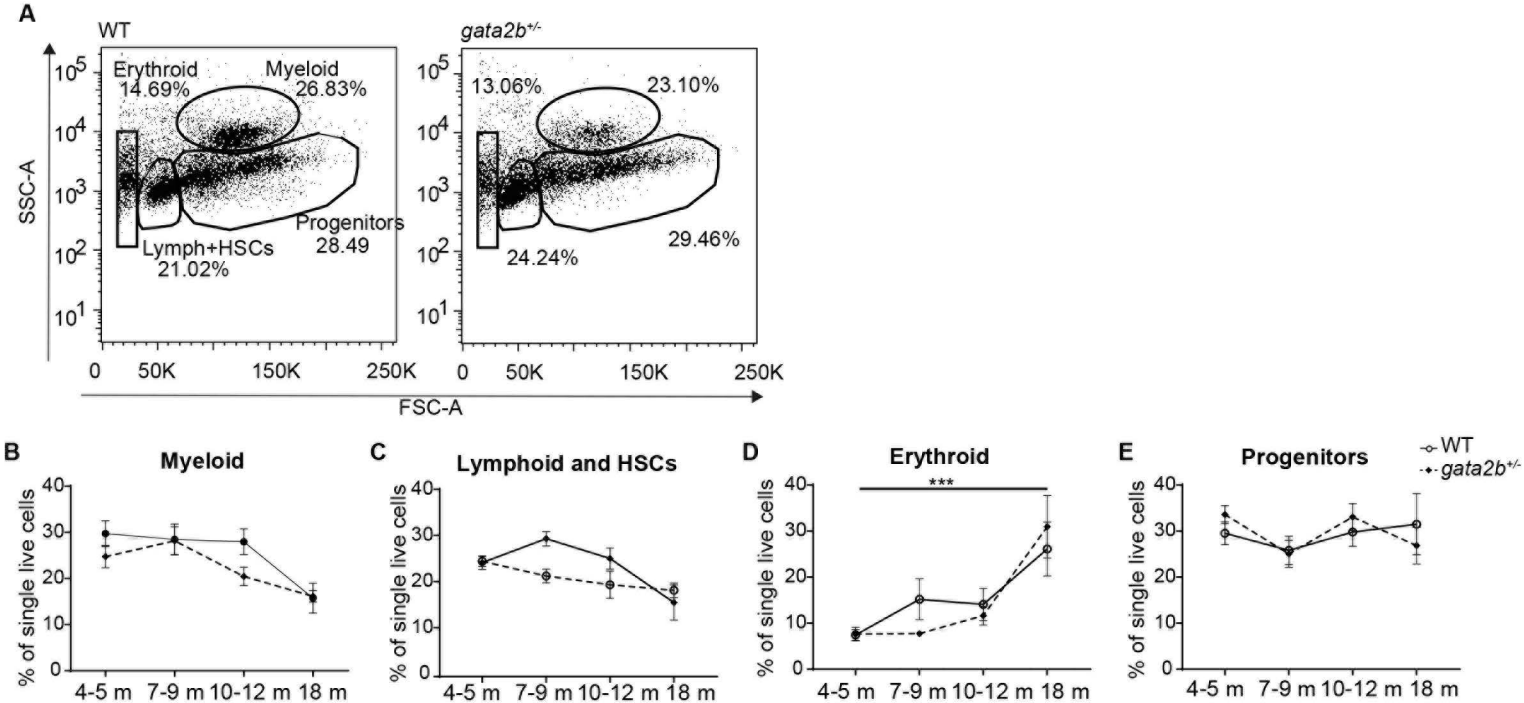
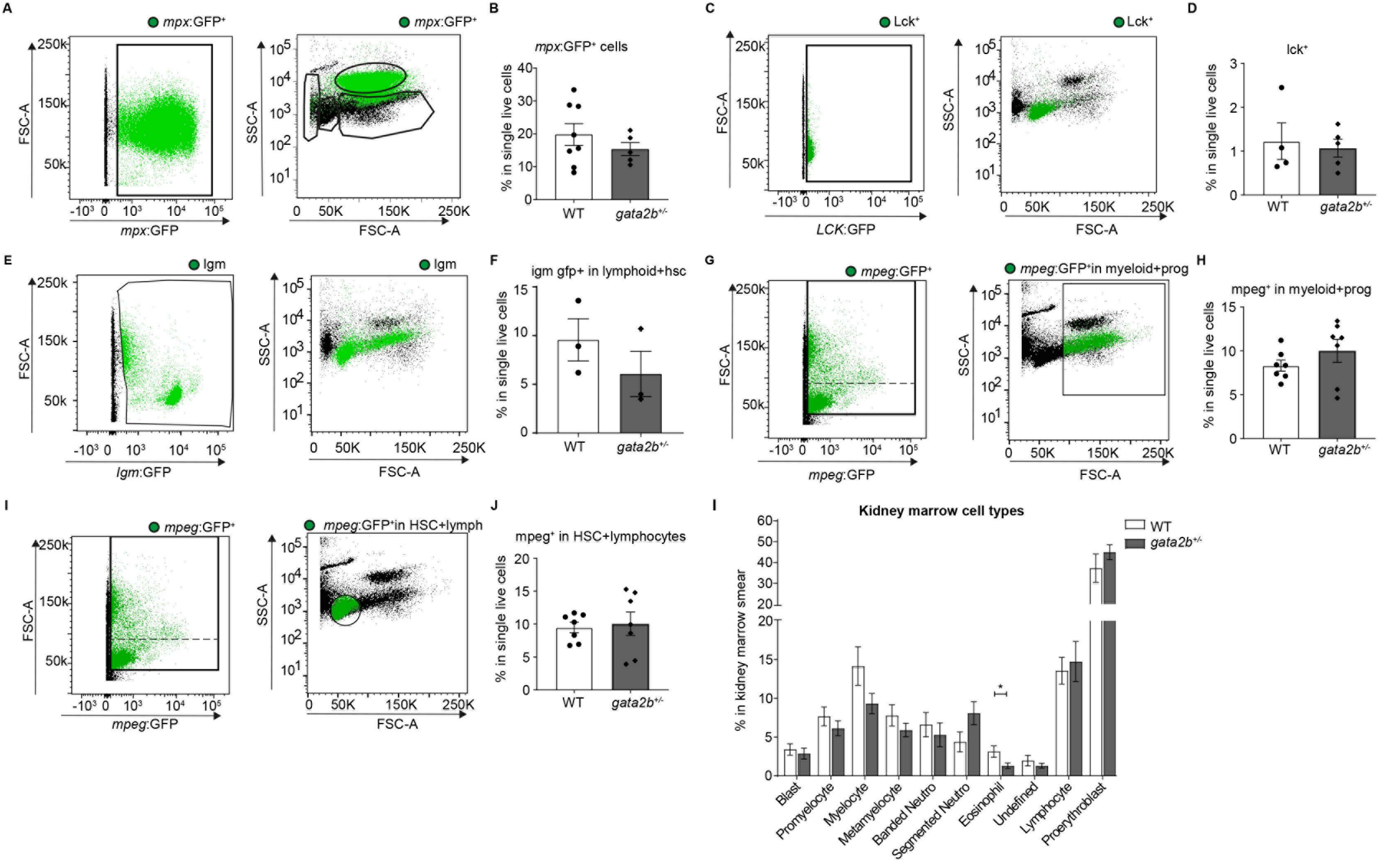

A
GATA2
WT
gata2b+/-
UMAP_2
UMAP_1
MIN
MAX
B
Normalized signal (range 0 – 170)
WT
gata2b+/-
Genes
gata1a
Links
score
0.050
0.025
0.000
25175000
25182500
11 position (bp)
C
snATAC-seq
gata1a activity
% of cells
HSPCs
Progenitors
HSPCs + thromb
Ery prog.1
Ery prog.2
Myeloid prog.1
Myeloid prog.2
Neutrophils
Monocytes
B cells + prog.
B cells
T cells
NK
Epithelial
Endocrine
D
scRNA-seq
gata1a expression
% of cells
HSPCs
Progenitors
Proliter prog.
HSPCs + thromb
Ery prog.1
Ery prog.2
Ery prog.3
Myeloid prog.1
Myeloid prog.2
Neutrophils
Monocytes
B cells
T cells
NK
NK2
Epithelial
Stromal
Endocrine
Adipocytes
E
GATA1::TAL1
WT
gata2b+/-
UMAP_2
UMAP_1
MIN
MAX
F
zfpm1
WT
gata2b+/-
UMAP_2
UMAP_1
MIN
MAX
Erythroid Progenitors 2
G
WT
H
gata2b+/-
UMAP_2
UMAP_1
I
Cluster distribution
WT
gata2b+/-
clusters
log2 Foldchange
Significant
n.s.
J
Differentially expressed transcripts in Erythroid Progenitors 2
gata2b+/- upregulated
gata2b+/- downregulated
chaf1a
slbp
cdca7a
pcna
mcm5
hells
ccnd1
rpa3
mcm2
nasp
mcm6
ENSDARG00000024204
cpox
hmbsb
urod
ppox
zgc:86764
nusap1
aurkb
plk1
ube2c
histh1l
cep55l
hmgb2a
ENSDARG00000051735
DNA replication
Mitotic cytokinesis
Erythroid Progenitors 2
K
WT
L
gata2b+/-
pseudotime
UMAP_2
UMAP_1
Erythroid Progenitors 2
WT
gata2b+/-
expression
M
hbba1
N
gata1a
O
mki67
P
hbba1
Q
gata1a
R
mki67
Pseudotime
S
Erythroid Progenitors 2
WT
gata2b+/-
log2 (FoldChange)
TF order
ETV4
ETV1
GABPA
IKZF1
ETS2
ETS1
RUNX1
GATA2
GATA1::TAL1
GATA5
threshold
Up
NS
Down

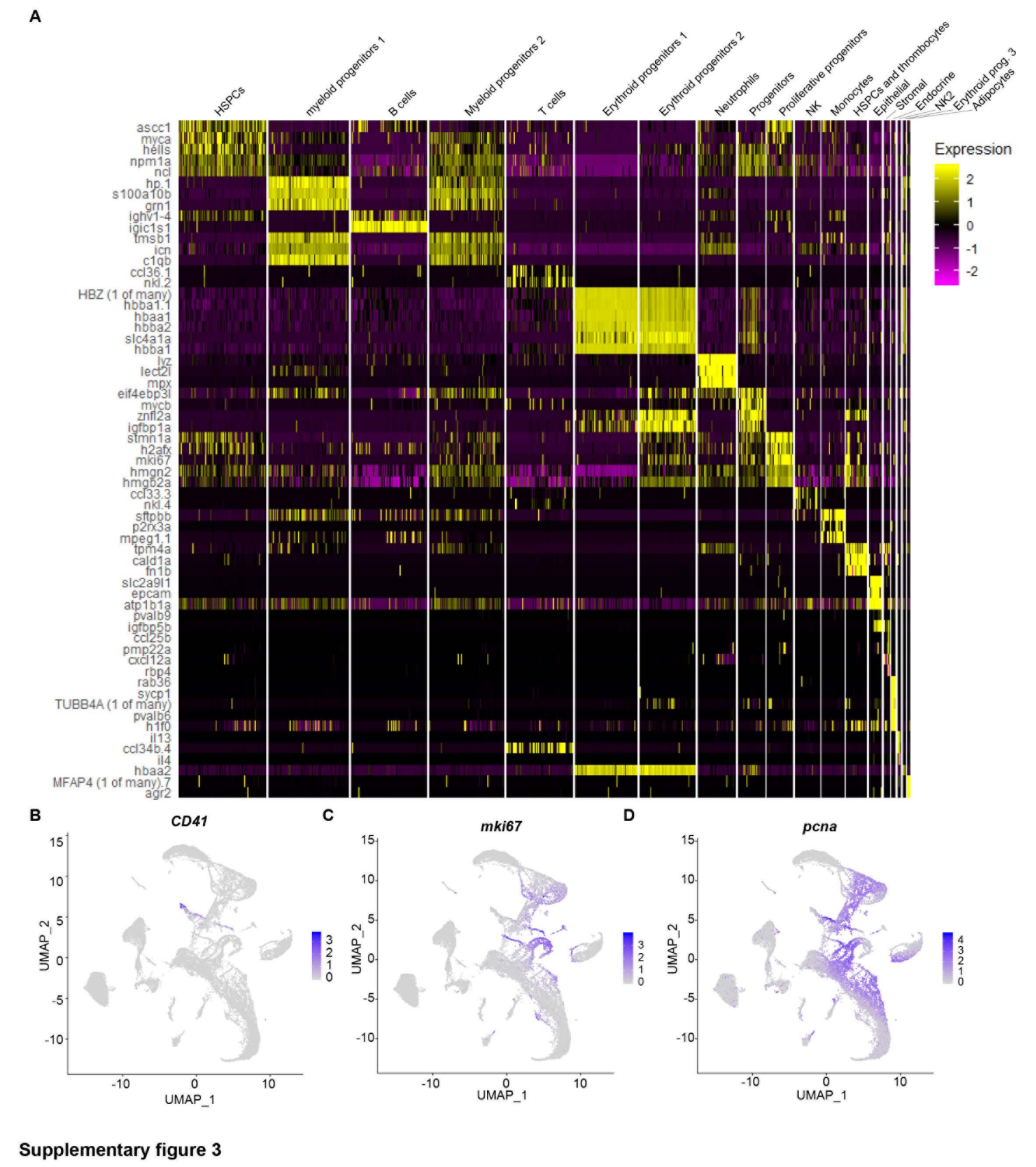
Figure 5



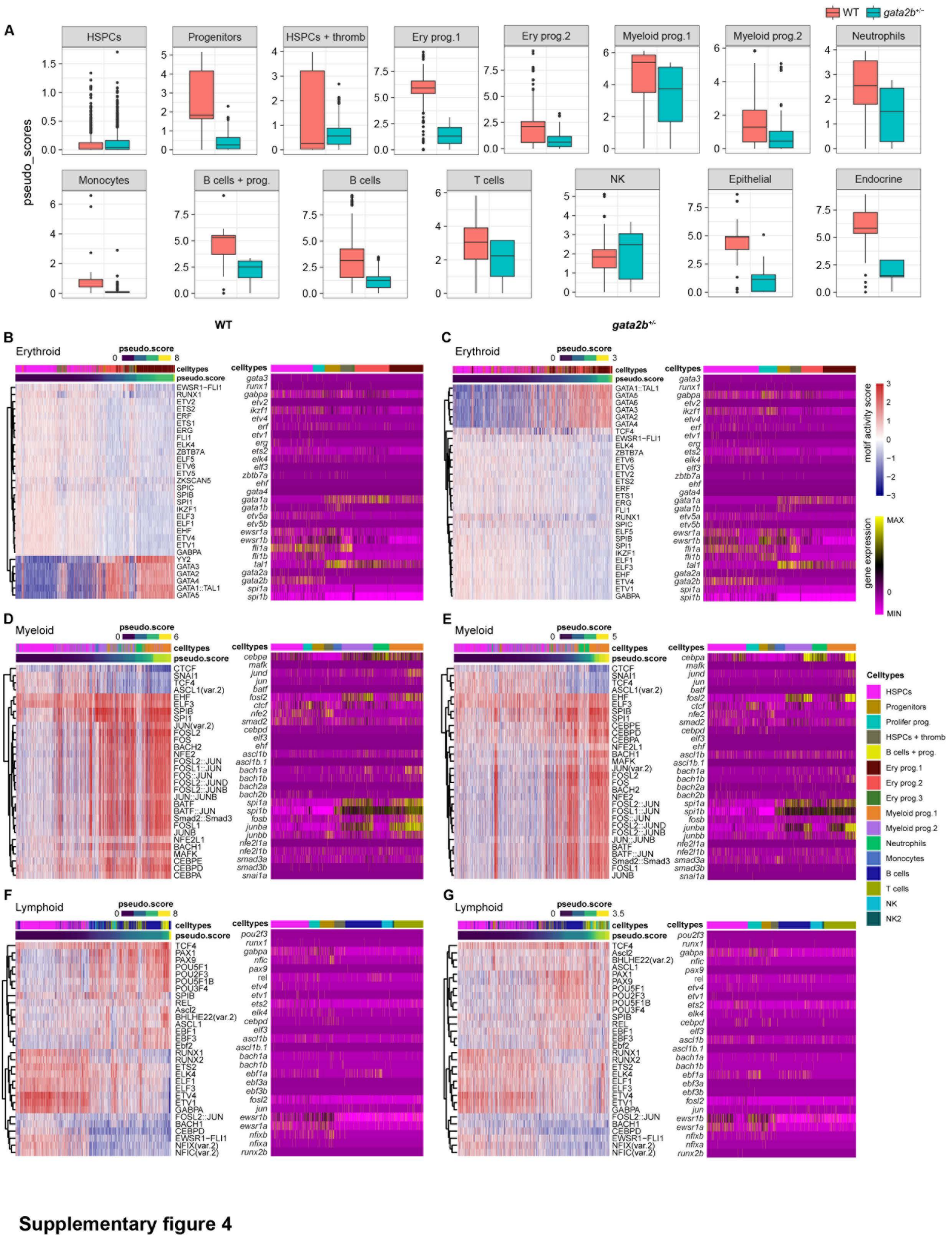
Supplementary figure 1



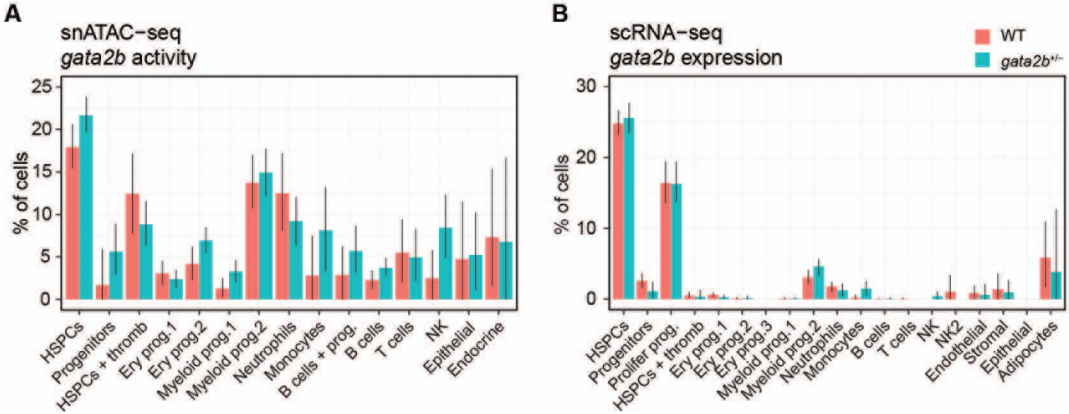
Supplementary figure 2



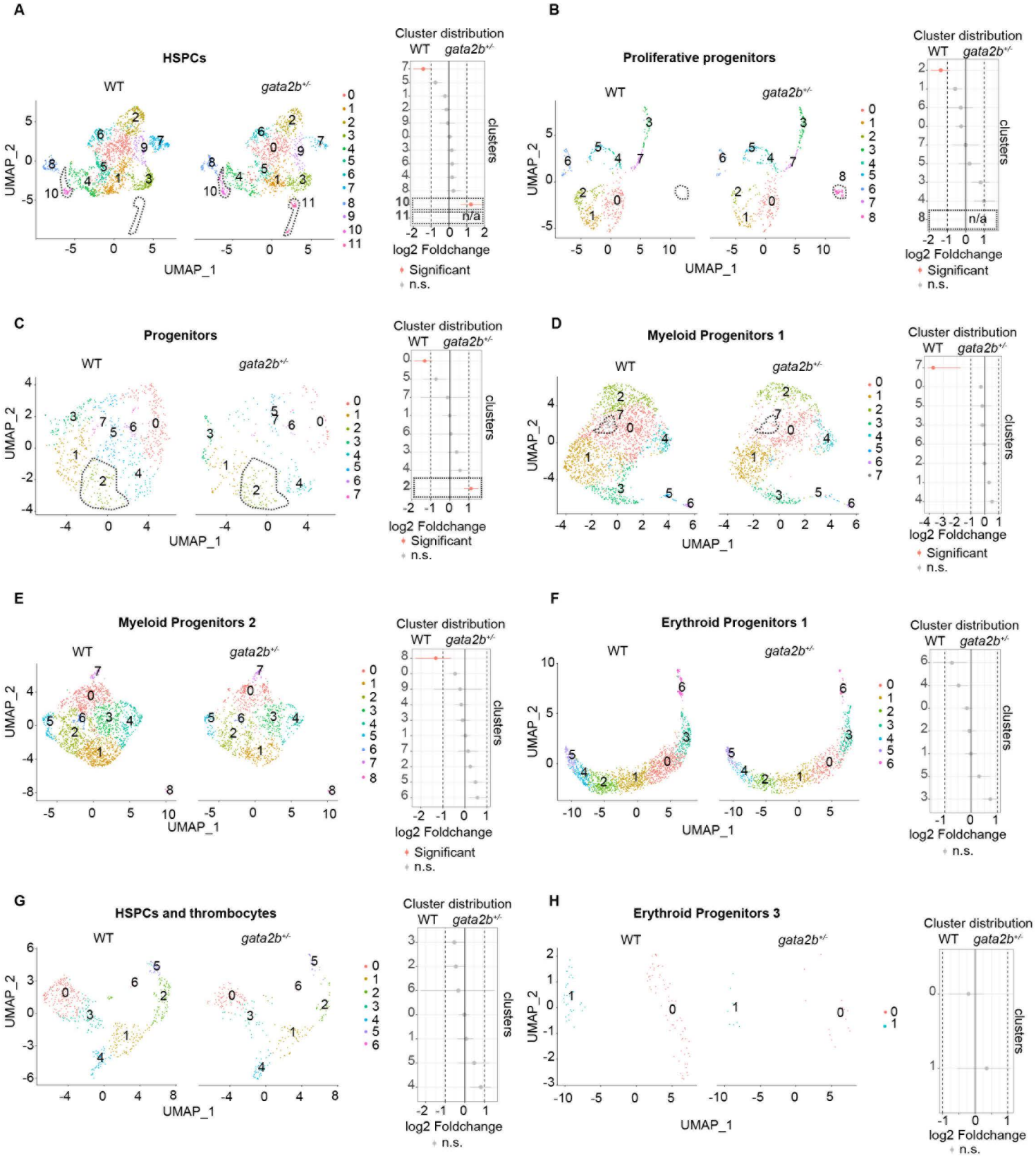
Supplementary figure 3



Supplementary figure 4



Supplementary Figure 5



Supplementary figure 6