GATA2 haploinsufficiency causes an epigenetic feedback mechanism resulting in myeloid and erythroid dysplasia

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

[*]equally contributed

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 delay 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) showing 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 the mechanisms causing the different aspects of GATA2 deficiency syndromes largely

62 undiscovered.

63 To better understand the biology of GATA2 haploinsufficiency syndromes, zebrafish
64 serves as an attractive alternative model. Zebrafish have the advantage of having two *GATA2*
65 orthologues; *Gata2a* and *Gata2b*. *Gata2a* is expressed predominantly in the vasculature[14]
66 and is required for programming of the hemogenic endothelium[15,16]. *Gata2b* is expressed in
67 hematopoietic stem/progenitor cells (HSPCs)[14] and homozygous deletion (*gata2b*$^{-/-}$)
68 redirects HSPC differentiation to the lymphoid lineage in expense of myeloid differentiation
69 causing a lymphoid bias with an incomplete B-cell differentiation in the kidney marrow (KM),
70 thus mimicking one of the GATA2 haploinsufficiency phenotypes found in patients.
71 Additionally, the most primitive HSC compartment was lost in *gata2b*$^{-/-}$ KM, but none of the
72 animals displayed signs of dysplasia[16,17]. Because patients carry heterozygous rather than
73 homozygous *GATA2* mutations, we specifically focused on how Gata2b haploinsufficiency
74 could be mechanistically linked to erythro-myelodysplasia, the major clinical hallmark of
75 GATA2 patients.

76 To validate the model, we first assessed hematopoietic cell differentiation in *gata2b*
77 heterozygous zebrafish (*gata2b*$^{+/-}$) KM and observed erythroid dysplasia in all *gata2b*$^{+/-}$ but
78 not in wild-type (WT) zebrafish and myeloid dysplasia in 25% of the *gata2b*$^{+/-}$ zebrafish.
79 Subsequent single-cell RNA and ATAC sequencing analysis revealed an auto-regulatory
80 feedback mechanism of Gata2b where *gata2b* chromatin was over-accessible, underlying
81 the hematopoietic lineage differentiation defects in adults. Further characterization of
82 *gata2b*$^{+/-}$ zebrafish showed reduction of HSC proliferation, failure to initiate the "GATA2 to
83 GATA1 switch" and depletion of *Gata1a* and its co-factor *FOG1* in erythroid progenitor cells.
84 Taken together, these results reveal a dosage-dependent function of *Gata2b* and provide a
85 plausible explanation for the hematopoietic defects observed in 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 1-6). On average 0.5% of WT erythroid cells showed dysplastic features, compared to
147 9.9% of *gata2b*$^{+/-}$ erythroid cells (Figure 1A and B), the latter representing 4.5% of the total
148 kidney marrow population of *gata2b*$^{+/-}$ zebrafish. Myeloid lineage dysplasia was seen in the
149 *gata2b*$^{+/-}$ KM in 25% of the fish (Figure 1A, panel 7 and 8, and B). In these samples, 30% of
150 myeloid cells were dysplastic compared to 0.3% in WT. Myeloid dysplasia was mostly
151 represented by multi-lobulated nuclei, whereas the erythroid abnormalities ranged from
152 nuclear deformities and double nuclei to irregular cytoplasm or an almost complete lack of
153 cytoplasm (Figure 1A). The remaining cell types were not affected morphologically by
154 Gata2b haploinsufficiency. These results indicate that *gata2b* heterozygosity induces
155 dysplasia, predominantly in erythroid and myeloid 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, no significant differences were found 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), suggesting aging results in an erythroid
169 biased hematopoietic system in zebrafish. To confirm the presence of a normal
170 hematopoietic lineage distribution in *gata2b*$^{+/-}$ kidney marrow we bred Gata2b-zebrafish
171 with transgenic GFP reporter zebrafish. Lineage analysis in transgenic lines specifically
172 marking neutrophils by *Tg(mpx:GFP)* (supplemental Figure 2A-B), T-cells by *Tg(lck:GFP)*
173 (supplemental Figure 2C-D), B-cells by *Tg(Igm:GFP)* (supplemental Figure 2E-F) and
174 *Tg(mpeg:GFP)* (supplemental Figure 2G-H) or monocytes by *Tg(mpeg:GFP)*[23,24]
175 (supplemental Figure 2I-J) showed no significant alterations in lineage distribution between
176 WT and *gata2b*$^{+/-}$ KM [25-28]. KM smear quantification by May–Grünwald-Giemsa (MGG)
177 staining revealed that *gata2b*$^{+/-}$ KM had a significant decrease in eosinophils compared to
178 WT (supplemental Figure 2K), representing less than 5% of the total KM cells. In summary,
179 based on flow cytometry analysis and transgenic reporters we conclude that the
180 differentiation in major hematopoietic lineages is not altered in *gata2b*$^{+/-}$ KM and that the
181 erythroid and myeloid dysplasia does not affect this.
182
183 *Single cell RNA- and ATAC sequencing analysis reveals a cell maturation delay in gata2b*$^{+/-}$
184 *KM*
185 To investigate the molecular mechanisms underlying dysplasia in *gata2b*$^{+/-}$ KM, we
186 assessed chromatin accessibility, a measure for gene activation, and transcriptional
187 differences in dysplastic cell populations. We flow-sorted 4 cell populations based on light
188 scatter and observed dysplastic cells in the progenitor and lymphoid + HSPCs population of
189 *gata2b*$^{+/-}$ KM, indicating that dysplastic cells could be viably sorted (Figure 1A, panel 6). We

190 did not identify a uniquely separated population of dysplastic cells, possibly caused by a
191 heterogeneity in their shape. Therefore, we sorted progenitor and lymphoid + HSPC
192 populations from WT and *gata2b*$^{+/-}$ KMs and performed single-cell (sc) RNA- and single
193 nucleus (sn) ATAC-sequencing to identify transcriptome and epigenetic regulation signatures
194 defining dysplastic cells (Figure 2A). After filtering out low quality cells, we obtained 28,996
195 cells (WT = 18,147, *gata2b*$^{+/-}$ = 10,849) for scRNA-seq analysis and 11,762 cells (WT = 4,190,
196 *gata2b*$^{+/-}$ = 7,572) for snATAC-seq analysis (Figure 2B).
197 scRNA-sequencing resulted in 19 clusters based on a nearest neighbor algorithm using
198 the R Seurat package[20] (Figure 2B). Each cluster was classified based on differentially
199 expressed genes (supplemental Figure 3A-D) and known differentiation markers[29-34] (Figure
200 2C). We then used this labeled scRNA-seq dataset as a reference to interpret and identify 16
201 clusters in the snATAC-seq dataset based on the R Signac package[21] (Figure 2B, supplemental
202 Figure 4A-C). Cluster proportion analysis indicated an overall similar distribution of cells
203 within clusters between *gata2b*$^{+/-}$ and WT (Figure 2C). However, a clear difference in
204 differentiation trajectory was observed between WT and *gata2b*$^{+/-}$ KM in scRNA-seq and
205 snATAC-seq data based on the R package Monocle3[22], and particularly note the difference in
206 pseudo-time scores between WT and *gata2b*$^{+/-}$ erythroid progenitors 1 and myeloid
207 progenitors 1, suggesting a delayed differentiation in *gata2b*$^{+/-}$ KM (Figure 2D and E,
208 supplemental figure 4D). Whereas the pseudo-time score for HSPCs was similar between WT
209 and *gata2b*$^{+/-}$ zebrafish, the pseudo-time scores of differentiation of erythroid and myeloid
210 lineages were reduced in *gata2b*$^{+/-}$ KM (Figure 2D and E). Moreover, the expression of
211 *gata2b* was reduced in *gata2b*$^{+/-}$ cells compared to WT in the terminal trajectory of
212 hematopoietic development, where erythroid and myeloid progenitors are located (Figure
213 2F and G). Based on these findings, we conclude that *gata2b*$^{+/-}$ KM showed an overall
214 lineage maturation delay. As expected, a clear distinction in open chromatin patterns of
215 lineage specific transcription factors (TF) between immature and mature cells was seen in
216 WT KM (Supplemental figure 5A, C and E, left panels). In contrast, this distinction was less
217 pronounced in *gata2b*$^{+/-}$ KM (Supplemental figure 5B, D and F, left panels). Specifically, the
218 ETV family of transcription factors and the GATA family transcription factors were affected in
219 erythroid differentiation, SPI1, SPIB, and CEBPA family of transcription factors were affected
220 in myeloid differentiation, and TCF4 was most affected in lymphoid differentiation and
221 showed significantly different patterns of activation (Supplemental figure 5A-F). Surprisingly,
222 this did not always lead to a reduction in gene expression (Supplemental figure5A-F, right

223 panels), particularly note the increased expression of *gata2b* in erythroid lineage
224 differentiation (Supplemental figure 5B, right panel).
225
226 *gata2b heterozygosity results in over-accessible gata2b chromatin leading to self-renewal*
227 *defects in HSPCs.*
228 To further investigate the mechanism underlying the effects of haploinsufficient *gata2b*
229 expression, we assessed the accessible peak distribution of *gata2b* and its upstream region.
230 First of all, the chromatin accessibility signals of *gata2b* were strongest in the HSPCs cluster
231 (Figure 3A, top line), then gradually disappeared with cell maturation (Figure 3A, all other
232 lines), which is consistent with the GATA2 function in HSC maintenance[1-3]. However, when
233 we compared the signals between WT on the left and *gata2b*$^{+/-}$ on the right, we observed
234 that the chromatin of the *gata2b* gene body and upstream regions was more accessible in
235 *gata2b*$^{+/-}$ HSPCs. Chromatin co-accessibility analysis, which was used to predict
236 *cis*-regulatory interaction in the genome, suggested that there were two regions, +3.5-4.1kb
237 termed enhancer 1(E1) and +5.4-6.2kb, termed enhancer 2 (E2), which could be strongly
238 interacting with each other (Figure 3A, black arrows in Links). Moreover, E1 had stronger
239 interaction with *gata2b* transcription start site (TSS) in *gata2b*$^{+/-}$ KM (Figure 3A, red arrows
240 in Links). These two upstream regions might work as enhancers for *gata2b* transcription.
241 When we assessed the cell distribution of the two enhancers and *gata2b* TSS (Figure 3B-G),
242 we observed that the accessibility signals of +5.4-6.2kb region-E2, (Figure 3B, C), was
243 accessible in HSPCs and *gata2b*$^{+/-}$ erythroid progenitors, while a +3.5-4.1kb region-E1,
244 (Figure 3D, E) was mainly accessible in HSPCs and *gata2b*$^{+/-}$ myeloid progenitors. This result
245 suggests that E1 and E2 might be responsible for different functions in hematopoiesis, and
246 particularly in *gata2b*$^{+/-}$ KM, E1 affects myeloid differentiation and E2 affects erythroid
247 differentiation. The percentage of cells with *gata2b* mRNA expression did not show a
248 significant difference as a result from the up-regulated chromatin accessibility and stronger
249 enhancer-TSS co-accessibility (Supplementary figure 6A and B), but we did see an increase
250 in the level of *gata2b* expression in the HSPC cluster as a result of the increase in
251 enhancer-TSS co-accessibility (Figure 3A, H, and I and Supplementary Figure 5B, right panel).
252
253 *The functional consequence of epigenetic changes of the gata2b locus in HSPCs is a*
254 *reduction in G2-M phase in cell cycle*
255 Because the most significant signal differences occurred in HSPC clusters and the cell

256 differentiation velocity differences appeared at "progenitors" clusters (Figure 2D), we
257 compared the different TF activity in progenitor-like clusters, including "HSPCs",
258 "progenitors", and "HSPCs and thrombocytes". *Gata2b*$^{+/-}$ progenitor-like clusters showed
259 significantly up-regulated chromatin accessibility of motifs of the BATF family of
260 transcription factors (Figure 4A). The BATF family of TFs are known to function in limiting the
261 self-renewal of HSCs in response to the accumulation of DNA damage[35], suggesting a defect
262 in self-renewal of *gata2b*$^{+/-}$ HSCs. To investigate this, proliferation was assessed in
263 CD41:GFP^{int} purified HSCs (Figure 4B). Although numerically unaffected by *gata2b*
264 heterozygosity (Figure 4C), cell cycle phase distribution was markedly different in *gata2b*$^{+/-}$
265 HSCs compared to WT (Figure 4D and E), with increased G1 phase and decreased G2-M
266 phase. As stem cells are known to have a short G1 phase to reduce cellular differentiation
267 potential, this may already be an indication that these HSCs lose self-renewal capacity and
268 differentiate, which is halted in the next differentiation stages[36].
269
270 *Epigenetic dysregulation results in delayed erythroid maturation in gata2b*$^{+/-}$ *kidney marrow*
271 Next, we sought to reveal the molecular mechanism behind the erythroid dysplasia in
272 *Gata2b*$^{+/-}$ KM. Compared to WT, *Gata2b*$^{+/-}$ progenitor-like clusters had significantly reduced
273 accessibility of GATA2 and GATA1::TAL1 motifs (Figure 4A). Despite the increase in *gata2b*
274 expression, this indicated an overall reduction of Gata2b function in activation of
275 downstream targets in the *gata2b*$^{+/-}$ zebrafish genome. Activation of GATA2 motifs were
276 mainly enriched in the erythroid progenitors (Figure 5A). An important downstream target
277 of Gata2 is Gata1, which is the master regulator of erythropoiesis[37,38]. We therefore
278 investigated *Gata1a* accessibility and expression in our dataset (Figure 5B-E). The *gata1a* TSS
279 was generally more accessible in *gata2b*$^{+/-}$ KM (Figure 5B), but not in erythroid progenitors 2
280 (Figure 5C). Furthermore, we found downregulation of *gata1a* mRNA expression (Figure 5D)
281 and the GATA1::TAL1 motif was less accessible (Figure 5E), further supporting the reduced
282 functionality of Gata1a in erythroid progenitors. This could be due to reduced expression of
283 *zfpm1* (also known as *Friend of GATA* (*FOG1)*), which can facilitate *GATA1* expression[39](Figure
284 5F). The increased expression of *gata2b* in HSPCs (Figure 3I) and reduced expression of
285 *gata1a* in erythroid progenitors (Figure 5D) suggested that the "GATA factor switch", which is
286 a shift from GATA2 to GATA1 occupation during erythroid lineage differentiation[40-42], might
287 be silenced, causing deficiencies in erythroid lineage differentiation.
288 Considering that the dysplastic cells in KM smears were of myeloid or erythroid origin,

289 we hypothesized that dysplastic cells occupy a proportion of erythroid or myeloid clusters.
290 For a more precise examination of small populations along the myeloid and erythroid
291 differentiation lineage, we sub-clustered ''HSPCs'', "proliferative progenitors'', "progenitors",
292 "myeloid progenitors 1", "myeloid progenitors 2", "erythroid progenitors 1", and "erythroid
293 progenitors 2" clusters and identified 5 significantly overrepresented sub-clusters in
294 *gata2b*$^{+/-}$ KM (Figure 5G-I, supplemental Figure 7 A-H). These sub-clusters represented 5.6%
295 of the total sequenced *gata2b*$^{+/-}$ cells, comparable to the proportion of dysplastic cells
296 observed in *gata2b*$^{+/-}$ KM. Differentially expressed genes (DEGs) analysis of overrepresented
297 sub-clusters in *gata2b*$^{+/-}$ KM compared to the rest of that cluster showed downregulation of
298 tubulin transcripts (not shown), suggesting a loss of cytoskeletal structure, a characteristic of
299 dysplasia. Besides, TF motif score analysis with snATAC-seq indicated a general reduction for
300 lineage-specific determination and differentiation motifs in *gata2b*$^{+/-}$ KM (supplemental
301 Figure 5A-F), suggesting that *gata2b*$^{+/-}$ cell maturation could be impaired resulting in a
302 differentiation block, as this is often the case in dysplasia[43].

303 To identify the mechanisms behind the erythroid differentiation delay in *gata2b*$^{+/-}$ KM,
304 we compared the DEGs and their function in the "erythroid progenitors 1" cluster. The result
305 showed that *gata2b*$^{+/-}$ "erythroid progenitors 1" cluster cells had upregulated signatures of
306 DNA replication together with the downregulation of transcripts necessary for mitosis
307 (Figure 5J), which can explain the origin of multi-lobulated nuclei and other nuclear
308 abnormalities observed in *gata2b*$^{+/-}$ dysplastic cells[44,45]. Pseudo-time trajectory inference
309 showed a diminished differentiation in *gata2b*$^{+/-}$ "erythroid progenitors 1" cluster with a
310 lower pseudo-time score (Figure 5K, L, and T). Pseudo-time dynamic gene expression level
311 also indicated the erythropoiesis differentiation delay in the *gata2b*$^{+/-}$ "erythroid progenitors
312 1" cluster with persistent high open chromatin, but lower expression of *gata1a*, which
313 should gradually decrease toward terminal erythroid maturation, and the decreased initial
314 level of *mki67*, indicative of incomplete differentiation (comparing Figure 5M-O to Figure
315 5P-R). This could be due to a diminished open chromatin structure at the locus leading to a
316 reduction in Gata2b binding sites to drive proper *Gata1a* expression, to support the
317 erythroid differentiation process. Furthermore, TF motif score analysis of the Erythroid
318 progenitors 1 cluster revealed a significant depletion for GATA family motifs, but an
319 enrichment for ETS and ETV family of transcription factor motifs in *gata2b*$^{+/-}$ "Erythroid
320 progenitors 1" cells (Figure 5S), suggesting an epigenetic dysregulation related to decreased
321 maturation of erythroblasts resulting from Gata2b haploinsufficiency.

322
323
324 ***Discussion***

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

331 Here, we used transgenic reporters, morphological phenotyping, and single-cell omics
332 sequencing to analyze *gata2b* heterozygous zebrafish. We showed that, while major
333 differentiation lineages remain intact, *gata2b* heterozygosity causes dysplasia in erythroid
334 and myeloid progenitors of zebrafish KM. snATAC-seq revealed an auto-regulatory feedback
335 mechanism for Gata2b haploinsufficiency, characterized by *gata2b* chromatin
336 over-accessibility and stronger co-accessibility of *gata2b* enhancers and TSS, indicating
337 epigenetic dysregulation in hematopoiesis in the zebrafish model. scRNA-seq analysis did
338 not identify a single population of dysplastic cells. Interestingly, a single erythroid progenitor
339 cluster was identified using snATAC-seq, and could indicate a specific dysplastic erythroid
340 population. The difference in recognition of the dysplastic cluster between scRNA-seq and
341 snATAC-seq may point to the underlying mechanism of the dysplasia where we have
342 evidence that the epigenetic compensatory mechanism of maintaining normal levels of
343 Gata2b may cause the dysplastic features during lineage differentiation, particularly in the
344 GATA2 to GATA1 switch and therefore affects erythroid lineage differentiation most. This also
345 underscores the importance of epigenetic analysis in other types of MDS and the fact that
346 many epigenetic regulators are mutated in MDS and AML[46]. In *gata2b*$^{+/-}$ erythroid
347 progenitors we found increased proliferative signatures together with decreased expression
348 of genes related to mitosis indicating an impaired cell cycle progression. Furthermore, we
349 found the depletion of GATA family factors indicating a failure of the "GATA factor switch"
350 that is indispensable for erythroid lineage differentiation[40-42]. We propose that these
351 alterations could play a role in the onset of dysplasia found in *gata2b*$^{+/-}$ zebrafish.

352 Interestingly, whereas *gata2b*$^{-/-}$ zebrafish display abrogated myeloid lineage
353 differentiation and a bias toward lymphoid differentiation [16,17], *gata2b*$^{+/-}$ did not simply
354 result in an intermediate phenotype between WT and *gata2b*$^{-/-}$. Instead, Gata2b

355 haploinsufficiency uniquely caused dysplasia, not observed in *gata2b*$^{-/-}$, potentially caused
356 by compensatory mechanisms in haploinsufficiency that are otherwise overcome in
357 *gata2b*$^{-/-}$ zebrafish[47]. This may also be the reason that missense mutations have a different
358 molecular mechanism, because these alleles are not prone to mRNA degradation and may
359 thus not be compensated by epigenetic deregulation. The differences in the homozygous
360 and heterozygous Gata2b knockout phenotype support a role for gene dosage underlying
361 the GATA2 deficiency phenotype, possibly explaining the phenotypic heterogeneity between
362 patients. Since both erythroid and myeloid dysplasia can be observed in GATA2 patients[4], we
363 propose that the presence of dysplastic cells in *gata2b*$^{+/-}$ resembles the clinical phenotypes
364 associated with GATA2 heterozygosity. In the future, the isolation of single dysplastic cells
365 could help us to further explore the effect of Gata2 haploinsufficiency in malignant
366 transformation. Nevertheless, it remains to be established how *gata2b*$^{+/-}$ HSPCs would
367 respond to secondary insults such as infections or severe bleeding.
368 In conclusion, while the major lineage differentiation remains intact, *gata2b*$^{+/-}$ zebrafish
369 possess a stressed proliferative HSPC compartment which leads to the generation of
370 erythroid and myeloid dysplastic cells. Taken together, our model provides insights into the
371 consequences of Gata2b dosage, illustrated the alteration of the microenvironment, and
372 reveal how changes in epigenetics affect the outcome in lineage differentiation after Gata2b
373 haploinsufficiency in zebrafish.
374
375 **Acknowledgements**
376 We thank Dr Monteiro (University of Birmingham) for careful reading of the manuscript. We
377 thank the Experimental Animal Facility of Erasmus MC for animal husbandry and the
378 Erasmus Optical Imaging Center for confocal microscopy services. This research is supported
379 by the European Hematology Association (junior and senior non clinical research fellowship)
380 (EdP), the Dutch Cancer Foundation KWF/Alpe d'HuZes (SK10321)(EdP), the Daniel den Hoed
381 Foundation for support of the Cancer Genome Editing Center (IT) and the Josephine Nefkens
382 Foundation for purchase of the Chromium 10x (IT).
383
384 **Authorship contributions**
385 EdP and EG conceived the study; EG, CK, HdL, JZ, DB, and EB performed experiments; EG, WZ,
386 CK, RH, KG and EdP analysed results; IPT provided resources and EG, WZ, CK and EdP wrote
387 the manuscript and IPT revised the manuscript.

388

389 **Disclosures**

390 The authors declare no conflicts of interests

391

392

393 **References**

394

395 1. Gao X, Johnson KD, Chang YI, et al. Gata2 cis-element is required for
396 hematopoietic stem cell generation in the mammalian embryo. *J Exp Med*. Dec 16
397 2013;210(13):2833-42. doi:jem.20130733 [pii]
398 20130733 [pii]
399 10.1084/jem.20130733
400 2. de Pater E, Kaimakis P, Vink CS, et al. Gata2 is required for HSC generation and
401 survival. *J Exp Med*. Dec 16 2013;210(13):2843-50. doi:jem.20130751 [pii]
402 20130751 [pii]
403 10.1084/jem.20130751
404 3. Tsai FY, Keller G, Kuo FC, et al. An early haematopoietic defect in mice lacking
405 the transcription factor GATA-2. *Nature*. Sep 15 1994;371(6494):221-6.
406 doi:10.1038/371221a0
407 4. Donadieu J, Lamant M, Fieschi C, et al. Natural history of GATA2 deficiency in a
408 survey of 79 French and Belgian patients. *Haematologica*. Aug
409 2018;103(8):1278-1287. doi:haematol.2017.181909 [pii]
410 1031278 [pii]
411 10.3324/haematol.2017.181909
412 5. Spinner MA, Sanchez LA, Hsu AP, et al. GATA2 deficiency: a protean disorder of
413 hematopoiesis, lymphatics, and immunity. *Blood*. Feb 6 2014;123(6):809-21.
414 doi:S0006-4971(20)36040-7 [pii]
415 2013/515528 [pii]
416 10.1182/blood-2013-07-515528
417 6. Ostergaard P, Simpson MA, Connell FC, et al. Mutations in GATA2 cause
418 primary lymphedema associated with a predisposition to acute myeloid leukemia
419 (Emberger syndrome). *Nat Genet*. Sep 4 2011;43(10):929-31. doi:ng.923 [pii]
420 10.1038/ng.923

421 7. Hsu AP, Sampaio EP, Khan J, et al. Mutations in GATA2 are associated with the
422 autosomal dominant and sporadic monocytopenia and mycobacterial infection
423 (MonoMAC) syndrome. *Blood*. Sep 8 2011;118(10):2653-5.
424 doi:S0006-4971(20)40686-X [pii]
425 2011/356352 [pii]
426 10.1182/blood-2011-05-356352
427 8. Dickinson RE, Griffin H, Bigley V, et al. Exome sequencing identifies GATA-2
428 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency.
429 *Blood*. Sep 8 2011;118(10):2656-8. doi:S0006-4971(20)40687-1 [pii]
430 10.1182/blood-2011-06-360313
431 9. Hahn CN, Chong CE, Carmichael CL, et al. Heritable GATA2 mutations
432 associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat*
433 *Genet*. Sep 4 2011;43(10):1012-7. doi:ng.913 [pii]
434 10.1038/ng.913
435 10. Mutsaers PG, van de Loosdrecht AA, Tawana K, Bodor C, Fitzgibbon J, Menko
436 FH. Highly variable clinical manifestations in a large family with a novel GATA2
437 mutation. *Leukemia*. Nov 2013;27(11):2247-8. doi:leu2013105 [pii]
438 10.1038/leu.2013.105
439 11. Wang X, Muramatsu H, Okuno Y, et al. GATA2 and secondary mutations in
440 familial myelodysplastic syndromes and pediatric myeloid malignancies.
441 *Haematologica*. Oct 2015;100(10):e398-401. doi:haematol.2015.127092 [pii]
442 100e398 [pii]
443 10.3324/haematol.2015.127092
444 12. Rodrigues NP, Janzen V, Forkert R, et al. Haploinsufficiency of GATA-2 perturbs
445 adult hematopoietic stem-cell homeostasis. *Blood*. Jul 15 2005;106(2):477-84.
446 doi:S0006-4971(20)53269-2 [pii]
447 10.1182/blood-2004-08-2989
448 13. Ling KW, Ottersbach K, van Hamburg JP, et al. GATA-2 plays two functionally
449 distinct roles during the ontogeny of hematopoietic stem cells. *J Exp Med*. Oct 4
450 2004;200(7):871-82. doi:jem.20031556 [pii]
451 20031556 [pii]
452 10.1084/jem.20031556
453 14. Butko E, Distel M, Pouget C, et al. Gata2b is a restricted early regulator of

454 hemogenic endothelium in the zebrafish embryo. *Development*. Mar 15
455 2015;142(6):1050-61. doi:142/6/1050 [pii]
456 DEV119180 [pii]
457 10.1242/dev.119180
458 15. Dobrzycki T, Mahony CB, Krecsmarik M, et al. Deletion of a conserved Gata2
459 enhancer impairs haemogenic endothelium programming and adult Zebrafish
460 haematopoiesis. *Commun Biol*. Feb 13 2020;3(1):71.
461 doi:10.1038/s42003-020-0798-3 [pii]
462 798 [pii]
463 10.1038/s42003-020-0798-3
464 16. Gioacchino E, Koyunlar C, Zink J, et al. Essential role for Gata2 in modulating
465 lineage output from hematopoietic stem cells in zebrafish. *Blood Adv*. Jul 13
466 2021;5(13):2687-2700. doi:S2473-9529(21)00348-7 [pii]
467 10.1182/bloodadvances.2020002993
468 17. Avagyan S, Weber MC, Ma S, et al. Single-cell ATAC-seq reveals
469 GATA2-dependent priming defect in myeloid and a maturation bottleneck in lymphoid
470 lineages. *Blood Adv*. Jul 13 2021;5(13):2673-2686. doi:S2473-9529(21)00347-5 [pii]
471 2020/ADV2020002992 [pii]
472 10.1182/bloodadvances.2020002992
473 18. Ma D, Zhang J, Lin HF, Italiano J, Handin RI. The identification and
474 characterization of zebrafish hematopoietic stem cells. *Blood*. Jul 14
475 2011;118(2):289-97. doi:S0006-4971(20)44778-0 [pii]
476 2010/327403 [pii]
477 10.1182/blood-2010-12-327403
478 19. Tamplin OJ, Durand EM, Carr LA, et al. Hematopoietic stem cell arrival triggers
479 dynamic remodeling of the perivascular niche. *Cell*. Jan 15 2015;160(1-2):241-52.
480 doi:S0092-8674(14)01638-9 [pii]
481 10.1016/j.cell.2014.12.032
482 20. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal
483 single-cell data. *Cell*. Jun 24 2021;184(13):3573-3587 e29.
484 doi:S0092-8674(21)00583-3 [pii]
485 10.1016/j.cell.2021.04.048
486 21. Stuart T, Srivastava A, Madad S, Lareau CA, Satija R. Single-cell chromatin state

487 analysis with Signac. *Nat Methods*. Nov 2021;18(11):1333-1341.
488 doi:10.1038/s41592-021-01282-5 [pii]
489 10.1038/s41592-021-01282-5
490 22. Cao J, Spielmann M, Qiu X, et al. The single-cell transcriptional landscape of
491 mammalian organogenesis. *Nature*. Feb 2019;566(7745):496-502.
492 doi:10.1038/s41586-019-0969-x [pii]
493 10.1038/s41586-019-0969-x
494 23. Boatman S, Barrett F, Satishchandran S, Jing L, Shestopalov I, Zon LI. Assaying
495 hematopoiesis using zebrafish. *Blood Cells Mol Dis*. Dec 2013;51(4):271-6.
496 doi:S1079-9796(13)00160-5 [pii]
497 10.1016/j.bcmd.2013.07.009
498 24. Traver D, Paw BH, Poss KD, Penberthy WT, Lin S, Zon LI. Transplantation and in
499 vivo imaging of multilineage engraftment in zebrafish bloodless mutants. *Nat Immunol*.
500 Dec 2003;4(12):1238-46. doi:ni1007 [pii]
501 10.1038/ni1007
502 25. Langenau DM, Ferrando AA, Traver D, et al. In vivo tracking of T cell
503 development, ablation, and engraftment in transgenic zebrafish. *Proc Natl Acad Sci U*
504 *S A*. May 11 2004;101(19):7369-74. doi:0402248101 [pii]
505 1017369 [pii]
506 10.1073/pnas.0402248101
507 26. Renshaw SA, Loynes CA, Trushell DM, Elworthy S, Ingham PW, Whyte MK. A
508 transgenic zebrafish model of neutrophilic inflammation. *Blood*. Dec 15
509 2006;108(13):3976-8. doi:S0006-4971(20)52189-7 [pii]
510 10.1182/blood-2006-05-024075
511 27. Ellett F, Pase L, Hayman JW, Andrianopoulos A, Lieschke GJ. mpeg1 promoter
512 transgenes direct macrophage-lineage expression in zebrafish. *Blood*. Jan 27
513 2011;117(4):e49-56. doi:S0006-4971(20)58625-4 [pii]
514 2010/314120 [pii]
515 10.1182/blood-2010-10-314120
516 28. Page DM, Wittamer V, Bertrand JY, et al. An evolutionarily conserved program of
517 B-cell development and activation in zebrafish. *Blood*. Aug 22 2013;122(8):e1-11.
518 doi:S0006-4971(20)54254-7 [pii]
519 2012/471029 [pii]

520 10.1182/blood-2012-12-471029
521 29. Carmona SJ, Teichmann SA, Ferreira L, et al. Single-cell transcriptome analysis
522 of fish immune cells provides insight into the evolution of vertebrate immune cell
523 types. *Genome Res*. Mar 2017;27(3):451-461. doi:gr.207704.116 [pii]
524 10.1101/gr.207704.116
525 30. Danilova N, Bussmann J, Jekosch K, Steiner LA. The immunoglobulin
526 heavy-chain locus in zebrafish: identification and expression of a previously unknown
527 isotype, immunoglobulin Z. *Nat Immunol*. Mar 2005;6(3):295-302. doi:ni1166 [pii]
528 10.1038/ni1166
529 31. Kortum AN, Rodriguez-Nunez I, Yang J, et al. Differential expression and ligand
530 binding indicate alternative functions for zebrafish polymeric immunoglobulin receptor
531 (pIgR) and a family of pIgR-like (PIGRL) proteins. *Immunogenetics*. Apr
532 2014;66(4):267-79. doi:10.1007/s00251-014-0759-4
533 32. Macaulay IC, Svensson V, Labalette C, et al. Single-Cell RNA-Sequencing
534 Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells. *Cell Rep*.
535 Feb 2 2016;14(4):966-977. doi:S2211-1247(15)01538-7 [pii]
536 10.1016/j.celrep.2015.12.082
537 33. Tang Q, Iyer S, Lobbardi R, et al. Dissecting hematopoietic and renal cell
538 heterogeneity in adult zebrafish at single-cell resolution using RNA sequencing. *J Exp*
539 *Med*. Oct 2 2017;214(10):2875-2887. doi:jem.20170976 [pii]
540 20170976 [pii]
541 10.1084/jem.20170976
542 34. Athanasiadis EI, Botthof JG, Andres H, Ferreira L, Lio P, Cvejic A. Single-cell
543 RNA-sequencing uncovers transcriptional states and fate decisions in
544 haematopoiesis. *Nat Commun*. Dec 11 2017;8(1):2045.
545 doi:10.1038/s41467-017-02305-6 [pii]
546 2305 [pii]
547 10.1038/s41467-017-02305-6
548 35. Wang J, Sun Q, Morita Y, et al. A differentiation checkpoint limits hematopoietic
549 stem cell self-renewal in response to DNA damage. *Cell*. Mar 2 2012;148(5):1001-14.
550 doi:S0092-8674(12)00145-6 [pii]
551 10.1016/j.cell.2012.01.040
552 36. Pauklin S, Vallier L. The cell-cycle state of stem cells determines cell fate

553 propensity. *Cell*. Sep 26 2013;155(1):135-47. doi:S0092-8674(13)01025-8 [pii]
554 10.1016/j.cell.2013.08.031
555 37. Welch JJ, Watts JA, Vakoc CR, et al. Global regulation of erythroid gene
556 expression by transcription factor GATA-1. *Blood*. Nov 15 2004;104(10):3136-47.
557 doi:S0006-4971(20)55869-2 [pii]
558 10.1182/blood-2004-04-1603
559 38. Wang F, Zhu Y, Guo L, et al. A regulatory circuit comprising GATA1/2 switch and
560 microRNA-27a/24 promotes erythropoiesis. *Nucleic Acids Res*. Jan
561 2014;42(1):442-57. doi:gkt848 [pii]
562 10.1093/nar/gkt848
563 39. Cantor AB, Orkin SH. Coregulation of GATA factors by the Friend of GATA (FOG)
564 family of multitype zinc finger proteins. *Semin Cell Dev Biol*. Feb 2005;16(1):117-28.
565 doi:S1084-9521(04)00102-8 [pii]
566 10.1016/j.semcdb.2004.10.006
567 40. Grass JA, Boyer ME, Pal S, Wu J, Weiss MJ, Bresnick EH. GATA-1-dependent
568 transcriptional repression of GATA-2 via disruption of positive autoregulation and
569 domain-wide chromatin remodeling. *Proc Natl Acad Sci U S A*. Jul 22
570 2003;100(15):8811-6. doi:1432147100 [pii]
571 1008811 [pii]
572 10.1073/pnas.1432147100
573 41. Pal S, Cantor AB, Johnson KD, et al. Coregulator-dependent facilitation of
574 chromatin occupancy by GATA-1. *Proc Natl Acad Sci U S A*. Jan 27
575 2004;101(4):980-5. doi:0307612100 [pii]
576 1010980 [pii]
577 10.1073/pnas.0307612100
578 42. Bresnick EH, Lee HY, Fujiwara T, Johnson KD, Keles S. GATA switches as
579 developmental drivers. *J Biol Chem*. Oct 8 2010;285(41):31087-93.
580 doi:S0021-9258(19)88840-3 [pii]
581 R110.159079 [pii]
582 10.1074/jbc.R110.159079
583 43. Monika Belickova M, Merkerova MD, Votavova H, et al. Up-regulation of
584 ribosomal genes is associated with a poor response to azacitidine in myelodysplasia
585 and related neoplasms. *Int J Hematol*. Nov 2016;104(5):566-573.

586 doi:10.1007/s12185-016-2058-3 [pii]
587 10.1007/s12185-016-2058-3
588 44. Nakayama Y, Yamaguchi N. Multi-lobulation of the nucleus in prolonged S phase
589 by nuclear expression of Chk tyrosine kinase. *Exp Cell Res*. Apr 1
590 2005;304(2):570-81. doi:S0014-4827(04)00703-7 [pii]
591 10.1016/j.yexcr.2004.11.027
592 45. Ullah Z, Lee CY, Depamphilis ML. Cip/Kip cyclin-dependent protein kinase
593 inhibitors and the road to polyploidy. *Cell Div*. Jun 2 2009;4:10. doi:1747-1028-4-10
594 [pii]
595 10.1186/1747-1028-4-10
596 46. Haferlach T, Nagata Y, Grossmann V, et al. Landscape of genetic lesions in 944
597 patients with myelodysplastic syndromes. *Leukemia*. Feb 2014;28(2):241-7.
598 doi:leu2013336 [pii]
599 10.1038/leu.2013.336
600 47. Rossi A, Kontarakis Z, Gerri C, et al. Genetic compensation induced by
601 deleterious mutations but not gene knockdowns. *Nature*. Aug 13
602 2015;524(7564):230-3. doi:nature14580 [pii]
603 10.1038/nature14580
604
605
606

607 **Figure legends**

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

609

610 (A) Representative pictures of kidney marrow smears after May–Grünwald-Giemsa staining
611 of WT KM smears (panel 1 and 2) and *gata2b*$^{+/-}$ KM smears (panel 3-8). 3) Blebbing in
612 cytoplasm of proerythroblast; 4) Irregular cytoplasm in erythroid precursor; 5) Lobed
613 nucleus and micronucleus in erythroblast; 6) Blebbing in cytoplasm of blast of sorted cell
614 after cytospin; 7) Binucleated promyelocyte; 8) Multinucleated promyelocyte.
615 (B) Frequency of dysplastic cells of the erythroid, myeloid and lymphoid lineage in KM
616 smears of WT (n=8) and *gata2b*$^{+/-}$ (n=8) zebrafish. ***: $P < 0.001$

617

618 ***Figure. 2: Single cell transcriptome and chromatin accessibility analysis reveals a***
619 ***differentiation block in gata2b$^{+/-}$ KM***
620 (A) The flow cytometry-based sorting strategy for kidney marrow cells for scRNA-seq and
621 snATAC-seq analysis.
622 (B) UMAP analysis of scRNA-seq (top panels) and snATAC-seq (bottom panels) showing all
623 cell types, which was 19 cell types in scRNA-seq and 16 cell types in snATAC-seq. UMAP atlas
624 was separated between WT and *gata2b*$^{+/-}$ cells.
625 (C) Cluster distribution quantification of scRNA-seq (left panel) and snATAC-seq (right panel)
626 between WT and *gata2b*$^{+/-}$ cells, with significantly differentially distributed clusters
627 represented in pink. A statistical significance threshold of FDR < 0.05 and log2 fold change >
628 1 were applied to determine the significance of the observed differences.

629 (D-E) Box plots representing the comparison of pseudo-time scores across progenitor,
630 myeloid, and erythroid clusters in D) scRNA-seq data and E) snATAC-seq data. *: $P < 0.05$,
631 **: $P < 0.01$, ***: $P < 0.0001$, NS: not significant.
632 (F) WT (left panel) and $^{+/-}$ (right panel) expression of *gata2b* in scRNA-seq data in
633 pseudo-time. Dots are depicted if *gata2b* expression is detected and colored according to
634 cell type. The curved line shows the dynamic expression of *gata2b* throughout pseudo-time
635 trajectory.
636 (G) WT (left panel) and *gata2b*$^{+/-}$ (right panel) chromatin accessibility of the *gata2b* locus in
637 snATAC-seq data in pseudo-time. Dots are depicted if open chromatin was detected and are
638 colored according to the cell types. The curved line shows the open chromatin at the *gata2b*
639 locus throughout pseudo-time trajectory.
640
641 ***Figure. 3: Differences in chromatin accessibility of Gata2b result in upregulation of Gata2b***
642 ***expression in gata2b$^{+/-}$ HSPCs***
643 (A) Coverage plots showing significant differences of *gata2b* chromatin accessibility
644 between WT and *gata2b*$^{+/-}$ cells with the location of the *gata2b* gene block in blue. The
645 normalized peak signal range is 0 to 40. Highlighted genomic regions within grey blocks
646 indicate the location of two enhancers, enhancer 1 (E1) and enhancer 2 (E2). The link
647 between the blocks illustrate predicted *cis*-regulatory interactions between the *gata2b* gene
648 body and its upstream region, based on examination of genome co-accessibility. Link colors
649 correspond to predicted co-accessibility scores, with yellow indicating stronger predicted

650 interactions. Black arrows indicated the co-accessibility between E2 and E1. Red arrows
651 indicated the co-accessibility between E1 and TSS which is stronger in *gata2b*$^{+/-}$ KM.
652 (B-G) UMAP feature plots depict the open chromatin of Enhancer 2 (E2; B and C), Enhancer
653 1 (E1; D and E), and transcription start site (TSS; F and G) of *gata2b*. UMAP atlas were
654 separated by WT (B, D and F) and *gata2b*$^{+/-}$ cells (C, E and G).
655 (H) Open chromatin of the *gata2b* locus per HSPCs cluster cell comparing WT and *gata2b*$^{+/-}$
656 HSPCs.
657 (I) Violin plot displaying the expression level of *gata2b* in HSPCs in WT (pink) and *gata2b*$^{+/-}$
658 (turquois).
659
660 ***Figure 4. Functional consequences in HSPCs upon Gata2b overexpression***
661 (A) The dot plot displays the log2 fold change values of differentially activated transcription
662 factor (TF) motifs in progenitor clusters, including "HSPCs", "progenitors", and "HSPCs and
663 thrombocytes", between WT and *gata2b*$^{+/-}$ cells. The TF motifs are ranked in increasing
664 order based on log2 fold change values. Dotted lines indicate the threshold for significance,
665 with absolute log2 fold change values > 0.58. The blue and red colors represent TF motifs
666 that are statistically significant (P value < 0.05) with absolute log2 fold change values > 0.58.
667 (B) Representative figure of identification of CD41:GFP^{int} population (in green) and
668 distribution in the FSC-A SSC-A kidney marrow population.
669 (C) Quantification of the percentage of CD41:GFP^{int} cells in WT (n=12) and *gata2b*$^{+/-}$ (n=12)
670 KM single live cells.

671 (D) Representative flow cytometry plots of cell cycle analysis by anti-Ki67 and DAPI staining

672 in WT and *gata2b*$^{+/-}$ CD41:GFP^{int} KM.

673 (E) Quantification of the percentages of WT (n=3) and *gata2b*$^{+/-}$ (n=3) CD41:GFP^{int} cells in

674 different cell cycle stages (* P value<0.05). Data represents as mean ± standard error of the

675 mean.

676

677 ***Figure. 5: Epigenetic changes resulting in the erythroid differentiation block in gata2$^{+/-}$ KM***

678 (A) UMAP feature plots depicting open chromatin of GATA2 motif sequence indicating

679 activity of GATA2. UMAP atlas were separated by WT (left panel) and *gata2b*$^{+/-}$ cells (right

680 panel). The top left corner of UMAP plots display the TF sequence logo corresponding to the

681 analyzed motif, generated by the function *MotifPlot*.

682 (B) Coverage plots showing differences of *gata1a* chromatin accessibility between WT and

683 *gata2b*$^{+/-}$ cells.

684 (C) Percentage of cells derived from WT and *gata2b*$^{+/-}$ KM with accessibility of the *gata1a*

685 locus across clusters in snATAC-seq data.

686 (D) Percentage of cells derived from WT and *gata2b*$^{+/-}$ KM expressing *gata1a* across clusters

687 in scRNA-seq data.

688 (E) UMAP plots depict the activity of combined motif GATA1::TAL1. UMAP atlas were

689 separated by WT (left panel) and *gata2b*$^{+/-}$ cells (right panel). The top left corner of the

690 UMAP plots display the TF sequence logo corresponding to the analyzed motif, generated by

691 the function *MotifPlot*.

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

693 (G-H) UMAP plots of the subclustering of the erythroid progenitors 1 cluster in scRNA-seq
694 data. UMAP atlas were separated by WT(G) and *gata2b*$^{+/-}$ cells (H)
695 (I) Quantification of distribution between WT and *gata2b*$^{+/-}$ erythroid progenitor 1
696 subclusters. Significantly differentially distributed clusters in pink. Dotted box indicates
697 overrepresented subcluster in *gata2b*$^{+/-}$ cells.
698 (J) STRING network of upregulated and downregulated transcripts in *gata2b*$^{+/-}$ "erythroid
699 progenitors 1". Only networks with more than 2 interactions were represented. Highlighted
700 in red DNA replication genes from KEGG pathways and highlighted in blue mitotic
701 cytokinesis genes from biological processes (Gene Ontology).
702 (K-L) Lineage trajectory analysis of WT (K) and *gata2b*$^{+/-}$ (L) erythroid progenitor 1
703 subclusters.
704 (M-R) WT (M-O) and *gata2b*$^{+/-}$ (P-R) pseudo-time expression of individual genes *hemoglobin*
705 *beta adult 1* (*hbba1*) (M and P), *GATA binding protein 1 alpha (gata1a)* (N and Q), and *mki67*
706 (O and R). Dotted red boxes indicate the different initial expression of *mki67* (O and R). Fold
707 change > 0.05 & adjusted P value < 0.05. FDR=False discovery rate. FD=Fold difference.
708 (S) Dot plot displaying log2 fold change values of differentially activated TF motifs in the
709 erythroid progenitor 1 cluster between WT and *gata2b*$^{+/-}$ cells in snATAC-seq data. The TF
710 motifs are ranked in descending order based on log2 fold change values. Dotted lines
711 indicate the threshold (absolute log2 fold change < -0.58 and >0.58) for significance, The
712 blue (downregulated) and red (upregulated) colors represent TFs that are statistically
713 significant (P value < 0.05).

714 (T) Box plots representing the comparison of pseudo-time scores between WT and
715 *gata2b*$^{+/-}$across all subclusters in erythroid progenitors 1. *: P < 0.05, **: P < 0.01, NS: no
716 significance.
717
718
719 ***Supplementary figure. 1: Lineage gating of total KM in WT and gata2b*$^{+/-}$ *zebrafish***
720 (A) Gating strategy of flowcytometry analysis of whole kidney marrow of WT and *gata2b*$^{+/-}$
721 zebrafish. Percentages represent the average of all zebrafish analyzed per genotype.
722 (B-E) Quantitation as percentages of the different cell populations in single viable cells of
723 myeloid (B), lymphoid and HSCs (C), erythrocytes (D) and progenitors (E) in WT and
724 *gata2b*$^{+/-}$ zebrafish kidney marrow over time. Statistical analysis showing as mean ±
725 standard error of the mean. ***: P < 0.001.
726
727 ***Supplementary figure. 2: Transgenic zebrafish reporter lines show no overall***
728 ***differentiation difference in gata2b*$^{+/-}$**
729 (A) *mpx* positive cells, depicted in green, with distribution in FSC-A SSC-A graph.
730 (B) Quantification of the percentage of mpx:GFP+ cells in WT (n = 8) and *gata2b*$^{+/-}$ (n = 5)
731 kidney marrow single live cells.
732 (C) *Lck* positive cells, depicted in green, with distribution in FSC-A SSC-A graph.
733 (D) Quantification of the percentage of lck:GFP+ cells in WT (n = 4) and *gata2b*$^{+/-}$ (n = 7)
734 kidney marrow single live cells.
735 (E) *Igm* positive cells, depicted in green, with distribution in FSC-A SSC-A graph.

736 (F) Quantification of the percentage of IgM:GFP+ cells in WT (n = 3) and *gata2b*$^{+/-}$ (n = 3)

737 kidney marrow single live cells.

738 (G-J) *mpeg* positive cells, depicted in green, mark monocytes (G) and phagocytic B-cells (I).

739 Quantification of the percentage of mpeg:GFP+ cells (H and J) in WT (n = 7) and *gata2b*$^{+/-}$ (n

740 = 7) kidney marrow single live cells.

741 (K) Frequency of differentiated cell types in KM cells in smears of WT (n=8) and *gata2b*$^{+/-}$

742 (n=8) zebrafish. Differentiated erythrocytes were excluded from quantification. *P

743 value<0.5. Data represents mean ± Standard error of the mean.

744

745 ***Supplementary Figure. 3: Transcriptomic heatmap and the expression of marker genes***

746 (A) Unbiased heatmap representing the expression level of the top 5 expressed transcripts

747 per cluster. Transcripts highly expressed in multiple clusters were not repeated in the list

748 (like hemoglobin in erythroid clusters). Transcripts identified only by their chromosome

749 location were not included.

750 (B-D) FeaturePlot gene expression analysis of *CD41* (B) in HSPCs and thrombocytes, *mki67*

751 (C) in Proliferative progenitors, and *pcna* (D) predominantly in undifferentiated clusters.

752

753 ***Supplementary Figure. 4: Expression and accessibility of marker genes used for cluster***

754 ***annotation and the pseudo-time trajectories inference***

755 (A) Feature plot showing the scRNA-seq expression of marker gene *hemoglobin beta adult 1*

756 (*hbba1*) in the erythroid lineage, *granulin1* (*grn1*) in the myeloid lineage, *cluster of*

757 *differentiation 37* (*CD37*) in the B cell lineage, and *GATA binding protein 2b* (*gata2b*) in
758 HSPCs to show cluster annotation.
759 (B) Feature plot showing the snATAC-seq gene activity of *hbba1* in the erythroid lineage,
760 *grn1* in the myeloid lineage, *CD37* in the B lineage and *gata2b* in HSPCs to show cluster
761 annotation.
762 (C) Coverage plot illustrating the chromatin accessibility peaks of *hbba1* in the erythroid
763 lineage, *grn1* in the myeloid lineage, *CD37* in the B lineage, and *gata2b* in hematopoietic
764 stem and progenitor cells (HSPCs) to show cluster annotation.
765 (D) UMAP plot of scRNA-seq data (upper panels) and snATAC-seq data (bottom panels)
766 colored by pseudo-time scores representing cell developmental trajectories split by WT (left
767 panels) and *gata2b*$^{+/-}$ (right panels) cells. Lower pseudo-time scores correspond to the root
768 of trajectories, while higher pseudo-time scores represent the terminal differentiation
769 trajectories. The lines plotted on the UMAP atlas illustrate the differentiation paths of cells.
770
771
772 ***Supplementary Figure. 5: Motif regulation of lineages differentiation along pseudo-time***
773 ***trajectories show reduced differentiation in gata2b$^{+/-}$ KM compared to WT.***
774 (A) Heatmap plots showing the dynamic motif activity (left panel) and relative gene
775 expression (right panel) of erythroid cells differentiation along pseudo-time trajectory in WT
776 KM. Left panel depicting clusters in snATAC-seq data, including "HSPCs", "Progenitors",
777 "HPSCs and thrombocytes", "Erythroid progenitors 1", and "Erythroid progenitors 2", while
778 right panel depicting clusters in scRNA-seq data, including "HSPCs", "Progenitors",

779 “Proliferative progenitors”, “HPSCs and thrombocytes”, “Erythroid progenitors 1”,
780 “Erythroid progenitors 2”, and “Erythroid progenitors 3”.
781 (B) Heatmap plots showing the dynamic motif activity (left panel) and relative gene
782 expression (right panel) during erythroid cell differentiation along pseudo-time trajectory in
783 *gata2b*$^{+/-}$ KM. The same clusters as those depicted in Figure S5A were used to generate the
784 heatmaps.
785 (C) Heatmap plots showing the dynamic motif activity (left panel) and relative gene
786 expression (right panel) during myeloid cells differentiation along pseudo-time trajectory in
787 WT KM. Left panel depicting clusters in snATAC-seq data, including “HSPCs”, “Progenitors”,
788 “HPSCs and thrombocytes”, “Myeloid progenitors 1”, “Myeloid progenitors 2”,
789 “Neutrophils”, and “Monocytes”, while right panel depicting clusters in scRNA-seq data,
790 including “HSPCs”, “Progenitors”, “Proliferative progenitors”, “HPSCs and thrombocytes”,
791 “Myeloid progenitors 1”, “Myeloid progenitors 2”, “Neutrophils”, and “Monocytes”.
792 (D) Heatmap plots showing the dynamic motif activity (left panel) and relative gene
793 expression (right panel) during myeloid cell differentiation along pseudo-time trajectory in
794 *gata2b*$^{+/-}$ KM. The same clusters as those depicted in Figure S5C were used to generate the
795 heatmaps.
796 (E) Heatmap plots showing the dynamic motif activity (left panel) and relative gene
797 expression (right panel) during lymphoid cell differentiation along pseudo-time trajectory in
798 WT KM. Left panel depicting clusters in snATAC-seq data, including “HSPCs”, “Progenitors”,
799 “HPSCs and thrombocytes”, “B cells and progenitors”, “B cells”, “T cells”, and “NK”, while

800 right panel depicting clusters in scRNA-seq data, including “HSPCs”, “Progenitors”,
801 “Proliferative progenitors”, “HPSCs and thrombocytes”, “B cells”, “T cells”, “NK”, and “NK2”.
802 (F) Heatmap plots showing the dynamic motif activity (left panel) and relative gene
803 expression (right panel) during lymphoid cells differentiation along pseudo-time trajectory
804 in *gata2b*$^{+/-}$ KM. The same clusters as those depicted in Figure S5E were used to generate
805 the heatmaps.
806
807 ***Supplementary Figure 6. Cell percentages with snATAC activity of the gata2b locus and***
808 ***gata2b expression per cluster.***
809 A) Cell percentages with detectable open chromatin of the *gata2b* locus per cluster. B) Cell
810 percentages with detectable *gata2b* expression per cluster.
811
812 ***Supplementary Figure 6. gata2b***$^{+/-}$ ***cells have a different distribution of subclusters***
813 ***compared to WT***
814 Split UMAP representing the cell distribution in the various subclusters (distinguishable by
815 different colors) in WT and *gata2b*$^{+/-}$ for HSPCs (A), Proliferative progenitors (B), Progenitors
816 (C), Myeloid progenitors 1 (D), Myeloid progenitors 2 (E), Erythroid progenitors 2 (F), HSPCs
817 and thrombocytes (G) , and Erythroid progenitors 3 (H), with respective quantification of
818 distribution between genotypes. Significantly differentially distributed subclusters in pink.
819 FDR<0.05 & Log2 fold change >1.
820

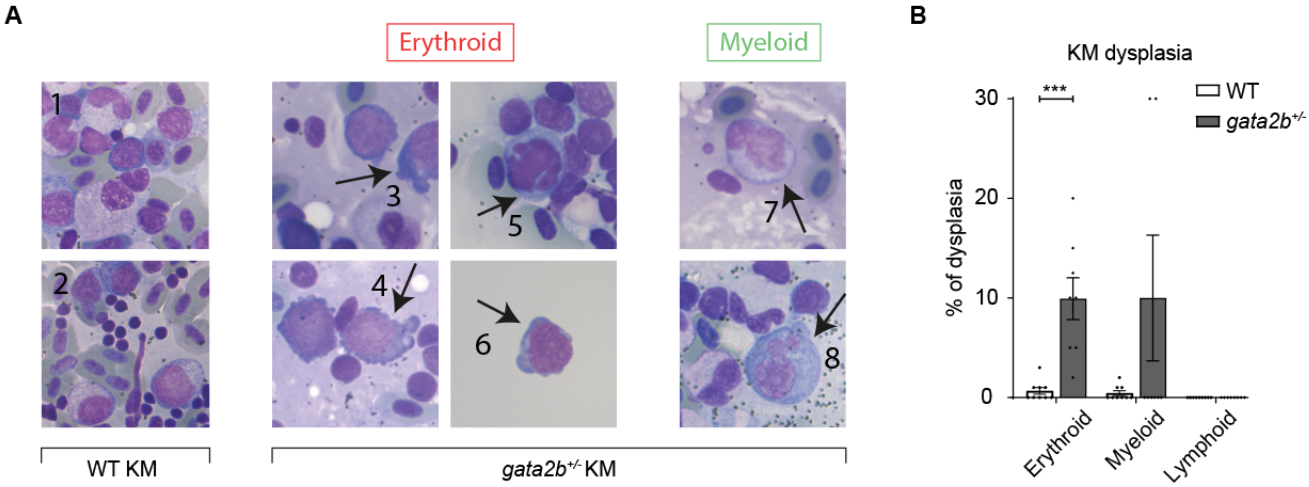
A
Erythroid
Myeloid
1
2
3
4
5
6
7
8
WT KM
gata2b+/- KM
B
KM dysplasia
WT
gata2b+/-
***
% of dysplasia
30
20
10
0
Erythroid
Myeloid
Lymphoid

Figure 1

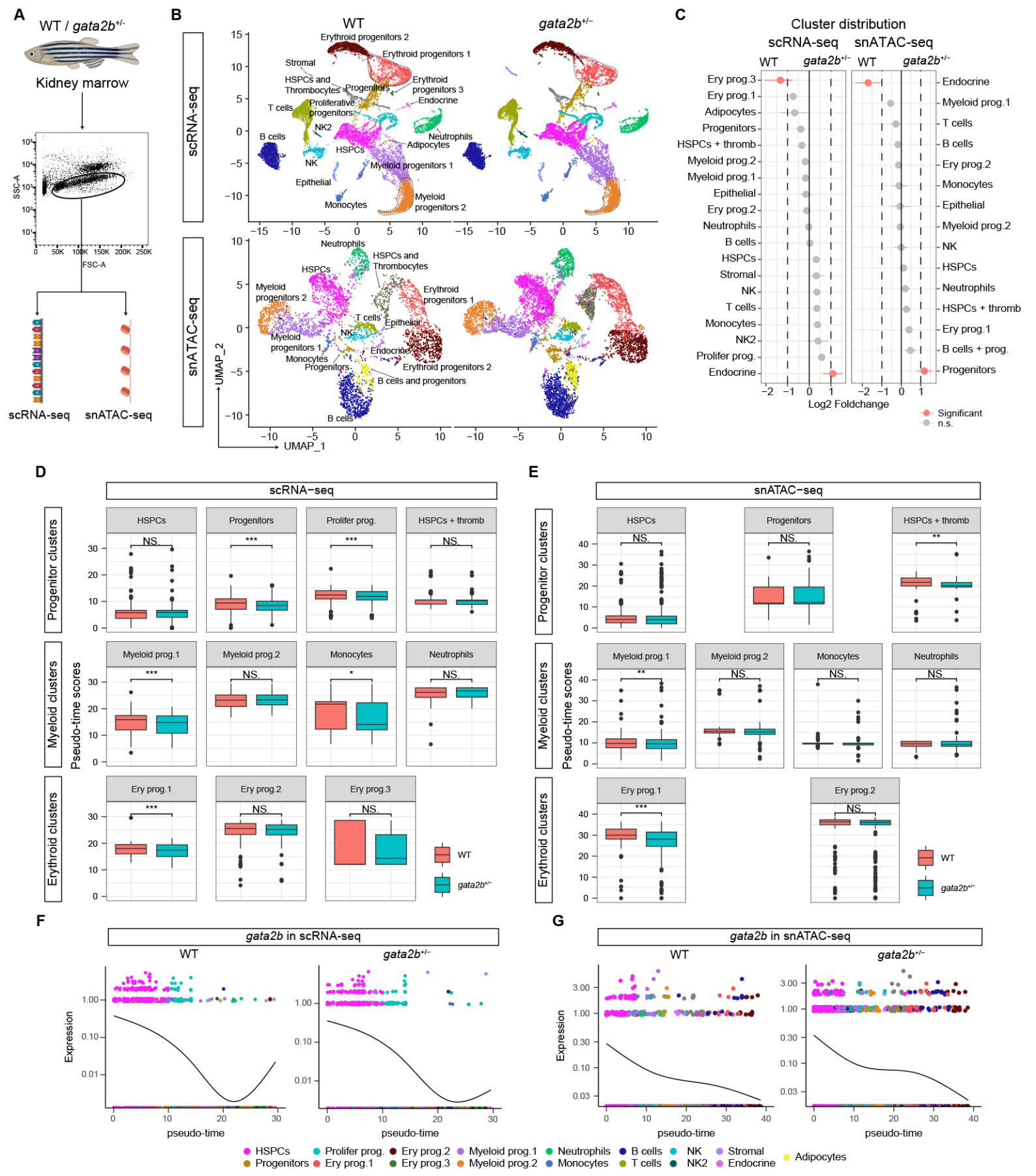
A
WT / gata2b+/-
Kidney marrow
SSC-A
FSC-A
scRNA-seq
snATAC-seq
B
WT
gata2b+/-
scRNA-seq
snATAC-seq
Erythroid progenitors 2
Erythroid progenitors 1
Stromal
HSPCs and Thrombocytes
Progenitors
Erythroid progenitors 3
Endocrine
T cells
Proliferative progenitors
NK2
B cells
Neutrophils
Adipocytes
HSPCs
NK
Myeloid progenitors 1
Epithelial
Monocytes
Myeloid progenitors 2
Neutrophils
HSPCs and Thrombocytes
HSPCs
Myeloid progenitors 2
Erythroid progenitors 1
T cells
Epithelial
NK
Myeloid progenitors 1
Monocytes
Endocrine
Progenitors
Erythroid progenitors 2
B cells and progenitors
B cells
UMAP_1
UMAP_2
C
Cluster distribution
scRNA-seq
snATAC-seq
WT
gata2b+/-
Ery prog.3
Ery prog.1
Adipocytes
Progenitors
HSPCs + thromb
Myeloid prog.2
Myeloid prog.1
Epithelial
Ery prog.2
Neutrophils
B cells
HSPCs
Stromal
NK
T cells
Monocytes
NK2
Prolifer prog.
Endocrine
Endocrine
Myeloid prog.1
T cells
B cells
Ery prog.2
Monocytes
Epithelial
Myeloid prog.2
NK
HSPCs
Neutrophils
HSPCs + thromb
Ery prog.1
B cells + prog.
Progenitors
Log2 Foldchange
Significant
n.s.
D
scRNA−seq
Progenitor clusters
HSPCs
Progenitors
Prolifer prog.
HSPCs + thromb
Myeloid clusters
Myeloid prog.1
Myeloid prog.2
Monocytes
Neutrophils
Erythroid clusters
Ery prog.1
Ery prog.2
Ery prog.3
Pseudo-time scores
NS.
***
*
WT
gata2b+/-
E
snATAC−seq
Progenitor clusters
HSPCs
Progenitors
HSPCs + thromb
Myeloid clusters
Myeloid prog.1
Myeloid prog.2
Monocytes
Neutrophils
Erythroid clusters
Ery prog.1
Ery prog.2
Pseudo-time scores
NS.
**
***
WT
gata2b+/-
F
gata2b in scRNA-seq
WT
gata2b+/-
Expression
pseudo-time
G
gata2b in snATAC-seq
WT
gata2b+/-
Expression
pseudo-time
HSPCs
Progenitors
Prolifer prog.
Ery prog.1
Ery prog.2
Ery prog.3
Myeloid prog.1
Myeloid prog.2
Neutrophils
Monocytes
B cells
T cells
NK
NK2
Stromal
Endocrine
Adipocytes

Figure 2

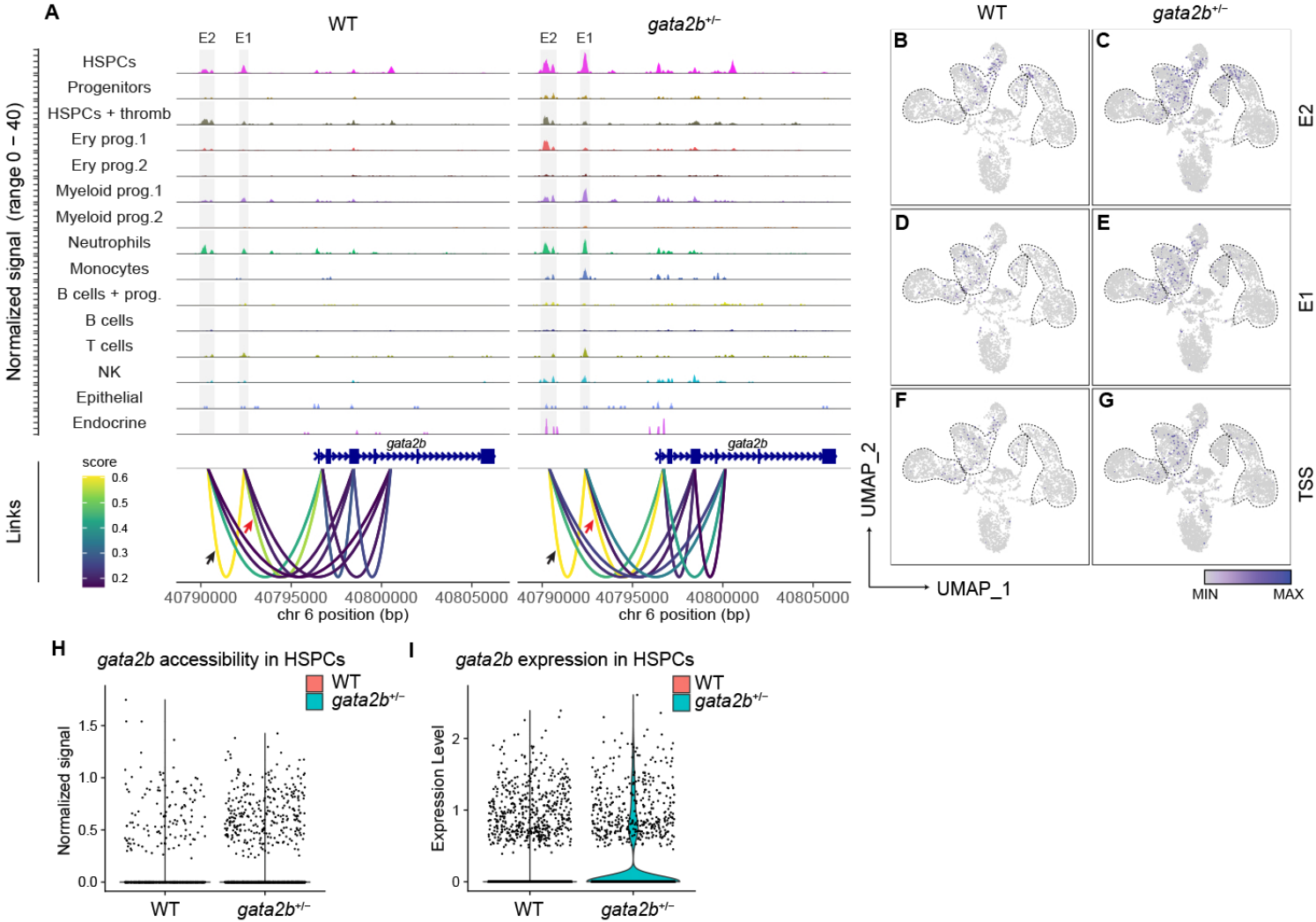
A
WT
gata2b+/–
E2
E1
HSPCs
Progenitors
HSPCs + thromb
Ery prog.1
Ery prog.2
Myeloid prog.1
Myeloid prog.2
Neutrophils
Monocytes
B cells + prog.
B cells
T cells
NK
Epithelial
Endocrine
Normalized signal (range 0 – 40)
gata2b
Links
score
0.6
0.5
0.4
0.3
0.2
40790000
40795000
40800000
40805000
chr 6 position (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
I
gata2b expression in HSPCs
Expression Level
2
1
0

Figure 3

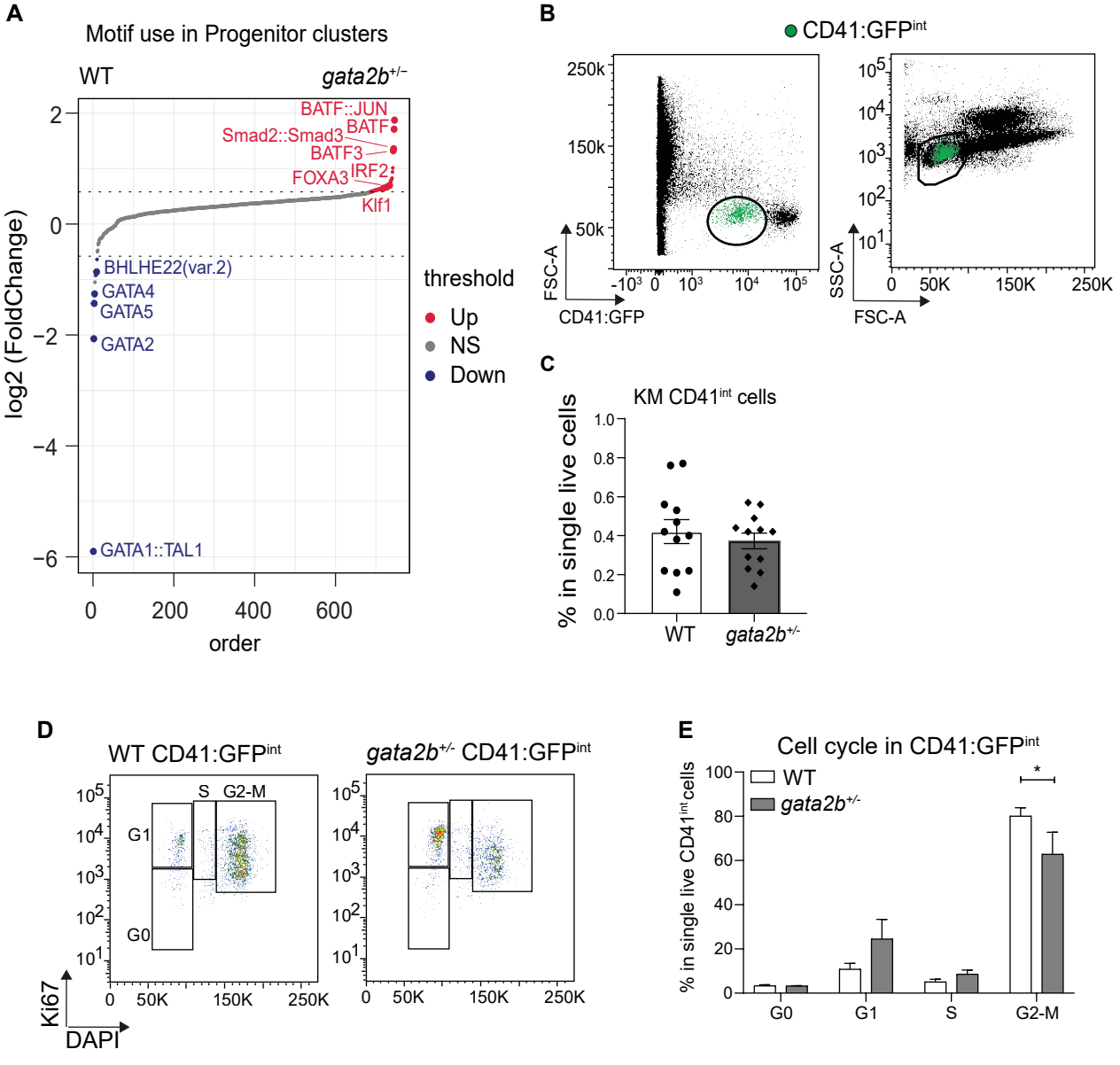
A
Motif use in Progenitor clusters
WT
gata2b+/-
BATF::JUN
BATF
Smad2::Smad3
BATF3
FOXA3
IRF2
Klf1
BHLHE22(var.2)
GATA4
GATA5
GATA2
GATA1::TAL1
log2 (FoldChange)
order
threshold
Up
NS
Down
B
CD41:GFPint
FSC-A
CD41:GFP
SSC-A
FSC-A
C
KM CD41int cells
% in single live cells
WT
gata2b+/-
D
WT CD41:GFPint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

Figure 4

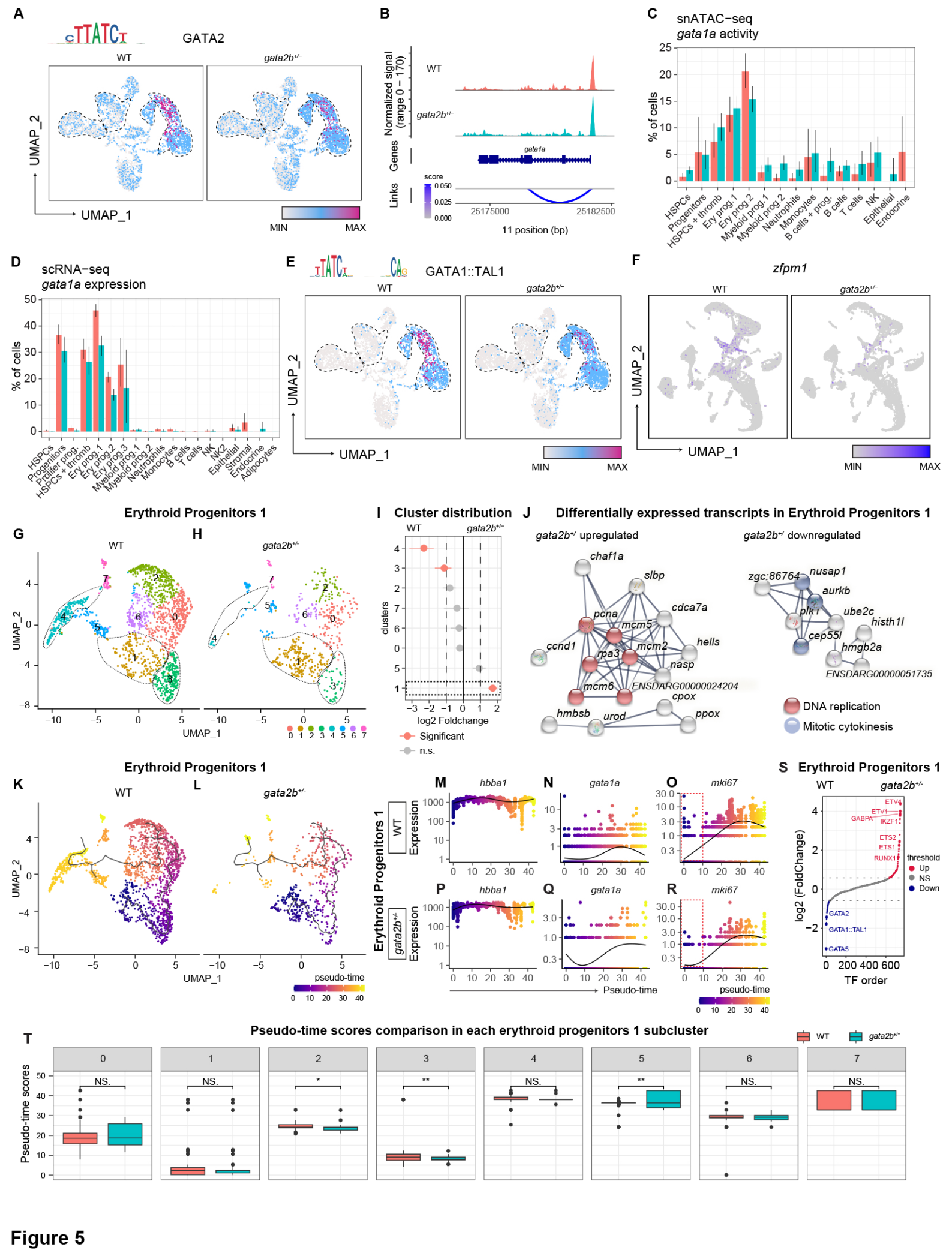
A
GATA2
WT
gata2b+/-
UMAP_2
UMAP_1
MIN
MAX
B
Normalized signal (range 0 – 170)
WT
gata2b+/-
Genes
gata1a
Links
score
0.050
0.025
0.000
25175000
25182500
11 position (bp)
C
snATAC−seq
gata1a activity
% of cells
HSPCs
Progenitors
HSPCs + thromb
Ery prog.1
Ery prog.2
Myeloid prog.1
Myeloid prog.2
Neutrophils
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 1
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 1
gata2b+/- upregulated
chaf1a
slbp
cdca7a
pcna
mcm5
ccnd1
rpa3
mcm2
hells
nasp
mcm6
ENSDARG00000024204
cpox
hmbsb
urod
ppox
gata2b+/- downregulated
zgc:86764
nusap1
aurkb
plk1
ube2c
histh1l
cep55l
hmgb2a
ENSDARG00000051735
DNA replication
Mitotic cytokinesis
Erythroid Progenitors 1
K
WT
L
gata2b+/-
UMAP_2
UMAP_1
pseudo-time
M
hbba1
N
gata1a
O
mki67
Erythroid Progenitors 1
WT
Expression
P
hbba1
Q
gata1a
R
mki67
gata2b+/-
Expression
Pseudo-time
pseudo-time
S
Erythroid Progenitors 1
WT
gata2b+/-
log2 (FoldChange)
ETV4
ETV1
GABPA
IKZF1
ETS2
ETS1
RUNX1
GATA2
GATA1::TAL1
GATA5
threshold
Up
NS
Down
TF order
T
Pseudo-time scores comparison in each erythroid progenitors 1 subcluster
WT
gata2b+/-
Pseudo-time scores
NS.
*
**

Figure 5