Gata2b haploinsufficiency causes aberrant transcriptional signatures in HSPCs resulting in myeloid and erythroid dysplasia in zebrafish

Emanuele Gioacchino[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

Corresponding author: e.depater@erasmusmc.nl

Abstract

The transcription factor *GATA2* has pivotal roles in hematopoiesis. Germline *GATA2* mutations result in GATA2 haploinsufficiency 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 are largely unknown. To explore the impact of GATA2 haploinsufficiency on hematopoiesis, we generated a zebrafish model carrying a heterozygous mutation in *gata2b*, an orthologue of *GATA2.* Morphological analysis revealed progression of myeloid and erythroid dysplasia in *gata2b*$^{+/-}$ kidney marrow (KM). Single cell RNA sequencing on KM cells showed that the erythroid dysplasia in *gata2b*$^{+/-}$ zebrafish was preceded by a differentiation block in erythroid progenitors, hallmarked by downregulation of cytoskeletal transcripts, aberrant proliferative signatures and ribosome biogenesis. Additionally, transcriptional and functional analysis of Gata2b haploinsufficient hematopoietic stem cells (HSCs) indicated that proliferative stress within the HSC compartment possibly contributes to the development of myeloid and erythroid dysplasia in *gata2b*$^{+/-}$ zebrafish.

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 loss of function of one allele, causing GATA2 haploinsufficiency. The clinical features of GATA2 haploinsufficiency are broad and include immunodeficiency, pulmonary-, vascular- and/or lymphatic dysfunctions and a strong propensity to develop MDS or AML[4,5]. GATA2 haploinsufficiency has been linked to the pathogenesis of Emberger syndrome[6], Monocytopenia and Mycobacterium Avium Complex (MonoMAC) syndrome[7], dendritic cell, monocyte, B- and natural killer 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 patients is approximately 80%[4,5]. There is no clear correlation 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 experimental models.

In mice, Gata2 has an essential regulatory activity in HSC generation and maintenance. *Gata2*-null mice are lethal at embryonic day (E) 10.5[3], whereas *Gata2* heterozygous (*Gata2*$^{+/-}$) mice survive to adulthood with normal blood values. However, *Gata2*$^{+/-}$ mice have a diminished HSC compartment in the bone marrow (BM) and *Gata2*$^{+/-}$ HSCs have a reduced repopulation capacity in competitive transplantation studies[12,13]. Whereas mouse models thus emerged as a precious 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 undiscovered. To better understand the biology of GATA2 haploinsufficiency syndromes, zebrafish serves as an attractive alternative model. Zebrafish have the advantage of having 2 *GATA2* orthologues; *gata2a* and *gata2b*. Gata2a is expressed predominantly in the vasculature[14] and is required for programming of the hemogenic endothelium[15,16]. Gata2b is expressed in hematopoietic stem/progenitor cells (HSPCs)[14] and homozygous deletion (*gata2b*$^{-/-}$) redirects HSPC differentiation to the lymphoid lineage in expense of myeloid differentiation causing a lymphoid bias with an incomplete B-cell differentiation in the KM, thus mimicking a portion of the GATA2 haploinsufficiency phenotypes found in patients. Additionally, the most primitive HSC compartment was lost in *gata2b*$^{-/-}$ KM, but none of the animals displayed signs of dysplasia[16,17]. Because patients carry heterozygous rather than homozygous *GATA2* mutations, we hypothesize that Gata2b haploinsufficiency could be a driver for leukemia onset. To investigate this, we assessed hematopoietic cell differentiation in *gata2b* heterozygous zebrafish (*gata2b*$^{+/-}$) KM. Upon morphological assessment of KM smears, we observed erythroid dysplasia in all *gata2b*$^{+/-}$ but not in WT zebrafish and myeloid dysplasia in 25% of the *gata2b*$^{+/-}$ zebrafish. Single cell transcriptome analysis revealed downregulation of cytoskeletal transcripts,

aberrant proliferative signatures and ribosome biogenesis in erythroid progenitor cells. Further characterization of the HSC transcriptome indicated increased proliferative stress in the hematopoietic system originating in the HSC compartment, possibly underlying the erythroid dysplasia in *gata2b*$^{+/-}$ zebrafish. These results highlight a concentration-dependent function of Gata2b in zebrafish hematopoiesis.

Material and methods

Generation and genotyping of Gata2b heterozygous zebrafish

Gata2b$^{+/-}$ and wild type (WT) clutch mates were used for all analyses[16] and animals were maintained under standard conditions. A knockout allele was generated by introducing a 28bp out-of-frame insertion in exon 3 as previously described[16].

Zebrafish embryos were kept at 28,5°C on a 14h/10h light-dark cycle in HEPES-buffered E3 medium. Zebrafish were anesthetized using tricaine and euthanized by ice-water. Animal studies were approved by the animal Welfare/Ethics Committee in accordance to Dutch legislation.

Kidney marrow isolation and analysis

Kidney marrow was dispersed mechanically using tweezers and dissociated by pipetting in phosphate buffered saline (PBS)/10% fetal calf serum (FCS) to obtain single-cell suspensions as previously described[16]. The KM cells were sorted in non-stick cooled micro tubes (Ambion) containing 10% FCS in PBS. Proliferation was assessed by anti-Ki67 staining in fixed (4% paraformaldehyde) KM cells. 7-AAD (7-amino-actinomycin D) (Stem-Kit Reagents) 0.5mg/L or DAPI 1mg/L was used for live/dead discrimination. FACS sorting and analysis were performed using FACS AriaIII (BD Biosciences).

May–Grünwald-Giemsa stain of KM smears

Kidney marrow smears were fixed in 100% MeOH before staining in May Grünwald solution (diluted 1:1 in phosphate buffer) and Giemsa solution (diluted 1:20 in phosphate buffer) followed by a last rinsing step in tap water.

Morphological analysis was performed by counting 200-500 hematopoietic cells of each kidney marrow smear; excluding mature erythrocytes and thrombocytes. Cells were categorized as: blast, myelocyte, neutrophil, eosinophil, lymphocyte or erythroblast. Furthermore, if dysplasia was observed within a specific lineage, the percentage of dysplastic cells within that lineage was determined by additional counting of at least 50 cells within that specific lineage.

Single cell RNA sequencing

Kidney marrow cells were isolated and $7x10^4$ viable cells were sorted from 2 pooled *Tg(CD41:GFP*[18]*; runx1:DsRed*[19]*)* wild type (WT) or *gata2b*$^{+/-}$ male zebrafish at 1 year of age. For additional replicates, $7x10^4$ single viable cells were sorted from kidney marrows of one WT or *gata2b*$^{+/-}$ female zebrafish between 18-20 months of age. cDNA was prepared using the manufacturers protocol (10x Chromium V3) and sequenced on a Novaseq 6000 instrument (Illumina). After sample processing and quality control analysis, 18147 cells for WT and 10849 cells for *gata2b*$^{+/-}$ were processed for further analysis. The read depth was over 20K reads for all replicates. Data was analyzed using the Seurat and Monocle3 R packages[20,21].

RNA extraction and RNA quality control

Total sample RNA isolation was performed according to the standard protocol of RNA isolation with Trizol and GenElute LPA (Sigma). Quality and quantity of the total RNA was determined on a 2100 Bioanalyzer (Agilent) using the Agilent RNA 6000 Pico Kit.

RNA Sequencing and gene set enrichment analysis (GSEA)

cDNA was prepared using the SMARTer Ultra Low RNA kit (Clontech) for Illumina Sequencing. The gene expression values were measured as FPKM (Fragments per kilobase of exon per million fragments mapped). Fragment counts were determined per gene with HTSeq-count, utilizing the strict intersection option, and subsequently used for differential expression analysis using the DESeq2 package, with standard parameters, in the R environment. GSEA (Gene Set Enrichment Analysis) was performed on the FPKM values using collection of curated gene sets from Molecular Signatures Database (MSigDB). Network analysis was performed using the output from GSEA analysis in Cytoscape Software.

Statistics

Statistical analysis was carried out in GraphPad Prism 8 (GraphPad Software). Unless otherwise specified, data were analyzed using unpaired, 2-tailed Student's t-test. Statistical significance was defined as $p<0.05$. Graphs are means ± standard error of mean (SEM) and the number of replicates is indicated in the figure legend.

Results

Gata2b$^{+/-}$ kidney marrow shows erythroid and myeloid dysplasia

We first assessed hematopoietic cell morphology in KM smears of WT and *gata2b$^{+/-}$* zebrafish with ages ranging from 9 months post fertilization (mpf) to 18 mpf. Morphological analysis showed that, while WT zebrafish had normal KM cell morphology, all *gata2b$^{+/-}$* KM samples had a considerable fraction of dysplastic cells in the erythroid lineage (Figure 1A, B, panel 2, 4, 6, 7 and 8). On average 0.5% of WT erythroid cells showed dysplastic features, compared to 9.9% of *gata2b$^{+/-}$* erythroid cells (Figure 1B, C), the latter representing 4.5% of the total kidney marrow population of *gata2b$^{+/-}$* zebrafish. Furthermore, we found evidence of myeloid lineage dysplasia in the *gata2b$^{+/-}$* KM in 25 % of the fish (Figure 1B, panel 1, 3 and 5 and C). In these samples, 30% of myeloid cells were dysplastic compared to 0.3% in WT. While the myeloid dysplasia was mostly represented by multi-lobulated nuclei, the erythroid abnormalities were ranging from nuclear deformities and double nuclei to irregular cytoplasm or an almost complete lack of cytoplasm (Figure 1B). The remaining cell types were not affected morphologically by Gata2b haploinsufficiency. These results indicate that *gata2b* heterozygosity induces dysplasia, predominantly in erythroid progenitors.

Gata2b$^{+/-}$ kidney marrow differentiation remains grossly intact over time

Next, we assessed hematopoietic lineage differentiation in *gata2b$^{+/-}$* KM by scatter analysis (Figure 1D). *Gata2b$^{+/-}$* zebrafish KM showed no significant difference in the distribution of either mature myeloid-, erythroid-, lymphoid- or progenitor- and HSPC populations compared to WT (Figure 1E-H). Since GATA2 haploinsufficiency manifestations might require longer periods of time to become evident[4,5], we tested the effect of *gata2b* heterozygosity after aging zebrafish up to 18 months. No significant differences between the proportions of the different lineages were detected over time in *gata2b$^{+/-}$* kidney marrow compared to WT (Figure 1E-H). However, the erythroid population increased, indiscriminately of genotype, after 12 months of age ($P<0.001$) (Figure 1G), while the remaining myeloid, lymphoid and HSPC and progenitor populations did not vary as dramatically (Figure 1E, F and H). Lineage analysis in transgenic lines specifically marking neutrophils by *Tg(mpx:GFP)* (Suppl Figure 1A-B), T-cells by *Tg(lck:GFP)* (Suppl. Figure 1C-D), B-cells by *Tg(Igm:GFP)* and *Tg(mpeg:GFP*) (Suppl. Figure 1E-F and I-J) or macrophages by *Tg(mpeg:GFP)* (Suppl. Figure 1G-H)[22,23] showed no significant alterations in lineage distribution[24-27]. However, KM smear quantification by MGG staining revealed that *gata2b$^{+/-}$* KM had a significant decrease in eosinophils compared to WT (Figure 1I), representing less than 5% of the total KM cells. In summary, based on

scatter analysis and transgenic reporters we conclude that the differentiation in major hematopoietic lineages is not altered in *gata2b*$^{+/-}$ KM.

Single cell transcriptome analysis reveals unique subpopulations in WT and gata2b$^{+/-}$ *KM indicative of a cell maturation arrest.*

To investigate molecular mechanisms underlying the dysplasia in *gata2b*$^{+/-}$ KM, we sought to identify the transcriptional differences in the dysplastic cell population. Through flow cytometry, we sorted 4 cell populations based on light scatter and observed dysplastic cells in the progenitor and lymphoid + HSPCs population of *gata2b*$^{+/-}$ KM, indicating that dysplastic cells could be viably sorted (Figure 1B, panel 7 and 8). However, flow cytometry did not identify a uniquely separated population of dysplastic cells, possibly caused by a heterogeneity in their shape. Therefore, we sorted progenitor and lymphoid + HSPC populations from WT and *gata2b*$^{+/-}$ KMs and performed single cell RNA sequencing (scRNA-seq) to identify transcriptome signatures defining dysplastic cells (Figure 2A). We identified 19 clusters based on nearest neighbor algorithm in the R Seurat package[20] (Figure 2B). Each cluster was classified based on differentially expressed genes (Suppl. Figure 2A-C) and known differentiation markers (Figure 2C-F)[28-33]. In short, cells of the erythroid lineage are known to express hemoglobin genes like *hbba1. itga2b (CD41)* was used as a marker for thrombocytes and *mpeg1.1* for monocytes and *lysozyme (lyz)* for neutrophils. Furthermore, the differentiated populations were devoid of expression of proliferation genes like *mki67* and *pcna* indeed suggesting that these are differentiated cells and not a progenitor compartment. The populations expressing markers like *CCAAT enhancer binding protein alpha* (*cebpa)* and *granulin1 (grn1)* were identified as myeloid progenitors 1 and 2. T cells were marked by the expression of *tox, il2rb* and *dusp2*, NK cells were marked by *nkl.3, nkl.4* and *ccl33.3* and B cells were marked by the expression of *CD37, CD79a*, *pax5* and high levels of the immunoglobulin *ighv1-4*. Corroborating a conserved lymphoid and myeloid differentiation program, the cluster proportion analysis indicated an overall similar distribution of cells within clusters between *gata2b*$^{+/-}$ and WT (Figure 2G, H).

Considering that dysplastic cells were readily recognized as myeloid or erythroid origin in KM smears, we hypothesized that dysplastic cells do not form one major cluster but instead, they occupy a proportion of erythroid or myeloid clusters. For a more precise examination of small populations, we subclustered erythroid and myeloid clusters (including their progenitors) and identified 5 significantly overrepresented sub-clusters in *gata2b*$^{+/-}$ KM (Figure 2I, J and K and Supplementary Figure 3A-H). These sub-clusters represented 5.6% of the total sequenced *gata2b*$^{+/-}$ cells, comparable to the proportion of dysplastic cells observed in *gata2b*$^{+/-}$ KM. Differential gene expression analysis of overrepresented sub-clusters in *gata2b*$^{+/-}$ KM compared to the rest of that

cluster showed downregulation of tubulin transcripts (not shown), suggesting a loss of cytoskeletal structure, a characteristic of dysplasia. Further investigation of clusters revealed that ''HSPCs'', ''proliferative progenitors'', "progenitors", "myeloid progenitors 1" and "erythroid progenitors 2" clusters contain underrepresented sub-clusters in *gata2b*$^{+/-}$ KM (Figure 2I, J and K and Supplementary Figure 3A-D). The loss of these cells suggests that *gata2b*$^{+/-}$ cell maturation could be impaired resulting in a differentiation block, as this is often the case in dysplasia[34]. In line with this hypothesis, *gata2b*$^{+/-}$ "erythroid progenitor 2" cluster cells had upregulated signatures of DNA replication together with downregulation of transcripts necessary for mitosis indicating an impaired cell cycle progression of these cells (Figure 2L). A block in the cell cycle progression can explain the origin of multi-lobulated nuclei and other nuclear abnormalities observed in *gata2b*$^{+/-}$ dysplastic cells[35,36]. To investigate if the loss of sub-clusters was caused by an erythroid differentiation block in *gata2b*$^{+/-}$ KM, we performed lineage trajectory analysis separately on WT and *gata2b*$^{+/-}$ "erythroid progenitor 2" clusters. Because immature erythroid markers were highest in sub-cluster 1, we chose this cluster as the starting point for the trajectory analysis. Compared to WT, differentiation was clearly diminished in *gata2b*$^{+/-}$ "erythroid progenitor 2" cells resulting in an incomplete trajectory and overall reduced pseudo-time value of 12.5 in comparison to 20 in WT "erythroid progenitor 2" cells. Next, we examined the pseudo-time expression of erythroid lineage specific genes in the lineage trajectory. As expected, all cells in both WT and *gata2b*$^{+/-}$ "erythroid progenitor 2" cluster showed high expression of hemoglobin gene *hbba1*. During erythroid differentiation, expression of *GATA1* is known to be highest in immature erythroid cells and gradually decreases towards mature erythroid cells[37]. WT "erythroid progenitor 2" cells indeed show high expression for the zebrafish *GATA1* orthologue *gata1a* at the start of the differentiation trajectory and decrease its expression during differentiation, *gata2b*$^{+/-}$ "erythroid progenitor 2" cells showed a complete opposite expression pattern for *gata1a*. Normal erythroid differentiation is dependent on the "GATA factor switch", which is characterized by a shift from *GATA2* to *GATA1* occupation during erythroid lineage differentiation[38-40]. Our results suggest that these mechanisms might be altered in *gata2b*$^{+/-}$ KM, causing deficiencies in erythroid lineage differentiation.

Because we identified transcriptome signatures indicative of increased proliferation in *gata2b*$^{+/-}$ "erythroid progenitor 2" cluster, we analyzed the pseudo-time expression of *bona fide* proliferation marker *mki67* in these cells. As we described before for *gata2b*$^{-/-}$ HSCs[16], we observed upregulation of *mki67* in *gata2b*$^{+/-}$ "erythroid progenitor 2" cluster cells in contrast to WT cells that gradually decreased *mki67* expression throughout lineage trajectory. Similarly, we performed lineage trajectory and pseudo-time analysis for ''HSPCs'', ''proliferative progenitors'', "progenitors" and "myeloid progenitors 1" clusters (Supplementary Figure 4A-P). In these clusters, we found

underrepresented sub-clusters indicative of an accumulation of the cell types in *gata2b*$^{+/-}$ compared to WT, but lineage differentiation was not severely affected from Gata2b haploinsufficiency based on pseudo-time values and the pseudo-time expression of cluster marker genes (Supplementary Figure 4A-P). Nevertheless, in "myeloid progenitors 1" cluster, the expression of proliferation marker *mki67* was upregulated in *gata2b*$^{+/-}$ cells in contrast to WT cells throughout lineage differentiation, indicating that altered proliferation is not solely specific to "erythroid progenitor 2" cluster but broadly present in other clusters of *gata2b*$^{+/-}$ cells (Supplementary Figure 4E and H). Overall, in single cell transcriptome analysis, dysplastic cells do not form a separate cluster but are likely scattered in the myeloid and erythroid progenitor clusters. Furthermore, our findings indicate that erythroid dysplasia in *gata2b*$^{+/-}$ KM possibly originates as a result of proliferative alterations and impaired differentiation.

Gata2b$^{+/-}$ *HSPCs have a high nucleic acid metabolism and downregulated cytoskeletal and protein ubiquitination transcripts*

In *gata2b*$^{+/-}$ KM, dysplasia was observed in both myeloid and erythroid cells, suggesting an aberrancy derived from a common progenitor. Additionally, Gata2b expression was present mostly in the HSPC cluster (Figure 2F) and complete deletion of *gata2b* was shown to predominantly affect HSPCs in zebrafish[16]. The single cell analysis of HSPCs showed great heterogeneity with scattered HSPCs expressing differentiation markers like *ighv1-4* indicative of B cell differentiation, *Nkl.2* indicative of NK-differentiation and *icn* indicative of myeloid differentiation (Suppl Figure 2A). To investigate the effect of Gata2b haploinsufficiency in the most immature cells of the HSPC compartment, we further purified the HSC population by flow sorting of the CD41:GFP^{low} expressing cells from WT or *gata2b*$^{+/-}$ KMs (Figure 3A), as these were shown to contain transplantable HSCs[18]. Although numerically unaffected by *gata2b* heterozygosity (Figure 3B), gene expression was markedly different in *gata2b*$^{+/-}$ HSCs compared to WT. To elucidate the effect of Gata2b haploinsufficiency in specific biological pathways, we performed gene-set enrichment (GSEA) and KEGG analysis that revealed 71 upregulated and 80 downregulated gene-sets ($P<0.05$) in *gata2b*$^{+/-}$ HSCs. We found an enrichment of gene-sets related to DNA replication and ribosome biogenesis in *gata2b*$^{+/-}$ HSCs (Figure 3C-E). Simultaneously, gene-sets related to cytoskeleton and protein degradation were underrepresented in *gata2b*$^{+/-}$ HSCs (Figure 3F-H). Similar to the *gata2b*$^{+/-}$ transcriptome signatures we observed in scRNA-seq, we found upregulation of proliferation related genes such as *pcna* and *mcm2-8* in *gata2b*$^{+/-}$ CD41:GFP^{low} cells, indicating these signatures originate from the most immature compartment of *gata2b*$^{+/-}$ KM (Supplementary Figure 4Q). To understand the consequences of

under- and over- represented gene-sets found in *gata2b*$^{+/-}$ HSCs, we performed network analysis using Cytoscape software. In this analysis, gene-sets (dots) are connected with lines if they share similar genes and form networks if they are assigned in similar biological processes. Network analysis further revealed the enrichment of gene-sets related to proliferation and transcription at the expense of gene-sets related to protein ubiquitination and cytoskeleton structure (Figure 3I). Because our results showed an aberrant proliferative signature, we assessed proliferation by flow cytometry in *gata2b*$^{+/-}$ CD41:GFP^{low} cells. We found a relative increase in cells in G1 phase of the cell cycle and a significant decrease in cells in G2-M phase of the cell cycle phase in *gata2b*$^{+/-}$ HSCs indicating that despite increased proliferative signatures cell cycle progression is impaired in *gata2b*$^{+/-}$ HSCs (Figure 3J and K). These data suggest that *gata2b*$^{+/-}$ HSPCs have an aberrant expression of the genes controlling cell structure and replication and fail to progress through cell cycle, which is possibly the source of dysplasia observed in the KM of *Gata2b*$^{+/-}$ zebrafish.

Discussion

In humans, a balanced *GATA2* expression is essential for proper hematopoiesis. Consequently, more than 80% of *GATA2* mutation carriers progress to hematological malignancy by 40 years of age[4]. While the clinical consequences of *GATA2* mutations became obvious over the last decades, the regulation of GATA2 activity and its contribution to human bone marrow failure syndromes and progression to hematological malignancy remain incompletely understood.
Here, we used transgenic reporters, morphological phenotyping and RNA sequencing to analyze *gata2b* heterozygous zebrafish. We showed that, while major differentiation lineages remain intact, *gata2b* heterozygosity causes dysplasia in erythroid and myeloid progenitors of zebrafish KM. ScRNA-seq analysis did not identify a single population of dysplastic cells indicating that these cells are likely scattered through erythroid and myeloid clusters. In fact, after sub-clustering we identified uneven distribution of erythroid and myeloid sub-clusters between WT and *gata2b*$^{+/}$ KM. Further examination with lineage trajectory analysis captured a differentiation block in one cluster of erythroid progenitors, concomitant with aberrant expression of *gata1a*. This suggested that alterations in GATA switch mechanisms may also contribute to the deterioration of erythroid dysplasia in *gata2b*$^{+/-}$ KM. Additionally, in the same erythroid progenitor cluster, an increased proliferative signature together with decreased expression of genes related to mitosis indicated an impaired cell cycle progression. We propose that these alterations could play a role in the onset of dysplasia found in *gata2b*$^{+/-}$ zebrafish.

Because we found both myeloid and erythroid dysplasia in *gata2b*$^{+/-}$ zebrafish and the expression of *gata2b* was highest in immature HSPCs compare to differentiated cells, this suggested that Gata2b haploinsufficiency affects the HSPC compartment. However, even after sub-clustering, we did not observe dramatic differences between the distribution of WT and *gata2b*$^{+/-}$ cells in the HSPC cluster. Because the HSPC cluster also contained lineage primed cells, we purified the most primitive HSPCs (CD41 intermediate cells or HSCs) from WT and *gata2b*$^{+/-}$ KMs and performed additional RNA sequencing experiments to identify the effect of Gata2b haploinsufficiency in HSCs. After analyzing HSCs, an increased proportion of cells in G1 phase of the cell cycle were found in *gata2b*$^{+/-}$ HSCs compared to WT and these cells transcriptionally showed loss of quiescence as observed in Gata2b knockout (*gata2b*$^{-/-}$) HSCs[16]. Furthermore, the proliferation assay indicated that cell cycle progression is diminished in these cells in G2-M phase. Because erythroid progenitors showed upregulation of proliferation genes and downregulation of mitosis genes, we propose that Gata2 haploinsufficiency causes proliferative stress in HSPCs. As cells become multi-lobulated towards the beginning of G2-M phase of the cell cycle, a block in cell cycle progression in G2-M phase could possibly explain the presence of multi-lobulated dysplastic cells in *gata2b*$^{+/-}$ KM.

Interestingly, whereas *gata2b*$^{-/-}$ zebrafish display abrogated myeloid lineage differentiation and a bias toward lymphoid differentiation[16,17], *gata2b*$^{+/-}$ did not simply result in an intermediate phenotype between WT and *gata2b*$^{-/-}$. Instead, Gata2b haploinsufficiency uniquely cause dysplasia, not observed in *gata2b*$^{-/-}$. The differences in the homozygous and heterozygous *gata2b* knockout phenotype support a role for gene dosage underlying GATA2 deficiency phenotype, possibly explaining the phenotypic heterogeneity between patients. Since both erythroid and myeloid dysplasia can be observed in GATA2 patients[4], we propose that the presence of dysplastic cells in *gata2b*$^{+/-}$ resembles the clinical phenotypes associated with *GATA2* heterozygosity. In the future, the isolation of single dysplastic cells could help us to further explore the effect of Gata2 haploinsufficiency in malignant transformation. Nevertheless, it remains to be established how *gata2b*$^{+/-}$ HSPCs would respond to secondary insults such as infections or severe bleeding.

In conclusion, while the major lineage differentiation remains intact, *gata2b*$^{+/-}$ zebrafish possess a stressed proliferative HSPC compartment which leads to the generation of erythroid and myeloid dysplastic cells. Taken together, our model provides insights into the consequences of Gata2b dosage, and reveals transcriptional networks affected by Gata2b haploinsufficiency in zebrafish.

Acknowledgements

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

Authorship contributions

EdP and EG conceived the study; EG, CK, HdL, JZ, DB, and EB performed experiments; EG, CK, RH, KG and EdP analysed results; IT provided resources and EG, CK and EdP wrote the manuscript and IT revised the manuscript.

Disclosures

The authors declare no conflicts of interests

Figure legends

Figure. 1: *Gata2b$^{+/-}$ kidney marrow shows erythroid dysplasia and normal lineage distribution*

Representative pictures of kidney marrow smears after May–Grünwald-Giemsa staining of A) WT KM smears and B) *gata2b$^{+/-}$* KM smears. 1) Binucleated erythroblast; 2) Blebbing in cytoplasm of proerythroblast; 3) Binucleated promyelocyte; 4) Lobed nucleus and micronucleus in erythroblast; 5) Multinucleated promyelocyte; 6) Irregular cytoplasm in erythroid precursor; 7) Lobed nucleus in erythrocytes of sorted cell after cytospin; 8) Blebbing in cytoplasm of blast of sorted cell after cytospin. C) Frequency of dysplastic cells of the erythroid, myeloid and lymphoid lineage in KM smears of WT (n=9) and *gata2b$^{+/-}$* (n=9) zebrafish. D) Gating strategy of FACS analysis of whole kidney marrow of WT and

gata2b$^{+/-}$ zebrafish. Percentages represent the average of all zebrafish analyzed per genotype. Quantitation as percentages of the different cell populations in single viable cells of E) myeloid F) lymphoid and HSCs G) erythrocytes H) progenitors in WT and *gata2b*$^{+/-}$ zebrafish kidney marrow over time. Mean ± SEM. I) Frequency of differentiated cell types in KM cells in smears of WT (n=9) and *gata2b*$^{+/-}$ (n=9) zebrafish. Differentiated erythrocytes were excluded from quantification. * P value<0.5, ***P value<0.001. Data represents mean ± Standard error of the mean.

Figure. 2: Single cell analysis reveals unique *gata2b*$^{+/-}$ cells scattered among clusters and a differentiation block in erythroid progenitors

A) Kidney marrow light scatter sorting strategy for scRNA sequencing analysis and B) UMAP analysis of scRNA-seq data showing different cell types based on expression analysis. FeaturePlot gene expression analysis of C) *hemoglobin beta adult 1* (*hbba1*) expressed in the erythrocyte lineage D*) granulin1* (*grn1)* in the myeloid lineage E) *cluster of differentiation 37* (*CD37)* expressed in the B cell lineage and F) *GATA binding protein 2b* (*gata2b)* in HSPCs. G) Split UMAP of all cells separating WT and *gata2b*$^{+/-}$ cells. H) Quantification of distribution between WT and *gata2b*$^{+/-}$ cells in each cluster. Significantly differentially distributed clusters in pink. FDR<0.05 & Log2 fold change >1. I) Split UMAP of the subclustering of the erythroid progenitors 2 cluster. K) Quantification of distribution between WT and *gata2b*$^{+/-}$ erythroid progenitor 2 subcluster cells. Significantly differentially distributed clusters in pink. Dotted box indicates overrepresented subcluster in *gata2b*$^{+/-}$ cells L) STRING network of upregulated and downregulated transcripts in *gata2b*$^{+/-}$ "erythroid progenitors 2". Only networks with more than 2 interactions were represented. Highlighted in red DNA replication genes from KEGG pathways and highlighted in blue mitotic cytokinesis genes form biological processes (gene Ontology). M) Lineage trajectory analysis of WT and N) *gata2b*$^{+/-}$ erythroid progenitor 2 subclusters. Note the difference in pseudotime scale between the WT and *gata2b*$^{+/-}$ trajectories. O-R) WT and S-U) *gata2b*$^{+/-}$ pseudotime expression of individual genes (O +S) *hemoglobin beta adult 1* (*hbba1*), (P+T) *GATA binding protein 1 (gata1),* (R+U) *mki67.* Fold change >0.05 & adjusted P value <0.05. FDR=False discovery rate. FD=Fold difference.

Figure. 3: *Gata2b*$^{+/-}$ HSPCs have a high nucleic acid metabolism and downregulated cytoskeletal and protein ubiquitination transcripts

A) Representative figure of identification of CD41:GFP^{int} population (in green) and distribution in the FSC-A SSC-A kidney marrow population. B) Quantification of the percentage of CD41:GFP^{int} cells in

WT and *gata2b*$^{+/-}$ kidney marrow single live cells. C-H) Representative gene set enrichments plots showing the profile of the enrichment score (ES) and positions of gene set members on the rank ordered list. I) Cytoscape 3.8.2 enrichment map depicting significantly ($p < 0.01$ and FDR<0.15) upregulated (red) and downregulated (blue) gene sets. The network clusters in yellow are generated using the AutoAnnotation application of Cytoscape 3.8.2. J) Representative flow cytometry plots of cycle analysis by anti-Ki67 and DAPI staining in WT and *gata2b*$^{+/-}$ CD41:GFP^{int} KM. K) Quantification of the percentages of WT (n=3) and *gata2b*$^{+/-}$ (n=3) CD41:GFP^{int} cells in different cell cycle stages (* P value<0.5). Data represents mean ± Standard error of the mean.

References

1 Gao, X. *et al.* Gata2 cis-element is required for hematopoietic stem cell generation in the mammalian embryo. *J Exp Med* **210**, 2833-2842, doi:jem.20130733 [pii]

10.1084/jem.20130733 (2013).

2 de Pater, E. *et al.* Gata2 is required for HSC generation and survival. *J Exp Med* **210**, 2843-2850, doi:jem.20130751 [pii]

10.1084/jem.20130751 (2013).

3 Tsai, F. Y. *et al.* An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature* **371**, 221-226, doi:10.1038/371221a0 (1994).

4 Donadieu, J. *et al.* Natural history of GATA2 deficiency in a survey of 79 French and Belgian patients. *Haematologica* **103**, 1278-1287, doi:haematol.2017.181909 [pii]

10.3324/haematol.2017.181909 (2018).

5 Spinner, M. A. *et al.* GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics, and immunity. *Blood* **123**, 809-821, doi:S0006-4971(20)36040-7 [pii]

10.1182/blood-2013-07-515528 (2014).

6 Ostergaard, P. *et al.* Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). *Nat Genet* **43**, 929-931, doi:ng.923 [pii]

10.1038/ng.923 (2011).

7 Hsu, A. P. *et al.* Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood* **118**, 2653-2655, doi:blood-2011-05-356352 [pii]

10.1182/blood-2011-05-356352 (2011).

8 Dickinson, R. E. *et al.* Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. *Blood* **118**, 2656-2658, doi:S0006-4971(20)40687-1 [pii]

10.1182/blood-2011-06-360313 (2011).

9 Hahn, C. N. *et al.* Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat Genet* **43**, 1012-1017, doi:ng.913 [pii]

10.1038/ng.913 (2011).

10 Mutsaers, P. G. *et al.* Highly variable clinical manifestations in a large family with a novel GATA2 mutation. *Leukemia* **27**, 2247-2248, doi:leu2013105 [pii]

10.1038/leu.2013.105 (2013).

11 Wang, X. *et al.* GATA2 and secondary mutations in familial myelodysplastic syndromes and pediatric myeloid malignancies. *Haematologica* **100**, e398-401, doi:haematol.2015.127092 [pii]

10.3324/haematol.2015.127092 (2015).

12 Rodrigues, N. P. *et al.* Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. *Blood* **106**, 477-484, doi:2004-08-2989 [pii]

10.1182/blood-2004-08-2989 (2005).

13 Ling, K. W. *et al.* GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells. *J Exp Med* **200**, 871-882, doi:jem.20031556 [pii]

10.1084/jem.20031556 (2004).

14 Butko, E. *et al.* Gata2b is a restricted early regulator of hemogenic endothelium in the zebrafish embryo. *Development* **142**, 1050-1061, doi:10.1242/dev.119180 (2015).

15 Dobrzycki, T. *et al.* Deletion of a conserved Gata2 enhancer impairs haemogenic endothelium programming and adult Zebrafish haematopoiesis. *Commun Biol* **3**, 71, doi:10.1038/s42003-020-0798-3

10.1038/s42003-020-0798-3 [pii] (2020).

16 Gioacchino, E. *et al.* Essential role for Gata2 in modulating lineage output from hematopoietic stem cells in zebrafish. *Blood Adv* **5**, 2687-2700, doi:S2473-9529(21)00348-7 [pii]

10.1182/bloodadvances.2020002993 (2021).

17 Avagyan, S. *et al.* Single-cell ATAC-seq reveals GATA2-dependent priming defect in myeloid and a maturation bottleneck in lymphoid lineages. *Blood Adv* **5**, 2673-2686, doi:S2473-9529(21)00347-5 [pii]

10.1182/bloodadvances.2020002992 (2021).

18 Ma, D., Zhang, J., Lin, H. F., Italiano, J. & Handin, R. I. The identification and characterization of zebrafish hematopoietic stem cells. *Blood* **118**, 289-297, doi:blood-2010-12-327403 [pii]

10.1182/blood-2010-12-327403 (2011).

19 Tamplin, O. J. *et al.* Hematopoietic stem cell arrival triggers dynamic remodeling of the perivascular niche. *Cell* **160**, 241-252, doi:10.1016/j.cell.2014.12.032 (2015).

20 Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821, doi:S0092-8674(19)30559-8 [pii]

10.1016/j.cell.2019.05.031 (2019).

21 Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386, doi:10.1038/nbt.2859 (2014).
22 Boatman, S. *et al.* Assaying hematopoiesis using zebrafish. *Blood Cells Mol Dis* **51**, 271-276, doi:S1079-9796(13)00160-5 [pii]

10.1016/j.bcmd.2013.07.009 (2013).
23 Traver, D. *et al.* Transplantation and in vivo imaging of multilineage engraftment in zebrafish bloodless mutants. *Nat Immunol* **4**, 1238-1246, doi:10.1038/ni1007

ni1007 [pii] (2003).
24 Langenau, D. M. *et al.* In vivo tracking of T cell development, ablation, and engraftment in transgenic zebrafish. *Proc Natl Acad Sci U S A* **101**, 7369-7374, doi:10.1073/pnas.0402248101

0402248101 [pii] (2004).
25 Renshaw, S. A. *et al.* A transgenic zebrafish model of neutrophilic inflammation. *Blood* **108**, 3976-3978, doi:blood-2006-05-024075 [pii]

10.1182/blood-2006-05-024075 (2006).
26 Ellett, F., Pase, L., Hayman, J. W., Andrianopoulos, A. & Lieschke, G. J. mpeg1 promoter transgenes direct macrophage-lineage expression in zebrafish. *Blood* **117**, e49-56, doi:blood-2010-10-314120 [pii]

10.1182/blood-2010-10-314120 (2011).
27 Page, D. M. *et al.* An evolutionarily conserved program of B-cell development and activation in zebrafish. *Blood* **122**, e1-11, doi:blood-2012-12-471029 [pii]

10.1182/blood-2012-12-471029 (2013).
28 Carmona, S. J. *et al.* Single-cell transcriptome analysis of fish immune cells provides insight into the evolution of vertebrate immune cell types. *Genome Res* **27**, 451-461, doi:10.1101/gr.207704.116 (2017).
29 Danilova, N., Bussmann, J., Jekosch, K. & Steiner, L. A. The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z. *Nat Immunol* **6**, 295-302, doi:ni1166 [pii]

10.1038/ni1166 (2005).
30 Kortum, A. N. *et al.* Differential expression and ligand binding indicate alternative functions for zebrafish polymeric immunoglobulin receptor (pIgR) and a family of pIgR-like (PIGRL) proteins. *Immunogenetics* **66**, 267-279, doi:10.1007/s00251-014-0759-4 (2014).
31 Macaulay, I. C. *et al.* Single-Cell RNA-Sequencing Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells. *Cell Rep* **14**, 966-977, doi:10.1016/j.celrep.2015.12.082 (2016).
32 Tang, Q. *et al.* Dissecting hematopoietic and renal cell heterogeneity in adult zebrafish at single-cell resolution using RNA sequencing. *J Exp Med* **214**, 2875-2887, doi:10.1084/jem.20170976 (2017).
33 Athanasiadis, E. I. *et al.* Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. *Nat Commun* **8**, 2045, doi:10.1038/s41467-017-02305-6 (2017).

34 Monika Belickova, M. *et al.* Up-regulation of ribosomal genes is associated with a poor response to azacitidine in myelodysplasia and related neoplasms. *Int J Hematol* **104**, 566-573, doi:10.1007/s12185-016-2058-3

10.1007/s12185-016-2058-3 [pii] (2016).

35 Nakayama, Y. & Yamaguchi, N. Multi-lobulation of the nucleus in prolonged S phase by nuclear expression of Chk tyrosine kinase. *Exp Cell Res* **304**, 570-581, doi:S0014-4827(04)00703-7 [pii]

10.1016/j.yexcr.2004.11.027 (2005).

36 Ullah, Z., Lee, C. Y. & Depamphilis, M. L. Cip/Kip cyclin-dependent protein kinase inhibitors and the road to polyploidy. *Cell Div* **4**, 10, doi:1747-1028-4-10 [pii]

10.1186/1747-1028-4-10 (2009).

37 Moriguchi, T. & Yamamoto, M. A regulatory network governing Gata1 and Gata2 gene transcription orchestrates erythroid lineage differentiation. *Int J Hematol* **100**, 417-424, doi:10.1007/s12185-014-1568-0 (2014).

38 Grass, J. A. *et al.* GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proc Natl Acad Sci U S A* **100**, 8811-8816, doi:10.1073/pnas.1432147100

1432147100 [pii] (2003).

39 Pal, S. *et al.* Coregulator-dependent facilitation of chromatin occupancy by GATA-1. *Proc Natl Acad Sci U S A* **101**, 980-985, doi:10.1073/pnas.0307612100

0307612100 [pii] (2004).

40 Bresnick, E. H., Lee, H. Y., Fujiwara, T., Johnson, K. D. & Keles, S. GATA switches as developmental drivers. *J Biol Chem* **285**, 31087-31093, doi:S0021-9258(19)88840-3 [pii]

10.1074/jbc.R110.159079 (2010).

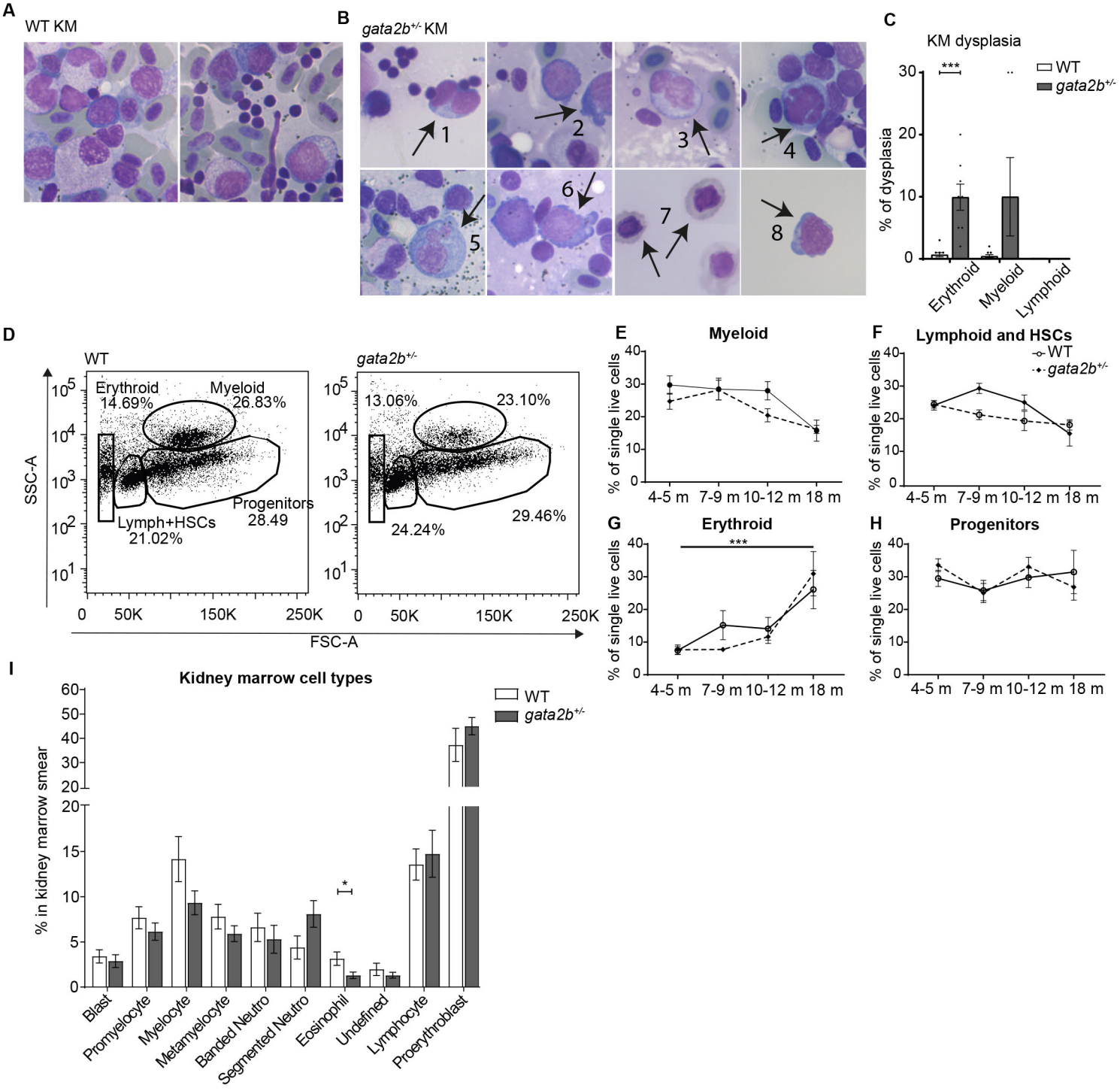
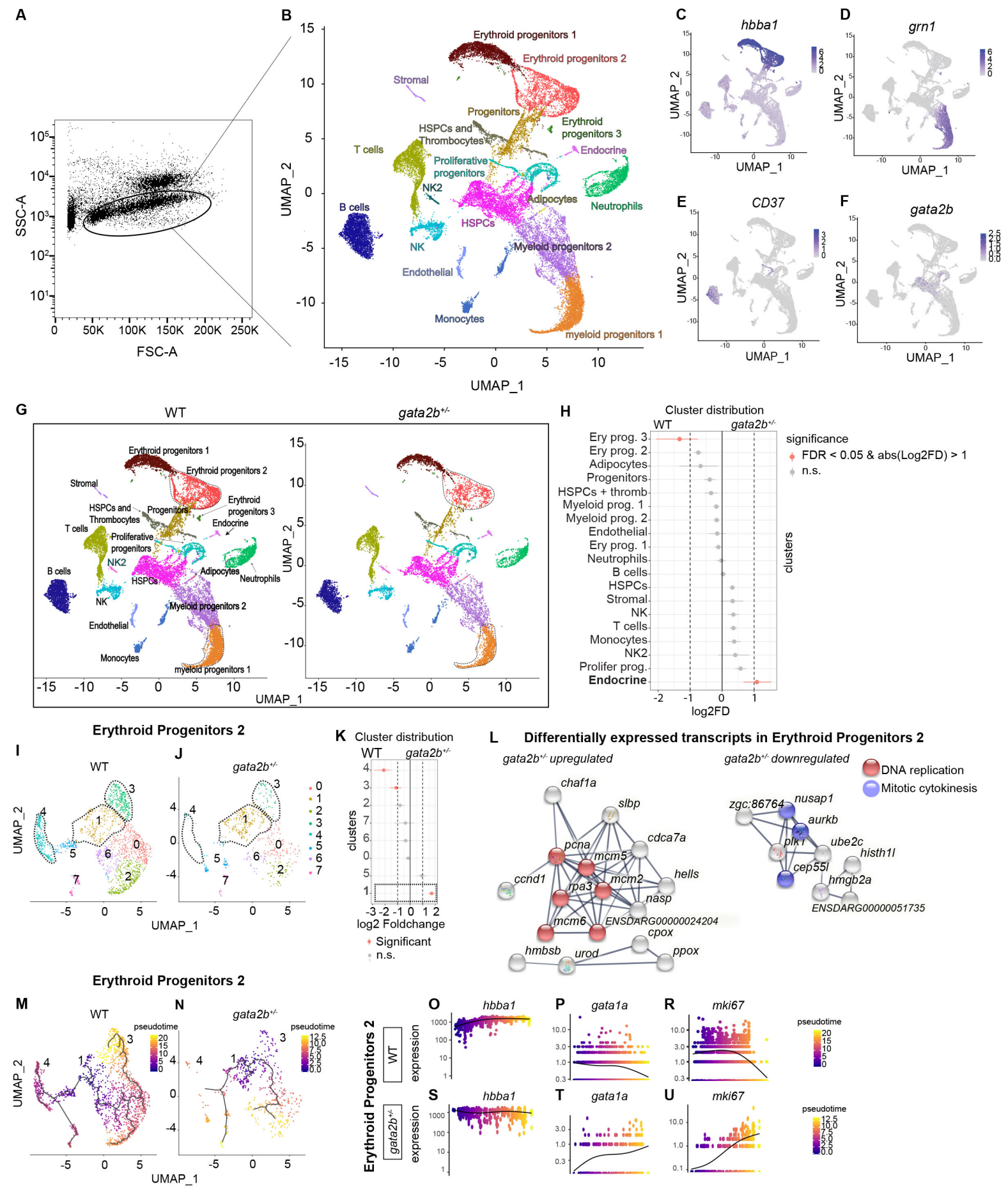

A
WT KM
B
gata2b+/- KM
1
2
3
4
5
6
7
8
C
KM dysplasia
WT
gata2b+/-
% of dysplasia
***
Erythroid
Myeloid
Lymphoid
D
WT
gata2b+/-
SSC-A
FSC-A
Erythroid 14.69%
Myeloid 26.83%
Progenitors 28.49
Lymph+HSCs 21.02%
13.06%
23.10%
29.46%
24.24%
E
Myeloid
F
Lymphoid and HSCs
G
Erythroid
H
Progenitors
% of single live cells
4-5 m
7-9 m
10-12 m
18 m
I
Kidney marrow cell types
% in kidney marrow smear
Blast
Promyelocyte
Myelocyte
Metamyelocyte
Banded Neutro
Segmented Neutro
Eosinophil
Undefined
Lymphocyte
Proerythroblast

Figure 1


A
SSC-A
FSC-A
B
Erythroid progenitors 1
Erythroid progenitors 2
Stromal
Progenitors
HSPCs and Thrombocytes
Erythroid progenitors 3
T cells
Endocrine
Proliferative progenitors
NK2
Adipocytes
Neutrophils
B cells
HSPCs
NK
Myeloid progenitors 2
Endothelial
Monocytes
myeloid progenitors 1
UMAP_1
UMAP_2
C
hbba1
D
grn1
E
CD37
F
gata2b
G
WT
gata2b+/-
H
Cluster distribution
WT
gata2b+/-
significance
FDR < 0.05 & abs(Log2FD) > 1
n.s.
Ery prog. 3
Ery prog. 2
Adipocytes
Progenitors
HSPCs + thromb
Myeloid prog. 1
Myeloid prog. 2
Endothelial
Ery prog. 1
Neutrophils
B cells
HSPCs
Stromal
NK
T cells
Monocytes
NK2
Prolifer prog.
Endocrine
clusters
log2FD
Erythroid Progenitors 2
I
J
K
Cluster distribution
log2 Foldchange
Significant
n.s.
L
Differentially expressed transcripts in Erythroid Progenitors 2
gata2b+/- upregulated
gata2b+/- downregulated
DNA replication
Mitotic cytokinesis
chaf1a
slbp
cdca7a
pcna
mcm5
ccnd1
rpa3
mcm2
hells
nasp
mcm6
ENSDARG00000024204
cpox
hmbsb
urod
ppox
zgc:86764
nusap1
aurkb
plk1
ube2c
histh1l
cep55l
hmgb2a
ENSDARG00000051735
Erythroid Progenitors 2
M
N
pseudotime
O
hbba1
P
gata1a
R
mki67
S
T
U
Erythroid Progenitors 2
WT
gata2b+/-
expression

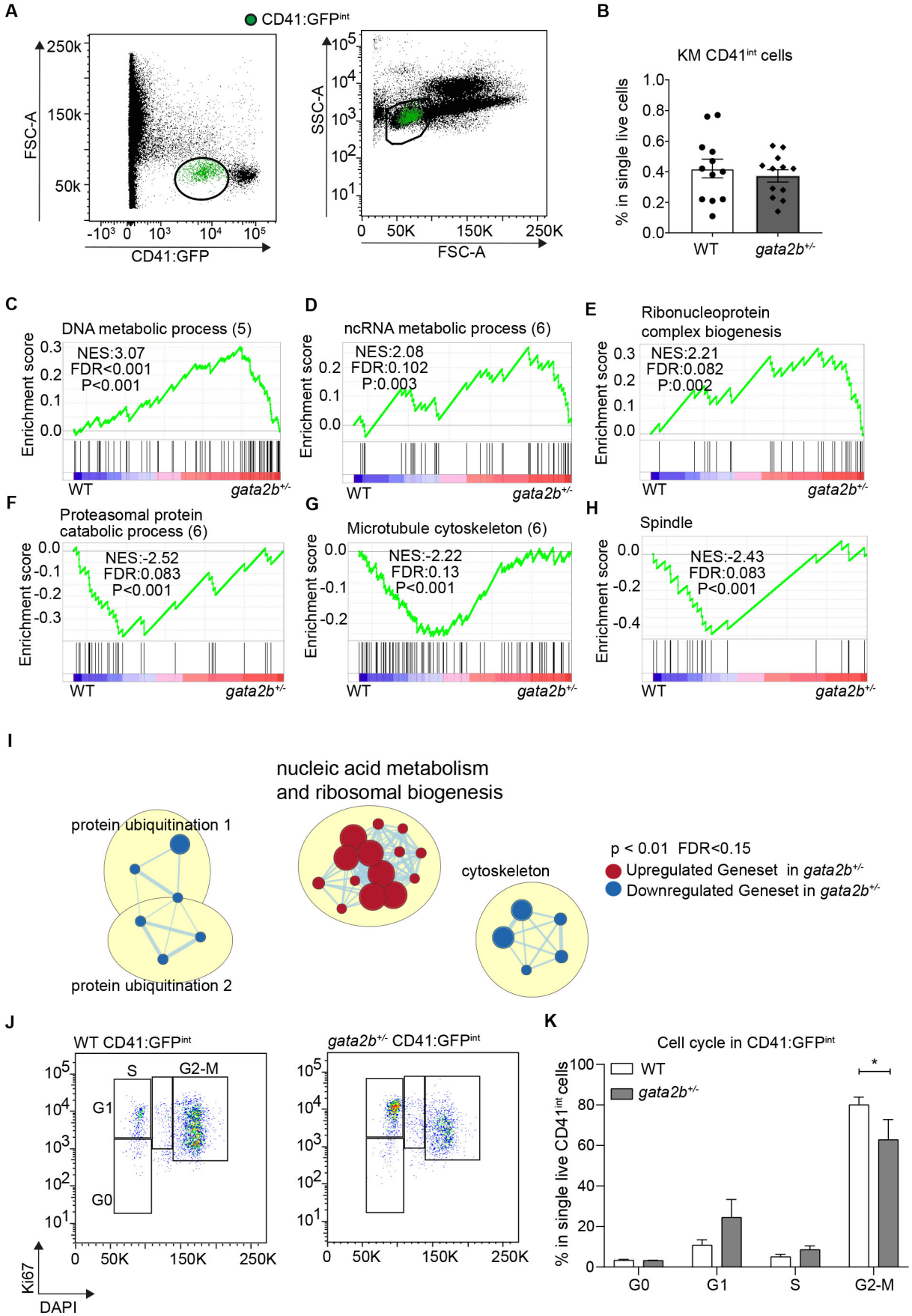

A
CD41:GFPint
FSC-A
250k
150k
50k
CD41:GFP
SSC-A
FSC-A
B
KM CD41int cells
% in single live cells
WT
gata2b+/-
C
DNA metabolic process (5)
NES:3.07
FDR<0.001
P<0.001
Enrichment score
D
ncRNA metabolic process (6)
NES:2.08
FDR:0.102
P:0.003
E
Ribonucleoprotein complex biogenesis
NES:2.21
FDR:0.082
P:0.002
F
Proteasomal protein catabolic process (6)
NES:-2.52
FDR:0.083
P<0.001
G
Microtubule cytoskeleton (6)
NES:-2.22
FDR:0.13
P<0.001
H
Spindle
NES:-2.43
FDR:0.083
P<0.001
I
nucleic acid metabolism and ribosomal biogenesis
protein ubiquitination 1
protein ubiquitination 2
cytoskeleton
p < 0.01 FDR<0.15
Upregulated Geneset in gata2b+/-
Downregulated Geneset in gata2b+/-
J
WT CD41:GFPint
gata2b+/- CD41:GFPint
S
G2-M
G1
G0
Ki67
DAPI
K
Cell cycle in CD41:GFPint
% in single live CD41int cells
WT
gata2b+/-
G0
G1
S
G2-M

Figure 3