Lineage skewing and genome instability underlie marrow failure in a zebrafish model of GATA2 deficiency

Mahony, Christopher B; Copper, Lucy; Vrljicak, Pavle; Noyvert, Boris; Constantinidou, Chrystala; Browne, Sofia; Pan, Yi; Palles, Claire; Ott, Sascha; Higgs, Martin R; Monteiro, Rui

DOI:
10.1016/j.celrep.2023.112571

Document Version
Publisher's PDF, also known as Version of record

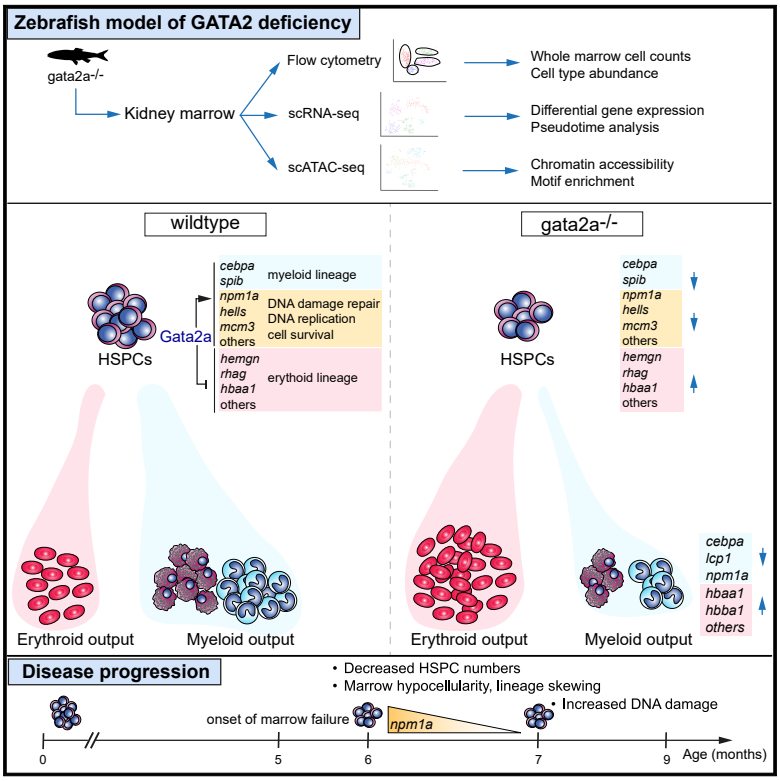
Citation for published version (Harvard):
Mahony, CB, Copper, L, Vrljicak, P, Noyvert, B, Constantinidou, C, Browne, S, Pan, Y, Palles, C, Ott, S, Higgs, MR & Monteiro, R 2023, 'Lineage skewing and genome instability underlie marrow failure in a zebrafish model of GATA2 deficiency', *Cell Reports*, vol. 42, no. 6, 112571. https://doi.org/10.1016/j.celrep.2023.112571

Cell Reports

Article

Lineage skewing and genome instability underlie marrow failure in a zebrafish model of GATA2 deficiency

Graphical abstract



Highlights

- Zebrafish *gata2a* enhancer mutants recapitulate human GATA2 deficiency disease
- A Gata2a/Cebpa axis maintains marrow cellularity and myeloid lineage output
- Gata2a regulates expression of DNA damage repair and replication-associated genes
- Progressive loss of *npm1a* in mutant HSPCs is associated with increased DNA damage

Authors

Christopher B. Mahony, Lucy Copper, Pavle Vrljicak, ..., Sascha Ott, Martin R. Higgs, Rui Monteiro

Correspondence

r.monteiro@bham.ac.uk

In brief

Mahony et al. use single-cell genomics to uncover the molecular defects and examine progression to marrow failure in a zebrafish model of GATA2 deficiency. They show that Gata2a/Cebpa regulates marrow cellularity and myeloid lineage output, whereas HSPC survival and genome stability are protected by maintaining expression of DNA damage repair and replication-associated genes.

Mahony et al., 2023, Cell Reports *42*, 112571
June 27, 2023
https://doi.org/10.1016/j.celrep.2023.112571

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Lineage skewing and genome instability underlie marrow failure in a zebrafish model of GATA2 deficiency

Christopher B. Mahony,[1] Lucy Copper,[1,2] Pavle Vrljicak,[3] Boris Noyvert,[4] Chrystala Constantinidou,[3,5] Sofia Browne,[1] Yi Pan,[4] Claire Palles,[1] Sascha Ott,[3,5] Martin R. Higgs,[1] and Rui Monteiro[1,6,*]

[1]Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK
[2]Cancer Research UK Birmingham Centre, Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK
[3]Division of Biomedical Sciences, Warwick Medical School, University of Warwick, Coventry, UK
[4]Centre for Computational Biology, Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK
[5]Bioinformatics Research Technology Platform, University of Warwick, Coventry, UK
[6]Lead contact
*Correspondence: r.monteiro@bham.ac.uk
https://doi.org/10.1016/j.celrep.2023.112571

SUMMARY

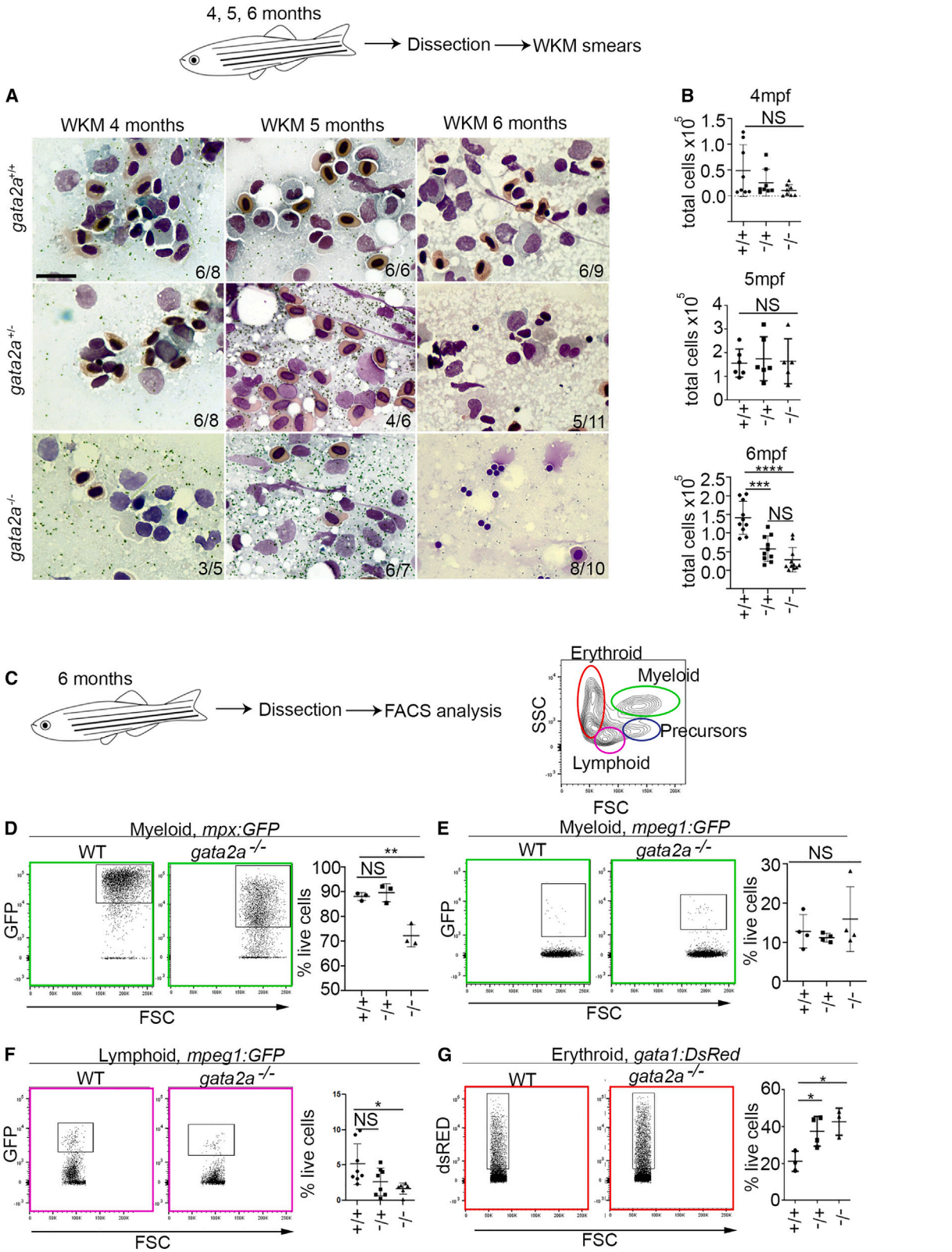
Inherited bone marrow failure associated with heterozygous mutations in GATA2 predisposes toward hematological malignancies, but the mechanisms remain poorly understood. Here, we investigate the mechanistic basis of marrow failure in a zebrafish model of GATA2 deficiency. Single-cell transcriptomics and chromatin accessibility assays reveal that loss of *gata2a* leads to skewing toward the erythroid lineage at the expense of myeloid cells, associated with loss of *cebpa* expression and decreased PU.1 and CEBPA transcription factor accessibility in hematopoietic stem and progenitor cells (HSPCs). Furthermore, *gata2a* mutants show impaired expression of *npm1a*, the zebrafish NPM1 ortholog. Progressive loss of *npm1a* in HSPCs is associated with elevated levels of DNA damage in *gata2a* mutants. Thus, Gata2a maintains myeloid lineage priming through *cebpa* and protects against genome instability and marrow failure by maintaining expression of *npm1a*. Our results establish a potential mechanism underlying bone marrow failure in GATA2 deficiency.

INTRODUCTION

The development of hematopoietic stem and progenitor cells (HSPCs) from the aortic endothelium during embryogenesis is a highly conserved process across many species. HSPCs mature and eventually seed their adult niche (the bone marrow in mammals and kidney marrow in zebrafish) where they self-renew, differentiate to all blood lineages, and sustain hematopoiesis.[1,2] Given its importance, lineage commitment and differentiation are tightly controlled by the action of key transcription factors.

GATA2 is one such factor, required for the formation and survival of HSPCs.[3–5] In humans, *GATA2* haploinsufficiency due to mutations in coding or enhancer regions causes hematopoietic disorders collectively referred to as GATA2 deficiency syndromes.[6,7] These patients suffer from recurrent mycobacterial or viral infections and lymphedema and develop marrow failure.[8] Most patients with inherited germline GATA2 mutations (~75%) develop early-onset myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML) at a median age of 20.[9,10] Importantly, existing mouse models fail to mimic the disease progression observed in humans: loss of one GATA2 allele does not induce GATA2 deficiency, and homozygous mutants are embryonic lethal.[4,11] Moreover, adult mutants with homozygous mutations in the GATA2 9.5 kb enhancer only reveal hematopoietic repopulation defects upon additional exogenous stress.[12] In contrast, by deleting a conserved intronic endothelial enhancer in the zebrafish *gata2a* locus (hereafter referred to as *gata2a*$^{-/-}$ mutants), we have generated a *bona fide* model of *GATA2* deficiency[13] that shows marrow hypocellularity, neutropenia, and susceptibility to infections and develops an AML-like phenotype in the adult kidney marrow.[13]

Here, we make use of this model to investigate disease progression and understand how loss of GATA2 leads to perturbed hematopoiesis, marrow failure, and predisposition to MDS/AML. We study the transcriptional changes that occur at the single-cell level in mutant hematopoietic cells and demonstrate the impact of reduced GATA2 levels: *gata2a*$^{-/-}$ mutants display marrow hypocellularity from 6 months post-fertilization (mpf) onward, accompanied by neutropenia, monocytopenia, and increased numbers of erythrocytes with a concomitant deregulation of gene expression and chromatin accessibility at erythroid and myeloid genes. In addition, mutant HSPCs show decreased expression of key DNA replication and DNA damage repair genes, including *npm1a* and the prereplication complex, suggesting that defective replication/repair plays a role in the bone marrow failure phenotype in GATA2 deficiency. In particular, *npm1a* expression is lost as disease progresses and correlates



(legend on next page)

with increased DNA damage in whole kidney marrow (WKM) and in *in vitro* cultures of hematopoietic progenitors. Taken together, we demonstrate that *gata2a* is required to initiate and maintain lineage priming in HSPCs, regulating HSPC survival and protecting against marrow failure by maintaining the expression of essential DNA damage repair genes, thus protecting against genome instability.

RESULTS

Gata2a directs myeloid lineage differentiation and maintains WKM homeostasis

Our previous studies revealed that deletion of the gata*2a* i4 enhancer impaired hemogenic endothelium (HE) programming, although HSPC numbers in the embryonic niche recover by 5 days post-fertilization (dpf).[13] However, adults exhibit hypocellular WKM and neutropenia, and 1 in 10 show >98% blasts in WKM by 9 mpf, a hallmark of AML. Together, this indicated that these enhancer deletion mutants displayed key characteristics of GATA2 deficiency in humans.[13] To uncover the underlying mechanisms and to shed light on human disease, we first determined the precise onset of marrow hypocellularity by analyzing *gata2a* mutants at 4, 5, and 6 mpf using WKM smears and cell counts (Figures 1A and 1B). These experiments revealed that hypocellularity became apparent by 6 mpf in both heterozygous *gata2a*$^{+/-}$ and homozygous *gata2a*$^{-/-}$ mutants, indicating that maintenance of normal WKM cell numbers is highly sensitive to *gata2a* dosage (Figures 1A and 1B). To examine how HSPC differentiation was affected, *gata2a*$^{-/-}$ mutants were crossed to transgenic lines, enabling the labeling of various hematopoietic cell types, and these populations were quantified by flow cytometry (Figures 1C–1G).[14] As expected, we found decreased numbers of neutrophils in *gata2a*$^{-/-}$ animals (Figure 1D) but no changes in macrophages (Figure 1E). In addition, B cell numbers were decreased (Figure 1F), while erythrocytes were increased (Figure 1G). Together, these data showed that Gata2a is required to maintain hematopoietic cell numbers in WKM and for correct lineage differentiation in the adult marrow. No significant skewing in lineage differentiation was found in *gata2a* heterozygous mutants, indicating that maintenance of cellular homeostasis is more sensitive to Gata2a dosage than lineage determination.

gata2a$^{-/-}$ HSPCs display lineage skewing toward erythroid fates at the expense of myeloid cell fates

To investigate how *gata2a* regulates WKM cellularity and lineage differentiation, we undertook transcriptional and chromatin profiling of WKM at single-cell resolution at 5 mpf, prior to the onset of detectable lineage skewing and hypocellularity. We took two adult fish per genotype, dissected and pooled WKM, and sorted live cells for single-cell RNA sequencing (scRNA-seq). scRNA-seq could readily identify cell clusters representing all major hematopoietic and kidney cell populations present in WKM (Figure 2A), consistent with previous studies.[18,19] Using these data, we defined "core HSPCs" (cHSPCs) by expression of *myb*, *lmo2*, *cebpa*, *meis1b*, *gata2b*, and *runx1* (Figure S1A). These cells were also enriched in other known HSPC markers such as *gfi1aa*, *dnmt3ba*, and *angpt1.* They also expressed *si:dkey-261h17.1*, identified as the zebrafish ortholog of *CD34* (https://zfin.org). Phylogeny and synteny analyses revealed conservation of the CD34 antigen as defined by ENSEMBL (Figures S2A–S2C). Therefore, we hereafter refer to this gene as *cd34-like* or *cd34l* (Figure S1A). We defined a second HSPC population as "myeloid-primed" HSPCs (mHSPCs) based on a lack of expression of markers such as *gata2b, runx1*, and *meis1b*, retention of myeloid-biased transcription factors (*lmo2*, *cebpa*, and *myb*), and higher expression levels of *c1qa* and *mafba* (Figure S1A). We then defined the remaining populations based on expression of key marker genes (*igfbp1a*, *hemgn*, *klf1*: erythroid cells [Erys] and erythroid progenitors [Ery-Ps]; *c1qa*, *mfap4*, *mafba*: monocytes/macrophages; *mpx*, *mmp13a*, *vamp8*: neutrophils; *nkl*.3, *il2rb*: natural killer [NK]/T cells; *pax5*, *cd37*, *cd79b*: B cells; *sox8b*, *tekt1*: multiciliated kidney cells, *spink2*.2; *viml*: kidney tubule cells) (Figure S1A).

Next, we took three adult fish per genotype, dissected WKM, and isolated nuclei for single-cell assay for transposase accessible chromatin sequencing (scATAC-seq). First, we identified cell populations based upon promoter chromatin accessibility using the same markers as the previous scRNA-seq experiment shown in Figure 2A (Figure S1B). We included *dnmt3ba* and *fgfr2*, as these were highly enriched in cHSPCs, and *c1qb*, as this was highly enriched in macrophages (Figure S1B). Then, we confirmed that chromatin accessibility was lost where the i4 enhancer had been deleted in *gata2a*$^{-/-}$ mutant cells (Figure S1C). We also ensured that sample distribution was even within specific clusters (Figures S3A and S3B). When we compared wild-type and *gata2a*$^{-/-}$ WKM, we found a decrease in neutrophils and a small increase in erythroid cells (Figure 2B). We then examined RNA expression and chromatin accessibility of key marker genes in cHSPCs, mHSPCs, and Ery-P to ensure that these populations were correctly defined in both experiments (Figure 2C). We found that *gata2a*, *cd34l*, and *meis1b* RNA expression and chromatin accessibility were enriched in

Figure 1. Loss of *gata2a* results in hypocellular marrow, neutropenia, B lymphopenia, and erythrocytosis at 6 mpf

(A) WKM smears and May Grunwald/Giemsa staining at 4, 5, and 6 mpf. Scale bar, 10 μm.

(B) Total WKM cell counts at 4 (n = 8/condition [biological replicates]), 5 (n = 6/condition [biological replicates] except *gata2a*$^{-/-}$, where n = 5) and 6 mpf (n = 10/condition [biological replicates] except *gata2a*$^{-/-}$, where n = 11). Error bars are mean ± SD.

(C) Flow cytometry gating strategy.

(D–G) Analysis of different lineages using either *mpx:GFP*, *mpeg1:GFP*, or *gata1:dsRed* transgenic lines. We examined neutrophils (*mpx:GFP*, myeloid gate),[15] macrophages (*mpeg1.1:GFP*, myeloid gate), B cells (*mpeg1.1:GFP*, lymphoid gate),[16] and erythrocytes (*gata1:dsRed*, erythroid gate)[17] at 6 mpf.

(D) *gata2a*$^{+/+}$ vs. *gata2a*$^{+/-}$: p = 0.82, *gata2a*$^{+/+}$ vs. *gata2a*$^{-/-}$: p = 0.0024, n = 3/condition (biological replicates).

(E) *gata2a*$^{+/+}$ vs. *gata2a*$^{+/-}$: p = 0.87, *gata2a*$^{+/+}$ vs. *gata2a*$^{-/-}$: p = 0.64, n = 4/condition (biological replicates).

(F) *gata2a*$^{+/+}$ vs. *gata2a*$^{+/-}$: p = 0.053, *gata2a*$^{+/+}$ vs. *gata2a*$^{-/-}$: p = 0.0145, n = 8/condition (biological replicates) except *gata2a*$^{-/-}$, where n = 6.

(G) *gata2a*$^{+/+}$ vs. *gata2a*$^{+/-}$: p = 0.0374, *gata2a*$^{+/+}$ vs. *gata2a*$^{-/-}$: p = 0.0146, n = 3/condition (biological replicates) except *gata2a*$^{+/-}$, where n = 4.

Analysis was completed using one-way ANOVA, not significant (NS) p > 0.05, *p < 0.05. (B and D–G) Each dot represents one sample. Data shown as mean ± SD.

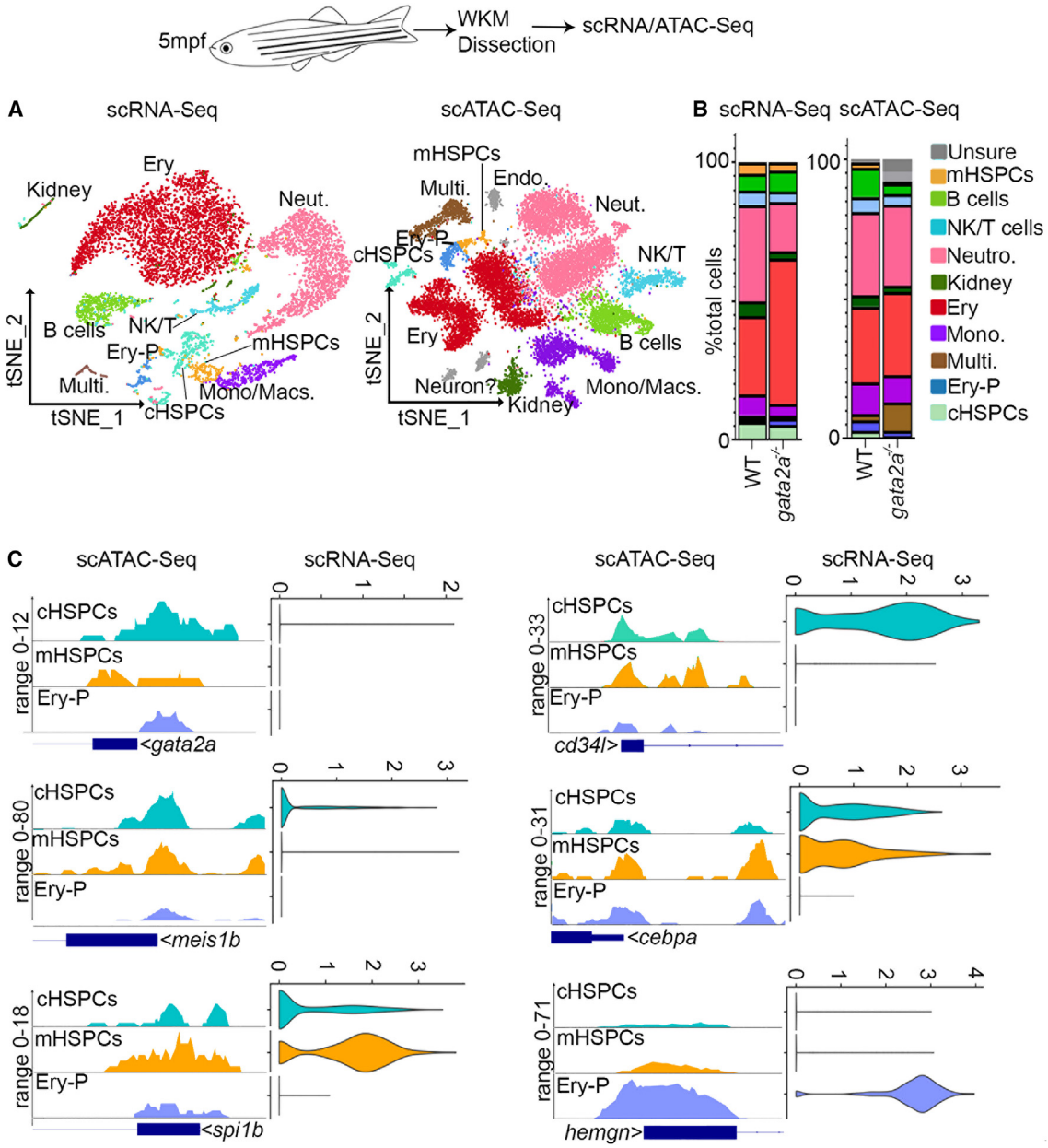


Figure 2. *gata2a* is required for lineage specification and transcriptional regulation in WKM hematopoietic populations

(A) scRNA-seq (left) and scATAC-seq (right) t-distributed stochastic neighbor embedding (tSNE) clustering of WKM hematopoietic populations. Analysis estimated 3,047 WT cells and 5,584 *gata2a*$^{-/-}$ cells for scRNA-seq (from 3 WT/*gata2a*$^{-/-}$ fish pooled to give one WT/mutant [MUT] sample that was sequenced) and 10,004 WT cells and 6,500 *gata2a*$^{-/-}$ cells for scATAC-seq (n = 3/condition [biological replicates]). Multi, multiciliated kidney (stromal) cells; cHSPCs, core HSPCs; Neut, neutrophils; NK/T, natural killer and T cells; Mono/Macs, monocytes/macrophages; Ery, erythroytes; mHSPCs, myeloid-biased HSPCs; Ery-P, erythroid progenitors; Endo, endothelial cells. Full list of differentially expressed genes can be found in Table S2.

(B) Percentage of WKM populations of total WT or *gata2a*$^{-/-}$ cells.

(C) Chromatin accessibility near promoter regions (scATAC-seq) and RNA expression (scRNA-seq, RNA assay) of indicated genes. Each dot represents a cell. cHSPC, core HSPC; mHSPC, myeloid-biased HSPC; Ery-P, erythroid progenitors.

See also Figures S1 and S2 and Tables S1 and S5.

cHSPCs (Figures 2C and S1B), *cebpa* and *spi1b* RNA expression and chromatin accessibility were enriched in mHSPCs, and *hemgn* RNA expression and chromatin accessibility were enriched in Ery-P (Figure 2C). Taken together, these data indicate that the lineage skewing observed at the cellular level at 6 mpf was already detectable at the molecular level (gene expression and chromatin accessibility) 1 month earlier in hematopoietic populations in WKM.

To identify the molecular changes underlying the lineage skewing, we examined differential gene expression in Ery-Ps, neutrophils, monocytes/macrophages, erythrocytes, NK/T cells, and B cells between genotypes (Figure S3C). Notably, erythrocytes from *gata2a*$^{-/-}$ animals exhibited a decrease in key erythrocytic genes (*alas2*, *hbaa2*, *hbba2*, *hbaa1*). This suggested that while their overall number was increased, erythrocyte function was likely impaired in the absence of gata2a. *gata2a*$^{-/-}$ monocytes showed a decrease in the myeloid markers *lcp1* and *cebpa* (Figure S3C). *gata2a*$^{-/-}$ mutant neutrophils also expressed significantly less *lcp1* and more *si:ch211-5k11.8* (orthologous to human HBZ-hemoglobin zeta, https://zfin.org) and hemoglobins (*hbba1*, *hbaa1*, *hbba2*, *hbba1.1)* (Figure S3C). Several globin genes including *hbba1* and *hbaa1* were upregulated in mutant B and NK/T cells (Figure S3C). Within the Ery-Ps, *hemgn* and hemoglobin genes were upregulated, corroborating the increase in erythrocytes observed previously (Figure S3C). Taken together, our findings demonstrate that myeloid and lymphoid hematopoietic populations in *gata2a*$^{-/-}$ mutants show downregulation of myeloid lineage genes and upregulation of genes of the erythroid lineage. This suggested that *gata2a* is required to maintain transcriptional states of myeloid and erythroid genes in the most differentiated hematopoietic populations. We next examined how loss of *gata2a* affected gene expression in cHSPCs and mHSPCs (Figures 3A and 3B). Gene Ontology (GO) term enrichment analysis highlighted only very few altered pathways in HSPCs, probably a reflection of the small number of affected genes detected in mutant cHSPCs (Figures 3A and S3D). Strikingly, however, gata*2a*$^{-/-}$ cHSPCs showed an increase in *hbba1.1* and *hbaa1* and a decrease in *lcp1* and *cebpa* (Figure 3A). Similar to the more differentiated hematopoietic populations (Figure 3B), *gata2a*$^{-/-}$ mHSPCs showed an increase in globin genes (e.g., *hbaa1.1*, *hbba2*) and *hemgn*[20] (Figure 3B). By contrast, *cebpa* and myeloid markers such as *lcp1* and *spi1b* were downregulated in mHSPCs (Figure 3B). Since *cebpa* is a key regulator of granulocyte and monocyte differentiation,[21–24] this suggests that the differentiation skewing toward the erythroid lineage in *gata2a*$^{-/-}$ mutants is likely driven by the loss of *cebpa* in mHSPCs (Figures 3B and 3D). We also found that critical DNA replication factors (e.g., *mcm3, mcm5*: Figure 3B) were downregulated in mHSPCs, suggesting impaired replication,[25,26] and broadly consistent with the cytopenias noted at 6 mpf. Accordingly, the GO term enrichment analysis showed that downregulated genes were enriched for pathways in DNA replication and chromatin organization (Figure 3C). In addition, we noted a decrease in expression of genes required for DNA repair and genome stability, such as *hells*,[27] *npm1a*,[28,29] and *ncl*[30,31] (Figure 3B).

To establish how our findings relate to molecular changes induced by GATA2 deficiency in humans, we compared our dataset with scRNA-seq data from a recent publication profiling HSCs from patients with GATA2 deficiency.[32] We took the differentially expressed genes in patient HSCs and compared them with differentially expressed genes in cHSPCs and mHSPCs of *gata2a*$^{-/-}$ mutants (Figure S3E). The genes overlapping both datasets included downregulated myeloid genes (e.g., *sp1b*, *lcp1*) and regulators of DNA damage and replication (e.g., *hells*, *npm1a*, *mcm3*, *mcm5*), as well as upregulation of erythroid genes (e.g., *cahz*, *hbba1.1*) (Figure S3E). Thus, GATA2 deficiency in the zebrafish model and in human patients shows overlapping dysregulation of key regulators of cellular function and myeloid and erythroid lineage differentiation.

To understand the molecular mechanism driving the hematopoietic phenotype in these mutants, we compared chromatin accessibility and RNA expression in cHSPCs and mHSPCs (Figures 3D and 3E). While we found striking differences in chromatin accessibility near the promoter region of *cebpa*, *npm1a*, *lcp1*, *hbba1*, and *hemgn*, only *cebpa* and *lcp1* showed significant changes in RNA expression in cHSPCs (Figure 3D). By contrast, we found little difference in chromatin accessibility in these two genes in mHSPCs, whereas their expression was clearly decreased in mutant cells (Figure 3E). Both *hbaa1.1* and *hemgn* were upregulated in mutant mHSPCs with few changes in chromatin accessibility in *gata2a*$^{-/-}$ mutants (Figure 3E). This suggested that changes in chromatin accessibility precede changes in RNA expression during hematopoietic differentiation cascade and that myeloid gene regulatory elements are primed for transcription by Gata2a.

Regulatory interactions between Gata2a, *cebpa*, and *npm1a*

To examine the regulatory relationship between *gata2a*, *cebpa*, and *npm1a*, we first analyzed publicly available GATA2 and CEBPA chromatin immunoprecipitation (ChIP)-seq data in HPC7 cells from datasets BioStudies:E-MTAB-3954, GEO:GSM552234, GEO:GSM1329815, and GEO:GSE47085,[33] using CODEX[34] (Figures S4A and S4B). Analysis of the *NPM1* locus showed no evidence for direct binding of GATA2 or CEBPA in the *NPM1* locus (Figure S4B), suggesting that the regulation of *NPM1* by GATA2 is likely indirect. In contrast, we found binding of GATA2 in an active *CEBPA* enhancer (+37 kb enhancer[35]) (Figure S4A), suggesting a direct regulation of *CEBPA* by GATA2. Indeed, *cebpa* expression in zebrafish primitive myeloid cells located in the posterior blood island (PBI) is decreased in *gata2a*$^{um27/um27}$ mutants,[36] a phenotype we could recapitulate by knocking down *gata2a* (*gata2a KD*) with a combination of three Alt-R gRNAs (Figure S4C). Notably, the proportion of embryos showing no *cebpa* expression in the PBI was also increased in *gata2a KD* embryos at 28 hours post-fertilization (hpf) (Figure S4D). This effect could be rescued by co-injection of murine GATA2 mRNA (Figure S4E), demonstrating that the phenotypes observed were directly attributable to loss of *gata2a*. Expression of *runx1*, normally downregulated in *gata2a*$^{-/-}$ embryos at 28 hpf,[13] was used as a positive control for this experiment (Figure S4F). We concluded that Gata2a regulates *cebpa* expression by binding to its enhancer directly, whereas *npm1a* expression is likely regulated indirectly.

To further examine how lineage skewing in *gata2a*$^{-/-}$ mutants progress across hematopoietic differentiation, we ordered the scRNA-seq data in pseudotime and analyzed the expression of a subset of erythroid and myeloid-specific genes (Figures 4A–4B′). We found that the erythroid gene *alas2* is expressed earlier in pseudotime and at higher levels in *gata2a*$^{-/-}$ cells compared with wild type (WT). By contrast, expression of the myeloid markers *cebpa*, *lcp1*, and *mpx* remains low in *gata2a*$^{-/-}$ mutants compared with WT cells. Taken together, these results

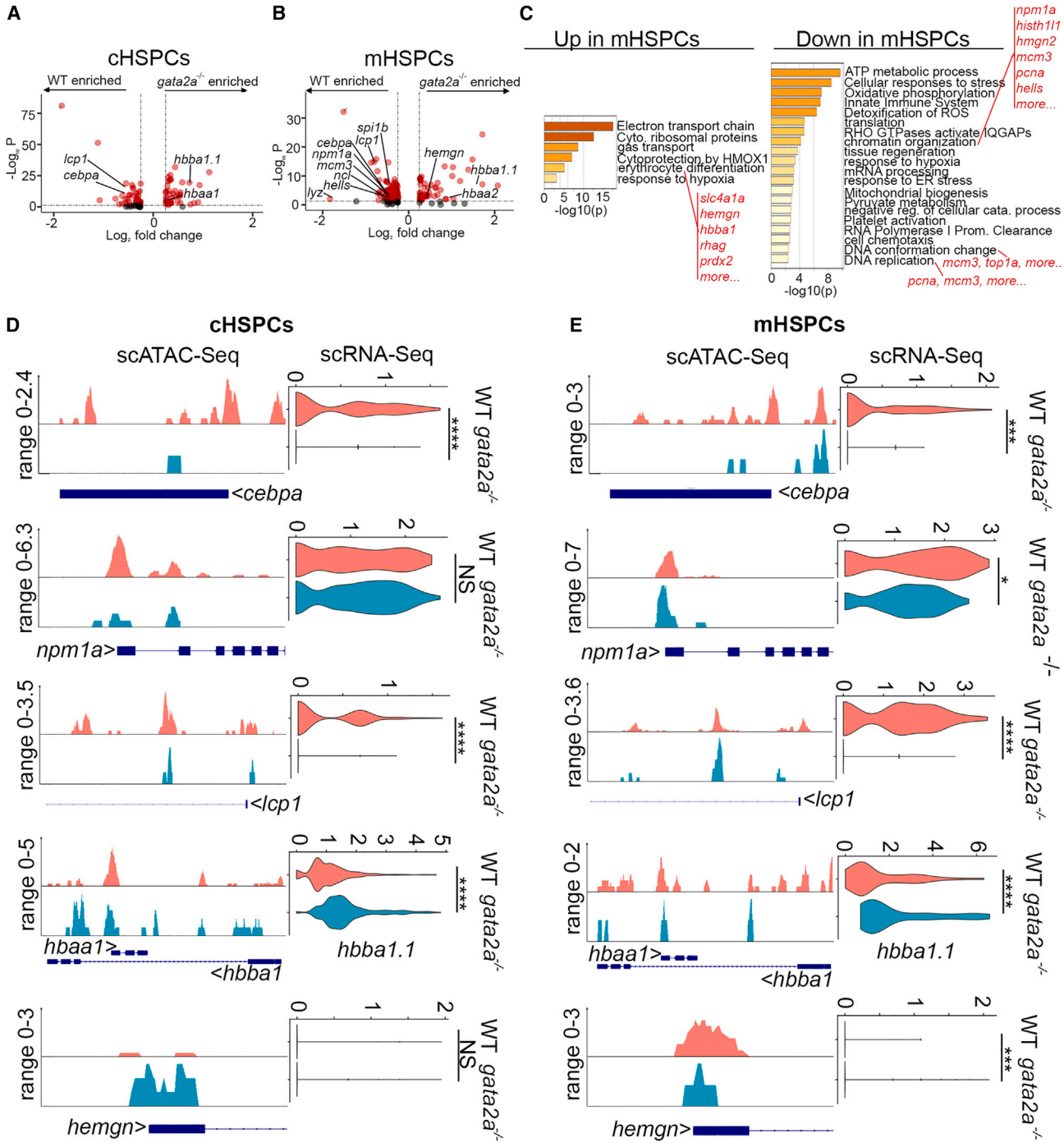


Figure 3. Differential regulation of chromatin accessibility and myeloid and erythroid gene expression by Gata2a in cHSPCs and mHSPCs

(A and B) Volcano plots of log2 fold change in gene expression vs. –log10(p value) in mHSPCs and cHSPCs. Each dot represents a gene. Dotted line indicates p = 0.05 and fold change = 0.25 with black genes falling below this threshold.

(C) GO term enrichment (performed using https://metascape.org) analysis of up- or downregulated genes in mHSPCs with a p <0.05. mHSPCs, myeloid HSPCs. Full list of differentially expressed genes can be found in Table S2.

(D and E) Chromatin accessibility near promoter regions (scATAC-seq) and RNA expression (scRNA-seq) of indicated genes in either cHSPCs (D) or mHSPCs (E). In (D), WT vs. *gata2a*$^{-/-}$: *cebpa* p = 3.36E−10; *npm1a* p > 0.05; *lcp1* p = 3.32E−15; *hbba1.1* p = 3.72E−18; *hemgn* p > 0.05. In (E), WT vs. *gata2a*$^{-/-}$: *cebpa* p = 0.00039; *npm1a* p = 0.034; *lcp1* p = 1.5E−15; *hbba1.1* p = 8.13E−8; *hemgn* p = 0.00085. cHSPC, core HSPC. mHSPCs, myeloid HSPCs.

NS p > 0.05, *p < 0.05, **p < 0.01, ****p < 0.0001. See also Figure S3 and Table S2.

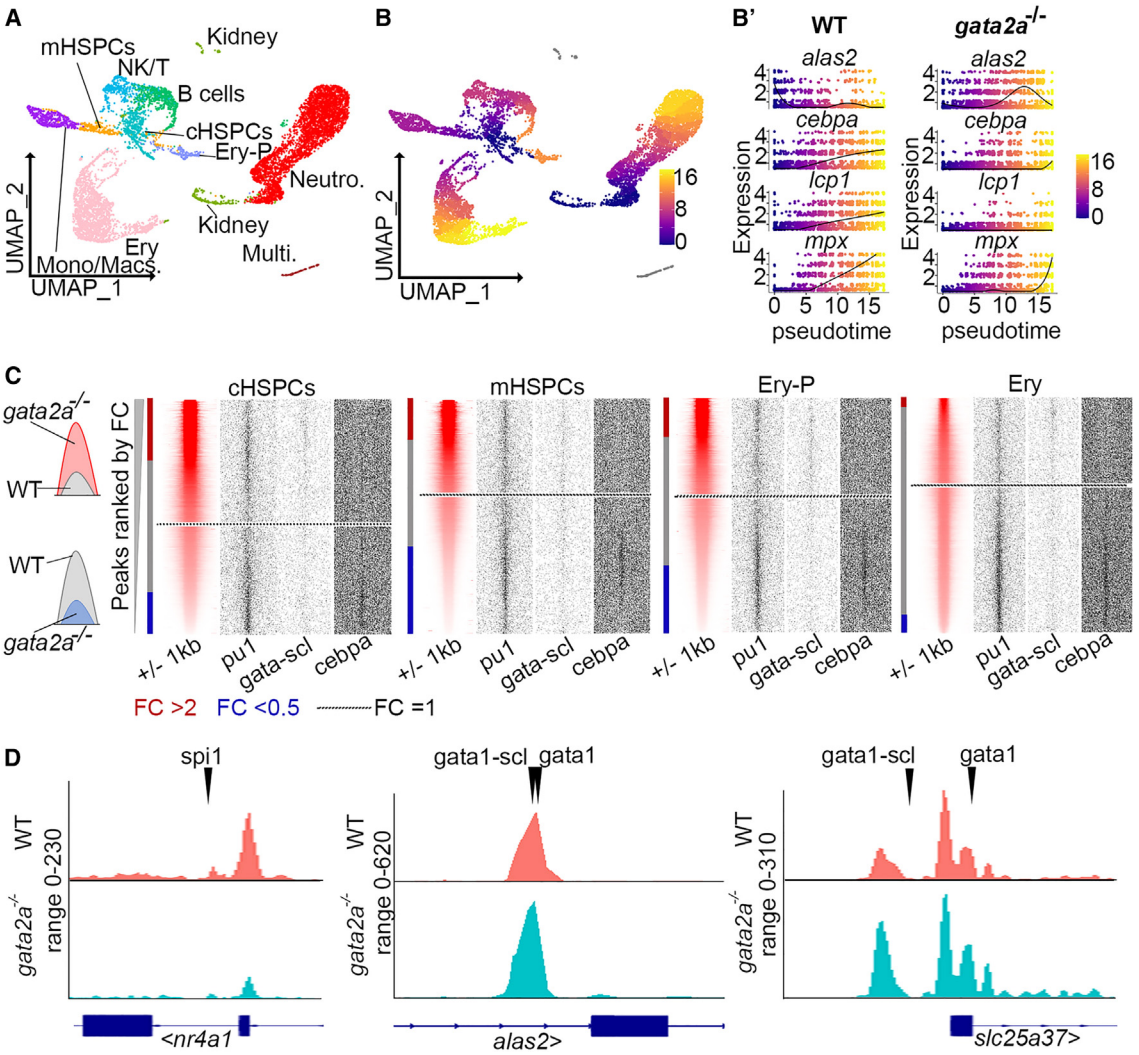


Figure 4. Chromatin accessibility alterations in *gata2a*$^{-/-}$ HSPCs promote erythrocyte differentiation at the expense of the myeloid lineage

(A) Uniform manifold approximation and projection (UMAP) of scRNA-seq data (from Figure 2A) with cell-type labels.

(B) Pseudotime ordering of cells and (B′) expression of *alas2*, *cebpa*, *lcp1*, and *mpx* in each cell with pseudotime progression.

(C) Heatmaps showing peaks ranked by fold change when comparing *gata2a*$^{-/-}$ and WT cells. Red bars depict peaks of FC >2 and blue when FC <0.5. Dotted line indicates FC = 1. FC, fold change; cHSPC, core HSPC; mHSPCs, myeloid HSPCs; Ery-P; erythroid progenitors; Ery, erythrocytes.

(D) Open (FC > 0.2) and closed (FC < −0.2) peaks from erythrocytes were scanned by HOMER for the indicated motifs and sites are indicated by arrowheads. See also Table S6.

corroborate the shift to erythroid at the expense of myeloid gene expression detected in the *gata2a*$^{-/-}$ mutants.

We then examined the scATAC-seq data for motif sites within peaks that are differentially accessible (“open” or “closed” peaks) in *gata2a*$^{-/-}$ mutants compared with WT (Figure 4C). A higher density of *spi1b* and *cebpa* motifs was inaccessible in cHSPCs, mHSPCs, Ery-Ps, and Erys (fold change [FC] < 1). Examination of “closed” peaks (log2FC < −0.2) in *gata2a*$^{-/-}$ erythrocytes for the presence of the *spi1b* motif revealed a match in a peak corresponding to *nr4a1* (Figure 4D), a key transcription factor required for myelopoiesis[37] and neutrophil survival in response to stress.[38] We also observed a higher density of accessible *gata-scl* motifs in this cell population (FC > 1). Scanning “open” peaks (log2FC > 0.2) in *gata2a*$^{-/-}$ erythrocytes for the presence of *gata1* and *gata1-scl* motifs identified matches near *slc25a37* (also known as *mitoferrin*, an erythrocyte solute carrier required for mitochondrial heme synthesis and erythrocyte differentiation)[39] and *alas2* (Figure 4D). Thus, loss of *gata2a* leads to changes in chromatin accessibility in cHSPCs that are found in their differentiated progeny and favor erythroid lineage commitment at the expense of the myeloid lineage, consistent with the erythroid skewing observed in the pseudotime analysis. We therefore concluded that *gata2a* expression in cHSPCs is required to establish a transcriptional network that controls erythroid and myeloid fate upon HSPC differentiation through

regulating gene expression and chromatin accessibility of lineage-specific genes.

The nucleophosmin ortholog *npm1a* is downregulated in mutant HSPCs

We next examined how loss of *gata2a* might drive disease progression in GATA2 deficiency. We observed that expression of the NPM1 ortholog *npm1a* was highly enriched in cHSPCs and mHSPCs (Figures S5A and S5A′) and decreased in *gata2a*$^{-/-}$ mHSPCs at 5 mpf (Figure 3B). Mutations in the C-terminal nuclear localization sequence of nucleophosmin (NPM1) are associated with myeloid and lymphoid malignancies[40] and are found in around 30% of AML cases.[41] Furthermore, NPM1 is associated with maintenance of genome stability and is required for DNA repair.[28,29] To investigate whether loss of *npm1a* might underly disease progression in GATA2 deficiency, we first asked whether its expression was detectable in HSPCs isolated from WKM via expression of low levels of GFP driven by the CD41 promoter (CD41:GFP^{low} HSPCs).[42] Because our experiment was performed using WT and *gata2a*$^{-/-}$ WKM on a *cd41:GFP* transgenic background, we first mapped *gfp* expression and confirmed that it was highly enriched in mHSPCs and cHSPCs (Figures S5A and S5A′). We grouped cHSPCs and mHSPCs into one population and partitioned it into three groups based upon expression levels of GFP: negative, low, and high (Figure S5B). We then identified the global markers associated with each population. GFP^{-} HSPCs displayed enrichment in erythroid markers (*znfl2a*, *mycb*, *hbae1*.3) and *npm1a* (Figure S5B). The GFP^{low} population was highly enriched in HSPC-associated markers, including *gata2b*, *cd34l*, *runx1t1*, and *npm1a*, consistent with previous work identifying HSPC activity in the GFP^{low} fraction by flow cytometry.[42] The GFP^{high} population expressed higher levels of thrombocyte-associated genes including *mpl* and *gp1bb* (Figures S5D and S5E) but lacked *npm1a* and *cd34l*. This confirmed that *npm1a* is expressed in the GFP^{low} HSPC population and allowed us to examine whether loss of *npm1a* in this population correlates with disease progression.

To address this, we dissected WKM of two *cd41:GFP* and two *cd41:GFP;gata2a*$^{-/-}$ adults at 7 mpf and sorted live CD41:GFP^{low} cells (Figure 5A). As expected, fluorescence-activated cell sorting (FACS) analysis showed a decrease in the number of CD41:GFP^{low} cells in the *gata2a*$^{-/-}$ mutants (Figure 5B). We then sorted these cells and profiled gene expression by scRNA-seq. This revealed two distinct clusters made up exclusively of either WT or *gata2a*$^{-/-}$ GFP^{low} HSPCs (Figure 5C). Differential expression analysis showed a strong down regulation of *npm1a* (Figures 5D and 5E) and upregulation of the innate immune response genes *lyz* and *lect2l* (Figures 5D and 5E), a likely consequence of a shift toward myeloid blast accumulation preceding MDS/AML onset. Consistent with the erythroid skewing found in 5 mpf WKM, *hbba2* and *hbaa2* were upregulated (Figures 5D and 5E). Thus, we concluded that disease progression toward marrow failure in *gata2a*$^{-/-}$ mutants correlated with a decrease in *npm1a* expression in HSPCs.

To investigate whether loss of Npm1a function affected HSPC survival, we treated zebrafish embryos from 24 to 48 hpf with the NPM1 inhibitor NSC348884 (NPMi)[43] and assessed expression of the neutrophil marker *mpx* (Figures S5F and S5G). This treatment induced a marked decrease in *mpx* expression in the caudal haematopoietic tissue (CHT) in a dose-dependent manner (Figure S5F), mimicking the loss of *mpx* in *npm1a* morphants[44] and indicating the effectiveness of this drug. Therefore, we next examined the number of HSPCs arriving in the CHT niche by treating Tg(Runx1P2:Citrine) embryos[45] with NPMi from 24 to 48 hpf and imaging the CHT region at 48 hpf. The NPM1i induced a robust decrease in the number of citrine^{+} cells (Figures 5F–5H), suggesting that *npm1a* is required to regulate HSPC cell numbers.

Loss of *npm1a* correlates with elevated levels of DNA damage

Loss of Npm1 function in hypomorphic murine models leads to genomic instability and impaired repair of DNA breaks.[29,46] Thus, we wanted to examine the prevalence of DNA damage in *gata2a*$^{-/-}$ mutant WKM. To achieve this, we performed staining for the DNA damage marker γH2AX[47] in WKM smears at 6 and 12 mpf (Figure 6). WT WKM smears showed no evidence of γH2AX staining at 6 or 12 mpf (Figure 6B). In contrast, higher levels of γH2AX staining were readily detected in *gata2a*$^{-/-}$ WKM from 6 to 12 mpf adults (Figures 6B and S6A). As a positive control, exposure to 10 Gy ionizing radiation induced high levels of nuclear γH2AX staining in freshly isolated WT WKM smears (Figures 6A and 6A′). These data were confirmed by flow cytometry analysis in 12mpf *gata2a*$^{-/-}$ WKM cell suspensions (Figures 6C and S6B), suggesting that loss of *gata2a* leads to elevated levels of DNA damage in hematopoietic cells. To independently confirm these results, we used the murine hematopoietic precursor cell line HPC7[48] to examine γH2AX levels following GATA or NPM1 inhibition for 24 h (Figures 6D and 6E). Treatment with 100 μM of the pan-GATA inhibitor pyrrothiogatain (GATAi)[49] or 1 μM NPMi[43] failed to induce a significant increase in γH2AX levels by immunoblotting (Figures 6D and 6E). However, using a combination of both inhibitors, or a concentration of GATAi that abolishes most of the GATA DNA-binding activity (200 μM; see Nomura et al.[49]), induced a significant increase in γH2AX levels (Figures 6D and 6E). In addition, we did not detect statistically significant changes in GATA2 protein except upon NPM1i treatment (Figure S6C), suggesting that loss of activity, rather than loss of protein, accounted for the increased γH2AX levels. We also confirmed that these treatments did not impact upon cell proliferation rates after 24 h of treatment (Figure S6D), indicating that the increased γH2AX levels are not due to decreased cell viability. Taken together, these data indicate that combined loss of NPM1 and GATA2 activity, or loss of the gata2a i4 enhancer, leads to increased DNA damage in HSPCs.

In summary, we demonstrate that *gata2a* regulates *cebpa* and *npm1a* expression in HSPCs. Loss of *cebpa* results in skewed differentiation, whereas loss of *npm1a* and other factors compromise DNA repair in HSPCs. This results in increased HSPC attrition, likely coupled with increased genomic instability over time in the surviving cells. We propose that this leads to accumulation of somatic mutations, driving disease progression from initial marrow failure toward a second phase of myeloid proliferation in HSPCs, ultimately resulting in the onset of MDS/AML in GATA2 deficiency.

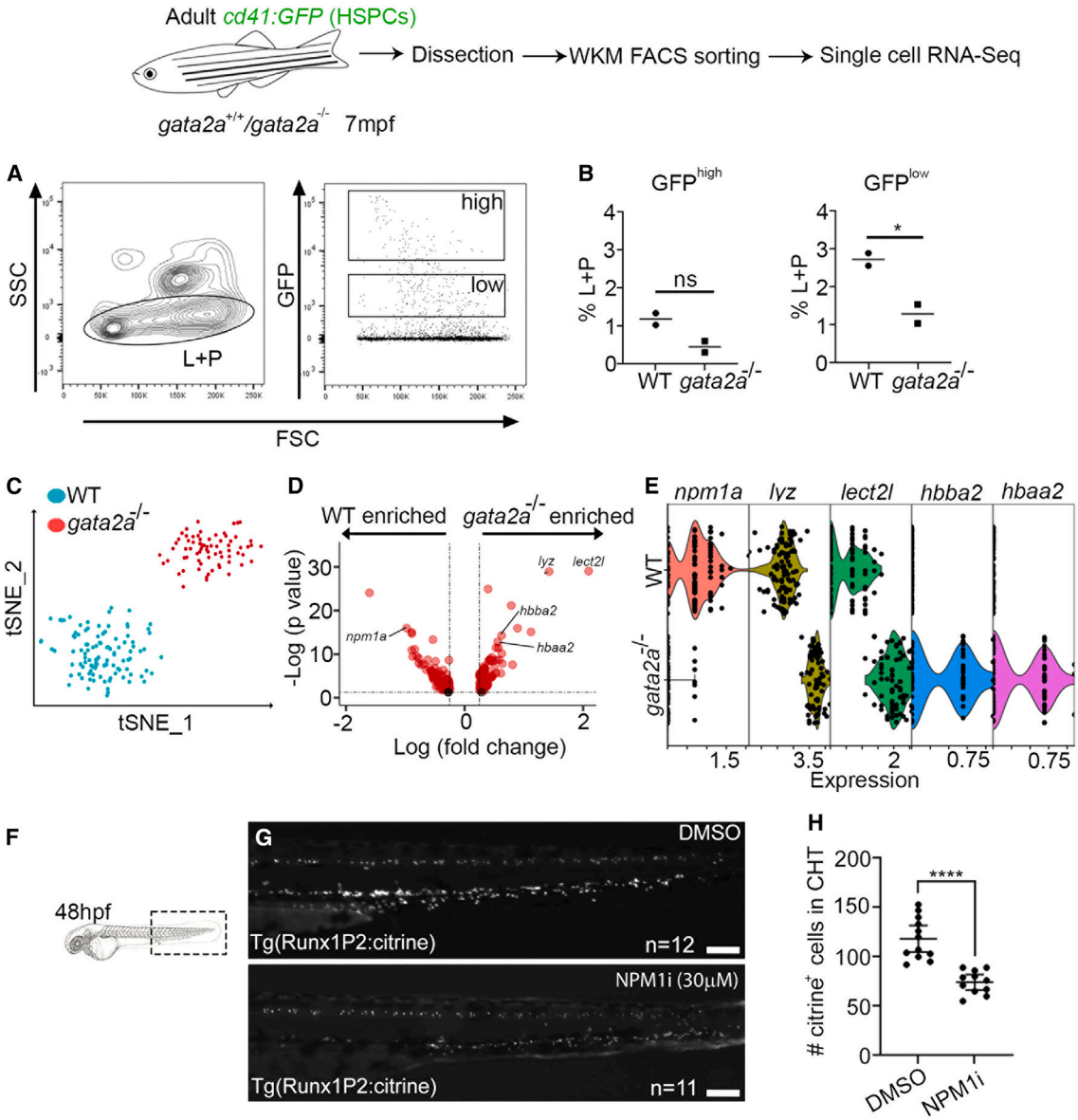


Figure 5. Severe npm1a downregulation in *gata2a*$^{-/-}$ HSPCs at 7 mpf

(A) FACS gating strategy to analyze/isolate *cd41:GFP*^{low} HSPCs from 7 mpf adults. L, lymphoid cells; P, progenitor cells.

(B) Two *cd41:GFP* and two *cd41:GFP;gata2a*$^{-/-}$ adults at 7 mpf and sorted live CD41:GFP^{low} cell analysis of GFP^{high} and GFP^{low} fractions. GFP^{high}: p = 0.0783, GFP^{low}: p = 0.0409, determined using a two-tailed unpaired t test.

(C) scRNA-seq and tSNE clustering of WT and *gata2a*$^{-/-}$ HSPCs. We chose cells with at least 1,000 UMI counts (106 WT and 72 *gata2a*$^{-/-}$ cells) and then exported extracted barcoded reads to be retained for further analysis in cell ranger/Loupe/R.

(D) Volcano plots of log2 FC vs. –log10(p value). Full list of differentially expressed genes in Table S4.

(E) Violin plot (each dot represents a single cell) of *npm1a*, *lyz*, *lect2l*, *hbba2*, and *hbaa2* expression in WT and *gata2a*$^{-/-}$ HSPCs. *gata2a*$^{-/-}$ vs. WT: npm1a; p = 1.06E−16.

(F) Imaging schematic. *lyz*: p = 1.15E−29; *lect2l*: p = 8.76E−30; *hbba2*: p = 5.3E−15; *hbaa2*: p = 1.73E−13.

(G) Imaging of the caudal haematopoietic tissue (CHT) region of Tg(Runx1p2:Citrine)$^{+}$ embryos at 48 hpf. Scale bar, 100 μm.

(H) Quantification. Error bars are mean with 95% confidence interval (CI). DMSO: n = 12 (biological replicates), NPM1i: n = 11 (biological replicates).

****p < 0.0001, determined using a Mann-Whitney test. See also Tables S4, S7, and S8.

DISCUSSION

We have shown that deletion of a conserved *gata2a* enhancer (the i4 enhancer) leads to hypocellularity and marrow failure at 6 mpf. This is preceded by increased expression of erythroid genes in HSPCs, myeloid and lymphoid lineages, and differentiation skewing associated with decreased *cebpa* and *spib* expression at 5 mpf. Furthermore, compromised chromatin

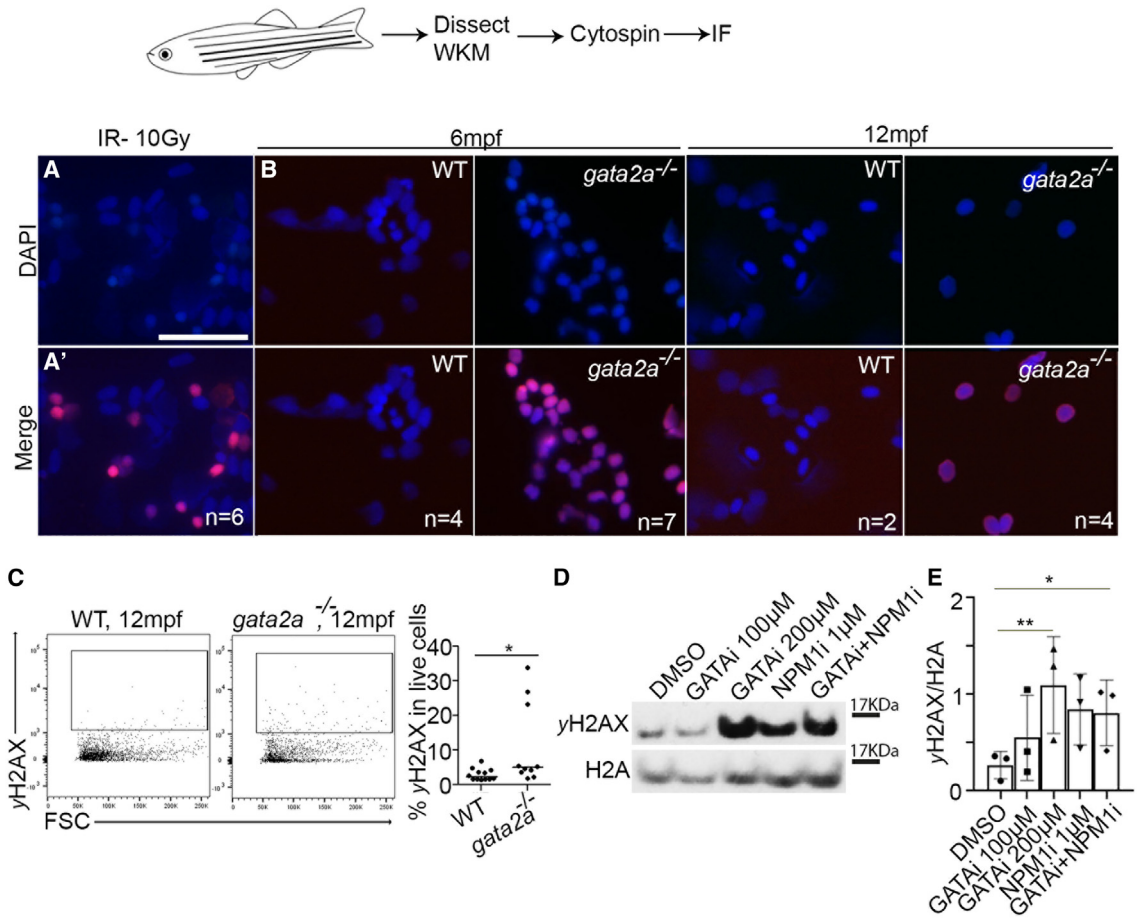


Figure 6. Loss of gata2a leads to increased DNA damage and mutation burden in WKM of *gata2a*$^{-/-}$ MUTs

(A and A′) Control irradiation performed at 10 Gy on WKM cells and immunofluorescence against DAPI and γH2AX. Scale, 25 μm.

(B) WKM from 6 and 12 mpf WT or *gata2a*$^{-/-}$ adults and immunofluorescence against DAPI and γH2AX. Imaging is at 40× magnification.

(C) Flow cytometry analysis of γH2AX expression in 12 mpf WKM. p = 0.0296, WT: n = 12 (biological replicates), *gata2a*$^{-/-}$: n = 10 (biological replicates).

(D) Western blot of γH2AX and H2A (loading control) in HPC7 cells after 1 μM NSC348884 (NPM1 inhibitor[43]) or 100/200 μM pyrrothiogatain (GATA inhibitor[49]) treatment for 24 h.

(E) Quantification of western blot data, n = 3/condition (biological triplicates).

*p < 0.05, determined by using an unpaired, two-tailed t test in (C). **p < 0.01 and *p < 0.05, determined by Friedman test in (E). See also Figure S6.

accessibility and decreased *npm1a* expression in mHSPCs (or loss of NPM1 function) correlate with increased DNA damage in WKM cells and in murine hematopoietic progenitor cells. These transcriptional changes lead to aberrant differentiation and defective DNA repair that ultimately trigger marrow failure. Why the hypocellularity phenotype manifests only between 5 and 6 mpf is unknown. We speculate that the shift in cellularity that we observed reflects a critical accumulation of replication stress in *gata2a*$^{-/-}$ HSPCs leading to functional decline in these cells rather than any specific maturation event. It has been demonstrated that cycling aged HSCs in mouse have increased levels of replication stress, cell cycle defects, and altered replication fork dynamics.[50] This was associated with decreased expression of mini-chromosome maintenance (MCM) components of the prereplication complex. Interestingly, we found decreased expression of two components of this complex, *mcm3* and *mcm5*, in *gata2a*-deficient mHSPCs and in GO term analysis, which showed an enrichment for DNA replication factors in this population (Figure 3). Therefore, we hypothesize that normal aging of HSPCs is accelerated in the absence of *gata2a*, leading to HSC attrition and subsequent decreased numbers of functional HSPCs by 5 to 6 mpf. Further experimental analysis investigating the loss of MCM components in the context of *gata2a* loss of function are required to investigate this intriguing possibility.

A gata2a/cebpa axis regulates marrow cellularity and myeloid lineage output

In our single-cell analyses, we find that chromatin remodeling in cHSPCs precedes a decrease in *cebpa* expression in *gata2a*$^{-/-}$ mHSPCs. Previous analyses of CEBPA function indicate that it acts as an HSC priming factor to promote myeloid differentiation and protects HSCs against apoptosis and maintains quiescence.[22,24] We therefore hypothesize that *gata2a* establishes a

transcriptional network in cHSPCs by indirectly regulating chromatin accessibility, possibly at a downstream enhancer that is required for the expression of *cebpa* in mHSPCs. The corresponding (+37 kb) murine enhancer is bound by GATA2,[35] and its cell-autonomous deletion results in neutropenia, marrow hypocellularity, and reduced long-term HSC (LT-HSC) numbers.[51,52] We conclude that as cHSPCs begin to differentiate to mHSPCs, *gata2a*$^{-/-}$ mutants lack the cues that maintain *cebpa* expression at normal levels, therefore skewing differentiation down the erythroid lineage at the expense of myeloid lineages.[22] This is illustrated by our pseudotime analysis of RNA expression and motif analysis of accessible chromatin where erythrocyte-specific motifs are enriched in open peaks and myeloid-specific motifs are associated with closed peaks in HSPCs. This is consistent with a previous report demonstrating that GATA2 and CEBPA mutations in myeloid hematopoietic progenitors cooperate to increase access to chromatin by erythroid transcription factors.[53]

Overlapping and distinct roles for Gata2a in HSPCs

Analysis of mutants for the *gata2a* zebrafish ortholog *gata2b* found that Gata2b also favored myeloid over lymphoid/erythroid lineage commitment in adult HSPCs.[54] An accompanying study showed decreased myeloid and increased lymphoid priming of hematopoietic progenitors by scATAC-seq in a similar *gata2b* mutant.[55] As both *gata2a* and *gata2b* are expressed in HSPCs and as loss of either gene leads to a defect in myelopoiesis/lymphopoiesis, we propose that they have overlapping roles in HSPC biology and might function as "allelic paralogs" whose deletion would recapitulate the lineage defects found in patients with germline heterozygous GATA2 deficiency. However, our *gata2a* enhancer deletion mutant[13] recapitulates a wider range of GATA2 deficiency syndrome symptoms (recurrent infections, lymphedema, cytopenia, marrow failure, and predisposition to MDS/AML), and thus more faithfully recapitulates GATA2 deficiency syndrome. In addition, both heterozygous and homozygous *gata2a* mutants show marrow hypocellularity (Figure 1) and decreased numbers of HSPCs (Figures 2 and 5), phenotypes that are absent from *gata2b* mutants.[54] Taken together, this suggests that Gata2a plays a role in maintaining marrow cellularity and HSPC survival that is not shared with Gata2b.

Genome instability and impaired DNA damage underlie disease progression in GATA2 deficiency

Appearance of MDS/AML in patients carrying germline GATA2 mutations is associated with secondary mutations in several genes.[10,56–58] While these studies provide an association between the genotype (mutated genes) and the phenotype (leukemia), they do not address how the identified mutations arise or how they contribute to disease progression. Here, we report that loss of *gata2a* leads to decreased *npm1a* expression, associated with elevated levels of DNA damage, which likely contributes to increased mutation burden in the mutant WKM. Although it remains unclear whether the regulation of *npm1a* by *gata2a* is direct, previous studies have shown that *NPM1* maintains genome integrity during cell cycle and inhibits apoptosis/tumor suppression through p53 and that its loss of function is associated with myeloid and lymphoid malignancies.[40,59,60] In addition, conditional deletion of Npm1 in murine HSCs leads to marrow failure and MDS that is exacerbated by p53 loss.[61] We therefore propose that progressive loss of *npm1a* (and that of other downregulated DNA repair genes such as the HELLS chromatin remodeler) in *gata2a*$^{-/-}$ HSPCs leads to impaired DNA damage repair. Over time, this leads to increased DNA damage and concomitant genomic instability that contributes to the acquisition of secondary mutations in cancer-associated genes. This is further corroborated by recent studies showing that cell-autonomous telomere shortening contributed to genome instability and increased tumor incidence.[62] Once HSPC clones acquire sufficient combinations of pathogenic mutations, they are conferred a competitive advantage that permits clonal expansion and eventually triggers MDS/AML in patients with GATA2 deficiency.

Our zebrafish *gata2a*$^{-/-}$ mutant therefore represents a tractable genetic model that will enable further investigations into how these acquired secondary mutations drive marrow failure and further disease progression to overt MDS/AML in GATA2 deficiency.

Limitations of the study

Our single-cell data allowed us to analyze differential gene expression in HSPCs (Figures 3A, 3B, and S3C) to find key genes regulated by Gata2a and skewing of gene expression. This analysis was limited by the number of HSPCs present in each whole-marrow sample. Therefore, it is likely that enriching for HSPCs might reveal additional information on differentially expressed genes at higher sensitivity.

Our data point toward a deficit in DNA repair and replication as an underlying cause for the marrow failure phenotype observed in the GATA2 deficiency model. We hypothesize that this deficit will result in genome instability in hematopoietic cells. Although preliminary evidence suggests that mutational burden is increased in mutant hematopoietic cells, a robust whole-genome sequencing (WGS) experiment comparing mutational burden from sibling WT or *gata2a*$^{-/-}$ mutant HSPCs will be required to gauge the extent of that instability and its development over time.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Zebrafish
  - Cell lines
- METHOD DETAILS
  - May-Grünwald (MG) and Giemsa staining of WKM smears
  - Flow cytometry and cell sorting

- Single cell RNA sequencing (scRNAseq)
- Single cell assay for transposable chromatin followed by sequencing (scATAC-seq)
- Immunofluorescence
- Western blot
- Cell proliferation assay
- Whole-mount *in situ* hybridisation
- CRISPR/Cas9 F0 injections and mRNA injections
- Live imaging

● QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.112571.

ACKNOWLEDGMENTS

We thank Constanze Bonifer and Jon Frampton for critical reading of the manuscript. We are grateful to the staff of the Biomedical Services Unit for fish husbandry and the staff at the flow cytometry facility, Dr. Adriana Flores-Langarica, Dr. Paola Pietroni, and Dr. Guillaume Desanti for cell sorting. We thank the staff of Genomics Birmingham (University of Birmingham) for genomic sequencing and Jon Frampton for the HPC7 cells. This research was supported by a Wellcome Trust ISSF Award to R.M. and a British Heart Foundation Fellowship to R.M. and C.B.M. (BHF IBSR FS/13/50/30436). C.B.M. was also supported by the Research and Development Fund (University of Birmingham). P.V. was supported by a Wellcome Trust Investigator Awards to S.O. (grant/award number: 212233/Z/18/Z). B.N. and Y.P. were supported through the Cancer Research UK Birmingham Centre award C17422/A25154. L.C. was funded by a PhD scholarship from the University of Birmingham and Cancer Research UK (C17422/A25154). M.R.H. was funded by an MRC Career Development Fellowship (MR/P009085/1).

AUTHOR CONTRIBUTIONS

Conceptualization, C.B.M., M.R.H., and R.M.; methodology, C.B.M. and R.M.; formal analysis, C.B.M., L.C., P.V., B.N., C.C., Y.P., and C.P.; investigation, C.B.M., L.C., and S.B.; resources, R.M., S.O., and C.P.; writing – original draft, C.B. and R.M.; writing – review & editing, C.B., S.O., M.R.H., and R.M.; supervision, S.O., M.R.H., and R.M.; funding acquisition, S.O. and R.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: April 14, 2022
Revised: March 14, 2023
Accepted: May 12, 2023

REFERENCES

1. Sawai, C.M., Babovic, S., Upadhaya, S., Knapp, D.J.H.F., Lavin, Y., Lau, C.M., Goloborodko, A., Feng, J., Fujisaki, J., Ding, L., et al. (2016). Hematopoietic stem cells are the major source of multilineage hematopoiesis in adult animals. Immunity *45*, 597–609. https://doi.org/10.1016/j.immuni.2016.08.007.
2. Crane, G.M., Jeffery, E., and Morrison, S.J. (2017). Adult haematopoietic stem cell niches. Nat. Rev. Immunol. *17*, 573–590. https://doi.org/10.1038/nri.2017.53.
3. de Pater, E., Kaimakis, P., Vink, C.S., Yokomizo, T., Yamada-Inagawa, T., van der Linden, R., Kartalaei, P.S., Camper, S.A., Speck, N., and Dzierzak, E. (2013). Gata2 is required for HSC generation and survival. J. Exp. Med. *210*, 2843–2850. https://doi.org/10.1084/jem.20130751.
4. Gao, X., Johnson, K.D., Chang, Y.I., Boyer, M.E., Dewey, C.N., Zhang, J., and Bresnick, E.H. (2013). Gata2 cis-element is required for hematopoietic stem cell generation in the mammalian embryo. J. Exp. Med. *210*, 2833–2842. https://doi.org/10.1084/jem.20130733.
5. Rodrigues, N.P., Boyd, A.S., Fugazza, C., May, G.E., Guo, Y., Tipping, A.J., Scadden, D.T., Vyas, P., and Enver, T. (2008). GATA-2 regulates granulocyte-macrophage progenitor cell function. Blood *112*, 4862–4873. https://doi.org/10.1182/blood-2008-01-136564.
6. Spinner, M.A., Sanchez, L.A., Hsu, A.P., Shaw, P.A., Zerbe, C.S., Calvo, K.R., Arthur, D.C., Gu, W., Gould, C.M., Brewer, C.C., et al. (2014). GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics, and immunity. Blood *123*, 809–821. https://doi.org/10.1182/blood-2013-07-515528.
7. Dickinson, R.E., Milne, P., Jardine, L., Zandi, S., Swierczek, S.I., McGovern, N., Cookson, S., Ferozepurwalla, Z., Langridge, A., Pagan, S., et al. (2014). The evolution of cellular deficiency in GATA2 mutation. Blood *123*, 863–874. https://doi.org/10.1182/blood-2013-07-517151.
8. McReynolds, L.J., Calvo, K.R., and Holland, S.M. (2018). Germline GATA2 mutation and bone marrow failure. Hematol. Oncol. Clin. North Am. *32*, 713–728. https://doi.org/10.1016/j.hoc.2018.04.004.
9. Wlodarski, M.W., Collin, M., and Horwitz, M.S. (2017). GATA2 deficiency and related myeloid neoplasms. Semin. Hematol. *54*, 81–86. https://doi.org/10.1053/j.seminhematol.2017.05.002.
10. Hirabayashi, S., Wlodarski, M.W., Kozyra, E., and Niemeyer, C.M. (2017). Heterogeneity of GATA2-related myeloid neoplasms. Int. J. Hematol. *106*, 175–182. https://doi.org/10.1007/s12185-017-2285-2.
11. Johnson, K.D., Hsu, A.P., Ryu, M.J., Wang, J., Gao, X., Boyer, M.E., Liu, Y., Lee, Y., Calvo, K.R., Keles, S., et al. (2012). Cis-element mutated in GATA2-dependent immunodeficiency governs hematopoiesis and vascular integrity. J. Clin. Invest. *122*, 3692–3704. https://doi.org/10.1172/JCI61623.
12. Soukup, A.A., Zheng, Y., Mehta, C., Wu, J., Liu, P., Cao, M., Hofmann, I., Zhou, Y., Zhang, J., Johnson, K.D., et al. (2019). Single-nucleotide human disease mutation inactivates a blood-regenerative GATA2 enhancer. J. Clin. Invest. *129*, 1180–1192. https://doi.org/10.1172/jci122694.
13. Dobrzycki, T., Mahony, C.B., Krecsmarik, M., Koyunlar, C., Rispoli, R., Peulen-Zink, J., Gussinklo, K., Fedlaoui, B., de Pater, E., Patient, R., and Monteiro, R. (2020). Deletion of a conserved Gata2 enhancer impairs haemogenic endothelium programming and adult Zebrafish haematopoiesis. Commun. Biol. *3*, 71. https://doi.org/10.1038/s42003-020-0798-3.
14. Traver, D., Paw, B.H., Poss, K.D., Penberthy, W.T., Lin, S., and Zon, L.I. (2003). Transplantation and in vivo imaging of multilineage engraftment in zebrafish bloodless mutants. Nat. Immunol. *4*, 1238–1246. https://doi.org/10.1038/ni1007.
15. Renshaw, S.A., Loynes, C.A., Trushell, D.M.I., Elworthy, S., Ingham, P.W., and Whyte, M.K.B. (2006). A transgenic zebrafish model of neutrophilic inflammation. Blood *108*, 3976–3978. https://doi.org/10.1182/blood-2006-05-024075.
16. Ferrero, G., Gomez, E., Lyer, S., Rovira, M., Miserocchi, M., Langenau, D.M., Bertrand, J.Y., and Wittamer, V. (2020). The macrophage-expressed gene (mpeg) 1 identifies a subpopulation of B cells in the adult zebrafish. J. Leukoc. Biol. *107*, 431–443. https://doi.org/10.1002/jlb.1a1119-223r.
17. Yaqoob, N., Holotta, M., Prem, C., Kopp, R., and Schwerte, T. (2009). Ontogenetic development of erythropoiesis can be studied non-invasively in GATA-1:DsRed transgenic zebrafish. Comp. Biochem. Physiol. Mol. Integr. Physiol. *154*, 270–278. https://doi.org/10.1016/j.cbpa.2009.06.024.
18. Moore, F.E., Garcia, E.G., Lobbardi, R., Jain, E., Tang, Q., Moore, J.C., Cortes, M., Molodtsov, A., Kasheta, M., Luo, C.C., et al. (2016). Single-cell transcriptional analysis of normal, aberrant, and malignant

hematopoiesis in zebrafish. J. Exp. Med. *213*, 979–992. https://doi.org/10.1084/jem.20152013.

19. Tang, Q., Iyer, S., Lobbardi, R., Moore, J.C., Chen, H., Lareau, C., Hebert, C., Shaw, M.L., Neftel, C., Suva, M.L., et al. (2017). Dissecting hematopoietic and renal cell heterogeneity in adult zebrafish at single-cell resolution using RNA sequencing. J. Exp. Med. *214*, 2875–2887. https://doi.org/10.1084/jem.20170976.
20. Peters, M.J., Parker, S.K., Grim, J., Allard, C.A.H., Levin, J., and Detrich, H.W., 3rd. (2018). Divergent Hemogen genes of teleosts and mammals share conserved roles in erythropoiesis: analysis using transgenic and mutant zebrafish. Biol. Open *7*, bio035576. https://doi.org/10.1242/bio.035576.
21. Dai, Y., Zhu, L., Huang, Z., Zhou, M., Jin, W., Liu, W., Xu, M., Yu, T., Zhang, Y., Wen, Z., et al. (2016). Cebpα is essential for the embryonic myeloid progenitor and neutrophil maintenance in zebrafish. J Genet Genomics *43*, 593–600. https://doi.org/10.1016/j.jgg.2016.09.001.
22. Suh, H.C., Gooya, J., Renn, K., Friedman, A.D., Johnson, P.F., and Keller, J.R. (2006). C/EBPalpha determines hematopoietic cell fate in multipotential progenitor cells by inhibiting erythroid differentiation and inducing myeloid differentiation. Blood *107*, 4308–4316. https://doi.org/10.1182/blood-2005-06-2216.
23. Avellino, R., and Delwel, R. (2017). Expression and regulation of C/EBPα in normal myelopoiesis and in malignant transformation. Blood *129*, 2083–2091. https://doi.org/10.1182/blood-2016-09-687822.
24. Hasemann, M.S., Lauridsen, F.K.B., Waage, J., Jakobsen, J.S., Frank, A.-K., Schuster, M.B., Rapin, N., Bagger, F.O., Hoppe, P.S., Schroeder, T., and Porse, B.T. (2014). C/EBPα is required for long-term self-renewal and lineage priming of hematopoietic stem cells and for the maintenance of epigenetic configurations in multipotent progenitors. PLoS Genet. *10*, e1004079. https://doi.org/10.1371/journal.pgen.1004079.
25. Shinya, M., Machiki, D., Henrich, T., Kubota, Y., Takisawa, H., and Mimura, S. (2014). Evolutionary diversification of MCM3 genes in Xenopus laevis and Danio rerio. Cell Cycle *13*, 3271–3281. https://doi.org/10.4161/15384101.2014.954445.
26. Snyder, M., He, W., and Zhang, J.J. (2005). The DNA replication factor MCM5 is essential for Stat1-mediated transcriptional activation. Proc. Natl. Acad. Sci. USA *102*, 14539–14544. https://doi.org/10.1073/pnas.0507479102.
27. Kollárovič, G., Topping, C.E., Shaw, E.P., and Chambers, A.L. (2020). The human HELLS chromatin remodelling protein promotes end resection to facilitate homologous recombination and contributes to DSB repair within heterochromatin. Nucleic Acids Res. *48*, 1872–1885. https://doi.org/10.1093/nar/gkz1146.
28. Koike, A., Nishikawa, H., Wu, W., Okada, Y., Venkitaraman, A.R., and Ohta, T. (2010). Recruitment of phosphorylated NPM1 to sites of DNA damage through RNF8-dependent ubiquitin conjugates. Cancer Res. *70*, 6746–6756. https://doi.org/10.1158/0008-5472.Can-10-0382.
29. Grisendi, S., Bernardi, R., Rossi, M., Cheng, K., Khandker, L., Manova, K., and Pandolfi, P.P. (2005). Role of nucleophosmin in embryonic development and tumorigenesis. Nature *437*, 147–153. https://doi.org/10.1038/nature03915.
30. Kawamura, K., Qi, F., Meng, Q., Hayashi, I., and Kobayashi, J. (2019). Nucleolar protein nucleolin functions in replication stress-induced DNA damage responses. J. Radiat. Res. *60*, 281–288. https://doi.org/10.1093/jrr/rry114.
31. Scott, D.D., and Oeffinger, M. (2016). Nucleolin and nucleophosmin: nucleolar proteins with multiple functions in DNA repair. Biochem. Cell. Biol. *94*, 419–432. https://doi.org/10.1139/bcb-2016-0068.
32. Wu, Z., Gao, S., Diamond, C., Kajigaya, S., Chen, J., Shi, R., Palmer, C., Hsu, A.P., Calvo, K.R., Hickstein, D.D., et al. (2020). Sequencing of RNA in single cells reveals a distinct transcriptome signature of hematopoiesis in GATA2 deficiency. Blood Adv. *4*, 2656–2670. https://doi.org/10.1182/bloodadvances.2019001352.
33. Org, T., Duan, D., Ferrari, R., Montel-Hagen, A., Van Handel, B., Kerényi, M.A., Sasidharan, R., Rubbi, L., Fujiwara, Y., Pellegrini, M., et al. (2015). Scl binds to primed enhancers in mesoderm to regulate hematopoietic and cardiac fate divergence. EMBO J. *34*, 759–777. https://doi.org/10.15252/embj.201490542.
34. Sánchez-Castillo, M., Ruau, D., Wilkinson, A.C., Ng, F.S., Hannah, R., Diamanti, E., Lombard, P., Wilson, N.K., and Gottgens, B. (2014). CODEX: a next-generation sequencing experiment database for the haematopoietic and embryonic stem cell communities. Nucleic Acids Res. *43*, D1117–D1123. https://doi.org/10.1093/nar/gku895.
35. Cooper, S., Guo, H., and Friedman, A.D. (2015). The +37 kb cebpa enhancer is critical for cebpa myeloid gene expression and contains functional sites that bind SCL, GATA2, C/EBPα, PU.1, and additional ets factors. PLoS One *10*, e0126385. https://doi.org/10.1371/journal.pone.0126385.
36. Peña, O.A., Lubin, A., Rowell, J., Hoade, Y., Khokhar, N., Lemmik, H., Mahony, C., Dace, P., Umamahesan, C., and Payne, E.M. (2021). Differential requirement of Gata2a and Gata2b for primitive and definitive myeloid development in zebrafish. Front. Cell Dev. Biol. *9*, 708113. https://doi.org/10.3389/fcell.2021.708113.
37. Tacke, R., Hilgendorf, I., Garner, H., Waterborg, C., Park, K., Nowyhed, H., Hanna, R.N., Wu, R., Swirski, F.K., Geissmann, F., and Hedrick, C.C. (2015). The transcription factor NR4A1 is essential for the development of a novel macrophage subset in the thymus. Sci. Rep. *5*, 10055. https://doi.org/10.1038/srep10055.
38. Prince, L.R., Prosseda, S.D., Higgins, K., Carlring, J., Prestwich, E.C., Ogryzko, N.V., Rahman, A., Basran, A., Falciani, F., Taylor, P., et al. (2017). NR4A orphan nuclear receptor family members, NR4A2 and NR4A3, regulate neutrophil number and survival. Blood *130*, 1014–1025. https://doi.org/10.1182/blood-2017-03-770164.
39. Chen, W., Paradkar, P.N., Li, L., Pierce, E.L., Langer, N.B., Takahashi-Makise, N., Hyde, B.B., Shirihai, O.S., Ward, D.M., Kaplan, J., and Paw, B.H. (2009). Abcb10 physically interacts with mitoferrin-1 (Slc25a37) to enhance its stability and function in the erythroid mitochondria. Proc. Natl. Acad. Sci. USA *106*, 16263–16268. https://doi.org/10.1073/pnas.0904519106.
40. Sportoletti, P., Grisendi, S., Majid, S.M., Cheng, K., Clohessy, J.G., Viale, A., Teruya-Feldstein, J., and Pandolfi, P.P. (2008). Npm1 is a haploinsufficient suppressor of myeloid and lymphoid malignancies in the mouse. Blood *111*, 3859–3862. https://doi.org/10.1182/blood-2007-06-098251.
41. Falini, B., Mecucci, C., Tiacci, E., Alcalay, M., Rosati, R., Pasqualucci, L., La Starza, R., Diverio, D., Colombo, E., Santucci, A., et al. (2005). Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. N. Engl. J. Med. *352*, 254–266. https://doi.org/10.1056/NEJMoa041974.
42. Ma, D., Zhang, J., Lin, H.-f., Italiano, J., and Handin, R.I. (2011). The identification and characterization of zebrafish hematopoietic stem cells. Blood *118*, 289–297. https://doi.org/10.1182/blood-2010-12-327403.
43. Qi, W., Shakalya, K., Stejskal, A., Goldman, A., Beeck, S., Cooke, L., and Mahadevan, D. (2008). NSC348884, a nucleophosmin inhibitor disrupts oligomer formation and induces apoptosis in human cancer cells. Oncogene *27*, 4210–4220. https://doi.org/10.1038/onc.2008.54.
44. Bolli, N., Payne, E.M., Grabher, C., Lee, J.-S., Johnston, A.B., Falini, B., Kanki, J.P., and Look, A.T. (2010). Expression of the cytoplasmic NPM1 mutant (NPMc+) causes the expansion of hematopoietic cells in zebrafish. Blood *115*, 3329–3340. https://doi.org/10.1182/blood-2009-02-207225.
45. Bonkhofer, F., Rispoli, R., Pinheiro, P., Krecsmarik, M., Schneider-Swales, J., Tsang, I.H.C., de Bruijn, M., Monteiro, R., Peterkin, T., and Patient, R. (2019). Blood stem cell-forming haemogenic endothelium in zebrafish derives from arterial endothelium. Nat. Commun. *10*, 3577. https://doi.org/10.1038/s41467-019-11423-2.
46. Andrade, N.S., Ramic, M., Esanov, R., Liu, W., Rybin, M.J., Gaidosh, G., Abdallah, A., Del'Olio, S., Huff, T.C., Chee, N.T., et al. (2020). Dipeptide repeat proteins inhibit homology-directed DNA double strand break repair

in C9ORF72 ALS/FTD. Mol. Neurodegener. *15*, 13. https://doi.org/10.1186/s13024-020-00365-9.

47. Sharma, A., Singh, K., and Almasan, A. (2012). Histone H2AX phosphorylation: a marker for DNA damage. Methods Mol. Biol. *920*, 613–626. https://doi.org/10.1007/978-1-61779-998-3_40.
48. Pinto do O, P., Kolterud, A., and Carlsson, L. (1998). Expression of the LIM-homeobox gene LH2 generates immortalized steel factor-dependent multipotent hematopoietic precursors. EMBO J. *17*, 5744–5756. https://doi.org/10.1093/emboj/17.19.5744.
49. Nomura, S., Takahashi, H., Suzuki, J., Kuwahara, M., Yamashita, M., and Sawasaki, T. (2019). Pyrrotricha acts as an inhibitor of GATA family proteins and inhibits Th2 cell differentiation in vitro. Sci. Rep. *9*, 17335. https://doi.org/10.1038/s41598-019-53856-1.
50. Flach, J., Bakker, S.T., Mohrin, M., Conroy, P.C., Pietras, E.M., Reynaud, D., Alvarez, S., Diolaiti, M.E., Ugarte, F., Forsberg, E.C., et al. (2014). Replication stress is a potent driver of functional decline in ageing haematopoietic stem cells. Nature *512*, 198–202. https://doi.org/10.1038/nature13619.
51. Avellino, R., Havermans, M., Erpelinck, C., Sanders, M.A., Hoogenboezem, R., Van De Werken, H.J.G., Rombouts, E., Van Lom, K., Van Strien, P.M.H., Gebhard, C., et al. (2016). An autonomous CEBPA enhancer specific for myeloid-lineage priming and neutrophilic differentiation. Blood *127*, 2991–3003. https://doi.org/10.1182/blood-2016-01-695759.
52. Avellino, R., Mulet-Lazaro, R., Havermans, M., Hoogenboezem, R., Smeenk, L., Salomonis, N., Schneider, R.K., Rombouts, E., Bindels, E., Grimes, L., and Delwel, R. (2022). Induced cell-autonomous neutropenia systemically perturbs hematopoiesis in *Cebpa* enhancer-null mice. Blood Adv. *6*, 1406–1419. https://doi.org/10.1182/bloodadvances.2021005851.
53. Di Genua, C., Valletta, S., Buono, M., Stoilova, B., Sweeney, C., Rodriguez-Meira, A., Grover, A., Drissen, R., Meng, Y., Beveridge, R., et al. (2020). C/EBPalpha and GATA-2 mutations induce bilineage acute erythroid leukemia through transformation of a neomorphic neutrophil-erythroid progenitor. Cancer Cell *37*, 690–704.e8. https://doi.org/10.1016/j.ccell.2020.03.022.
54. Gioacchino, E., Koyunlar, C., Zink, J., de Looper, H., de Jong, M., Dobrzycki, T., Mahony, C.B., Hoogenboezem, R., Bosch, D., van Strien, P.M.H., et al. (2021). Essential role for Gata2 in modulating lineage output from hematopoietic stem cells in zebrafish. Blood Adv. *5*, 2687–2700. https://doi.org/10.1182/bloodadvances.2020002993.
55. Avagyan, S., Weber, M.C., Ma, S., Prasad, M., Mannherz, W.P., Yang, S., Buenrostro, J.D., and Zon, L.I. (2021). Single-cell ATAC-seq reveals GATA2-dependent priming defect in myeloid and a maturation bottleneck in lymphoid lineages. Blood Adv. *5*, 2673–2686. https://doi.org/10.1182/bloodadvances.2020002992.
56. Wang, X., Muramatsu, H., Okuno, Y., Sakaguchi, H., Yoshida, K., Kawashima, N., Xu, Y., Shiraishi, Y., Chiba, K., Tanaka, H., et al. (2015). GATA2 and secondary mutations in familial myelodysplastic syndromes and pediatric myeloid malignancies. Haematologica *100*, e398–e401. https://doi.org/10.3324/haematol.2015.127092.
57. McReynolds, L.J., Yang, Y., Yuen Wong, H., Tang, J., Zhang, Y., Mulé, M.P., Daub, J., Palmer, C., Foruraghi, L., Liu, Q., et al. (2019). MDS-associated mutations in germline GATA2 mutated patients with hematologic manifestations. Leuk. Res. *76*, 70–75. https://doi.org/10.1016/j.leukres.2018.11.013.
58. West, R.R., Hsu, A.P., Holland, S.M., Cuellar-Rodriguez, J., and Hickstein, D.D. (2014). Acquired ASXL1 mutations are common in patients with inherited GATA2 mutations and correlate with myeloid transformation. Haematologica *99*, 276–281. https://doi.org/10.3324/haematol.2013.090217.
59. Okuwaki, M. (2008). The structure and functions of NPM1/Nucleophsmin/B23, a multifunctional nucleolar acidic protein. J. Biochem. *143*, 441–448. https://doi.org/10.1093/jb/mvm222.
60. Colombo, E., Marine, J.C., Danovi, D., Falini, B., and Pelicci, P.G. (2002). Nucleophosmin regulates the stability and transcriptional activity of p53. Nat. Cell Biol. *4*, 529–533. https://doi.org/10.1038/ncb814.
61. Morganti, C., Ito, K., Yanase, C., Verma, A., Teruya-Feldstein, J., and Ito, K. (2022). NPM1 ablation induces HSC aging and inflammation to develop myelodysplastic syndrome exacerbated by p53 loss. EMBO Rep. *23*, e54262. https://doi.org/10.15252/embr.202154262.
62. Lex, K., Maia Gil, M., Lopes-Bastos, B., Figueira, M., Marzullo, M., Giannetti, K., Carvalho, T., and Ferreira, M.G. (2020). Telomere shortening produces an inflammatory environment that increases tumor incidence in zebrafish. Proc. Natl. Acad. Sci. USA *117*, 15066–15074. https://doi.org/10.1073/pnas.1920049117.
63. Lin, H.F., Traver, D., Zhu, H., Dooley, K., Paw, B.H., Zon, L.I., and Handin, R.I. (2005). Analysis of thrombocyte development in CD41-GFP transgenic zebrafish. Blood *106*, 3803–3810. https://doi.org/10.1182/blood-2005-01-0179.
64. Traver, D., Paw, B., Poss, K., et al. (2023). Transplantation and in vivo imaging of multilineage engraftment in zebrafish bloodless mutants. Nat. Immunol. *4*, 1238–1246. https://doi.org/10.1038/ni1007.
65. Ellett, F., Pase, L., Hayman, J.W., Andrianopoulos, A., and Lieschke, G.J. (2011). mpeg1 promoter transgenes direct macrophage-lineage expression in zebrafish. Blood *117*, e49–e56. https://doi.org/10.1182/blood-2010-10-314120.
66. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420. https://doi.org/10.1038/nbt.4096.
67. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502. https://doi.org/10.1038/nbt.3192.
68. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., III, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell *177*, 1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.
69. Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat. Commun. *10*, 1523. https://doi.org/10.1038/s41467-019-09234-6.
70. Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496–502. https://doi.org/10.1038/s41586-019-0969-x.
71. Levine, J.H., Simonds, E.F., Bendall, S.C., Davis, K.L., Amir, E.a.D., Tadmor, M.D., Litvin, O., Fienberg, H.G., Jager, A., Zunder, E.R., et al. (2015). Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. Cell *162*, 184–197. https://doi.org/10.1016/j.cell.2015.05.047.
72. Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.-A., and Trapnell, C. (2017). Single-cell mRNA quantification and differential analysis with Census. Nat. Methods *14*, 309–315. https://doi.org/10.1038/nmeth.4150.
73. Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. *9*, 5233. https://doi.org/10.1038/s41598-019-41695-z.
74. Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386. https://doi.org/10.1038/nbt.2859.
75. Stuart, T., Srivastava, A., Madad, S., Lareau, C.A., and Satija, R. (2021). Single-cell chromatin state analysis with Signac. Nat. Methods *18*, 1333–1341. https://doi.org/10.1038/s41592-021-01282-5.
76. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

77. Saldanha, A.J. (2004). Java Treeview—extensible visualization of microarray data. Bioinformatics *20*, 3246–3248. https://doi.org/10.1093/bioinformatics/bth349.

78. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682. https://doi.org/10.1038/nmeth.2019.

79. Kalev-Zylinska, M.L., Horsfield, J.A., Flores, M.V.C., Postlethwait, J.H., Vitas, M.R., Baas, A.M., Crosier, P.S., and Crosier, K.E. (2002). Runx1 is required for zebrafish blood and vessel development and expression of a human RUNX1-CBF2T1 transgene advances a model for studies of leukemogenesis. Development *129*, 2015–2030.

80. Bennett, C.M., Kanki, J.P., Rhodes, J., Liu, T.X., Paw, B.H., Kieran, M.W., Langenau, D.M., Delahaye-Brown, A., Zon, L.I., Fleming, M.D., and Look, A.T. (2001). Myelopoiesis in the zebrafish, Danio rerio. Blood *98*, 643–651.

81. Liu, T.X., Rhodes, J., Deng, M., Hsu, K., Radomska, H.S., Kanki, J.P., Tenen, D.G., and Look, A.T. (2007). Dominant-interfering C/EBPalpha stimulates primitive erythropoiesis in zebrafish. Exp. Hematol. *35*, 230–239. https://doi.org/10.1016/j.exphem.2006.10.008.

82. Jowett, T., and Yan, Y.L. (1996). Double fluorescent in situ hybridization to zebrafish embryos. Trends Genet. *12*, 387–389.

83. Dobrzycki, T., Krecsmarik, M., Bonkhofer, F., Patient, R., and Monteiro, R. (2018). An optimised pipeline for parallel image-based quantification of gene expression and genotyping after in situ hybridisation. Biol. Open *7*, bio031096. https://doi.org/10.1242/bio.031096.

84. Kroll, F., Powell, G.T., Ghosh, M., Gestri, G., Antinucci, P., Hearn, T.J., Tunbak, H., Lim, S., Dennis, H.W., Fernandez, J.M., et al. (2021). A simple and effective F0 knockout method for rapid screening of behaviour and other complex phenotypes. Elife *10*, e59683. https://doi.org/10.7554/eLife.59683.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
H2A.XS139ph	GeneTex	GTX127342; RRID:AB_2833105
Alexa Fluor 594 goat anti-rabbit	Invitrogen	A11012; RRID:AB_2534079
γH2AX	Millipore	05-636; RRID:AB_309864
H2A	Millipore	07-146; RRID:AB_310394
GATA2	R&D Systems	MAB2046-SP; RRID:AB_2108424
Anti-rabbit HRP	Agilent	P0399; RRID:AB_2617141
Anti-goat HRP	Agilent	P0449; RRID:AB_2617143
Anti-Mouse HRP	Agilent	P0447; RRID:AB_2617137
Anti-Digoxigenin-AP	Roche	11093274910; RRID:AB_514497
Chemicals, peptides, and recombinant proteins		
StemPro-34 SFM	Invitrogen	10639011
Pen/Strep	Gibco	15140122
L-glutamine	Gibco	25030081
SCF	Peprotech	250-03
pan-GATA inhibitor pyrrothiogatain	Generon	AOB13027-5
NPM1 oligomerization inhibitor NSC348884	Selleck Chemicals	S8149-SEL
superfrost plus slide	Thermo Scientific	12372098
May-Grünwald	Sigma	MG500
Giemsa stain, modified solution	Sigma-Aldrich	48900-1L-F
DPX	Sigma-Aldrich	44581
trypan blue	Sigma-Aldrich	T6146
FCS	Gibco	26140079
Hoechst's	Invitrogen	11534886
PFA	Thermo Scientific	28908
Triton	Alfa Aesar	A16046
NaCl	Sigma	101856436
MgCl2	Sigma	102111007
sucrose	Sigma	102192646
PIPES	Alfa Aesar	J63234
DAPI	abcam	ab104139
DraIII	NEB	R3510
mMESSAGE mMACHINE™ T7 Transcription Kit	Thermofisher	AM1344
TE buffer (Tris-HCL and EDTA) Trizma base EDTA	Sigma-Aldrich Sigma-Aldrich	T1503 E6758
Ethanol	VWR	20821
20x SSC	BioRad	1610775
BM Purple AP Stain	Roche	11442074001
Blocking Reagent	Roche	11096176001
Formamide	VWR	24311.291
Tween 20	Sigma-Aldrich	P2287
Sodium Chloride	J.T.Baker	15368426
Maleic Acid	Sigma-Aldrich	M0375
Magnesium Chloride	Sigma-Aldrich	M9272
Hydrogen Peroxide	J.T.Baker	15587984

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
10x PBS	Gibco	14200-067
Citric Acid	Sigma-Aldrich	C0759
Torula RNA	Sigma-Aldrich	R6875
Heparin	Sigma-Aldrich	H3400
DMSO	Sigma-Aldrich	D2650
Tricaine	Sigma-Aldrich	E10521
Methylcellulose	Sigma-Aldrich	M0387
Deposited data		
RAW and analyzed data	This paper	GEO: GSE200503
scRNA-Seq: Single-cell RNA-seq reveals a distinct transcriptome signature of hematopoiesis in GATA2 deficiency	Wu et al.[32]	GEO: GSE135194
Gata2a ChIP-Seq (ArrayExpress)	Sanchéz-Castillo et al.[34]	BioStudies: E-MATB-3954 https://www.ebi.ac.uk/biostudies/arrayex press
Cebpa ChIP-Seq (ArrayExpress)	Sanchéz-Castillo et al.[34]	BioStudies: E-MATB-3954 https://www.ebi.ac.uk/biostudies/arrayexpress
DnaseI ChIP-Seq (ArrayExpress)	Sanchéz-Castillo et al.[34]	BioStudies: E-MATB-3954 https://www.ebi.ac.uk/biostudies/arrayexpress
Gata2 ChIP-Seq	Sanchéz-Castillo et al.[34]	GEO:GSM552234 https://www.ncbi.nlm.nih.gov/geo/
H3K27AC ChIPseq	Sanchéz-Castillo et al.[34]	GEO:GSM1329815 https://www.ncbi.nlm.nih.gov/geo/
H3K4me1 ChIp-Seq	Sanchéz-Castillo et al.[34]	GEO:GSE47085 https://www.ncbi.nlm.nih.gov/geo/
Analysis code	This study	https://zenodo.org/badge/latestdoi/595093929
Experimental models: Cell lines		
HPC-7	Pinto do Ó et al.[48]	RRID:CVCL_RB19
Experimental models: Organisms/strains		
zebrafish Tg(-6.0itga2b:EGFP)^{la2Tg}	Li et al.[63]	*N/A*
Zebrafish Tg(gata1a:DsRed)sd2Tg	Traver et al.[64]	N/A
Zebrafish Tg(mpx:GFP)c264Tg	Renshaw et al.[15]	N/A
Zebrafish Tg(mpeg1:GFP)gl22	Ellet et al.[65]	N/A
Zebrafish *gata2a*$^{\Delta i4/\Delta i4}$	Dobrzycki et al.[13]	N/A
Zebrafish Tg(runx1P2:Citrine)^{ox1Tg}	Bonkhofer et al.[45]	*N/A*
Recombinant DNA		
GATA2 ORF plasmid	Genscript	OMu19368D
Software and algorithms		
CellRanger (v3.1.0)	10x Genomics	https://www.10xgenomics.com/
Loupe(v6.0)	10x Genomics	https://www.10xgenomics.com/; RRID:SCR_018555
Seurat (v4.04)	Butler et al.[66]; Satija et al.[67]; Stuart et al.[68]	https://satijalab.org/seurat/index.html; RRID:SCR_007322
Metascape	Zhou et al.[69]	https://metascape.org/gp/index.html
Monocle3	Cao et al.[70]; Levine et al.[71]; Qiu et al.[72]; Traag et al.[73]; Trapnell et al.[74]	https://cole-trapnell-lab.github.io/monocle3/; RRID:SCR_018685
MACS3		https://github.com/macs3-project/MACS
CellRanger ATAC (v1.2.0)	10xGenomics	https://www.10xgenomics.com/
Signac (v1.4.0)	Stuart et al.[75]	https://stuartlab.org/signac/; RRID:SCR_021158

(*Continued on next page*)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
HOMER (v4.4)	Heinz et al.[76]	http://homer.ucsd.edu/homer/; RRID:SCR_010881
Java TreeView	Saldanha[77]	https://jtreeview.sourceforge.net/; RRID:SCR_016916
Fiji	Schindelin et al.[78]	https://imagej.net/software/fiji/; RRID:SCR_002285
GraphPad Prism (v9.0)	GraphPad software	https://www.graphpad.com/
AxioVision software	Zeiss	https://www.micro-shop.zeiss.com/en/us/system/axiovision+software-software+axiovision-software/6014/
Other		
ISH probe: *runx1*	Kalev-Zylinska[79]	N/A
ISH probe: *mpx*	Bennett et al.[80]	N/A
ISH probe: *cebpa*	Liu et al.[81]	N/A
crRNAs for *gata2a* (Dr.Cas9.GATA2A.AA) Sequence: ATAGCCAAGCTTCCCGAAG	IDT	N/A
crRNAs for *gata2a* (Dr.Cas9.GATA2A.AB Sequence: GAACACATCCACCTCGTCCG	IDT	N/A
crRNAs for *gata2a* (Dr.Cas9.GATA2A.AC) Sequence: TCTGGTGATACGTCTTTCGG	IDT	N/A
Cas9 protein (Alt-R® S.p. HiFi Cas9 Nuclease V3)	IDT	1081060

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Rui Monteiro, r.monteiro@bham.ac.uk.

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Single-cell RNA-seq and single-cell ATAC-seq data have been deposited in GEO and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- All original code has been deposited at github and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Zebrafish

AB* zebrafish strains, along with transgenic and mutant strains, were kept in a 14/10 h light/dark cycle at 28°C. We used the following transgenic animals: *Tg(-6.0itga2b:EGFP)*^{la2Tg} (referred to as *cd41*:GFP),[63] *Tg(gata1a:DsRed)*^{sd2Tg},[64] *Tg(mpx:GFP)*^{c264Tg},[15] *Tg(mpeg1:GFP)*^{gl22},[65] *Tg(runx1P2:Citrine)*^{ox1Tg}.[45] The gata2a mutant is an enhancer deletion mutant (*gata2a*$^{\Delta i4/\Delta i4}$),[13] here referred to as *gata2a*$^{-/-}$ for simplicity. Experimental procedures in adult fish were done in accordance with the Animal Scientific Procedure Act 1986 under an approved Home Office Project License (PPL2470547). A mix of male and female fish were used throughout this study. Adult fished used in this study were either 4, 5, 6, 7 or, 12mpf (as indicated on figures). Embryonic fish used in this study were 28hpf or 48hpf (as indicated on figures).

Cell lines

HPC7 murine multipotent progenitor cells were obtained from Prof J. Frampton (University of Birmingham), and maintained in StemPro-34 SFM (Invitrogen, 10639011) containing supplement, 1% Pen/Strep, 1 x L-glutamine and 100 ng/mL SCF. Cells

were treated with the pan-GATA inhibitor pyrrothiogatain (Generon, AOB13027-5[49]), the NPM1 oligomerization inhibitor NSC348884 (Selleck Chemicals, S8149-SEL)[43] or a combination of both for 24h (concentrations stated in the figure legend).

METHOD DETAILS

May-Grünwald (MG) and Giemsa staining of WKM smears

Whole kidney marrow tissue was dissected from sacrificed animals and smeared on a superfrost plus slide (Thermo Scientific) and left to air dry for at least 1h at RT. MG (Sigma) was diluted 1:1 and Giemsa (Sigma) 1:10 with distilled water. Slides were incubated with diluted MG for 5 min at RT, then washed and repeated with diluted Giemsa for 30 min. Slides were washed well with distilled water and left to air dry before being mounted by DPX (Sigma) and imaged using a Leica DM750 scope with Leica ICC50W digital camera attachment.

Flow cytometry and cell sorting

Cells were counted using a haemocytometer (Marionfeld) and dead cells were stained using trypan blue (Sigma, cat: T6146) diluted 1:10. Cells were dissociated and re-suspended in 10% FCS (GIbco) 1xPBS and analyzed using a BD LSRFORTESSA X-20 and dead cells were excluded using Hoechst's (Invitrogen, cat: 11534886), used at 1:10000. Cells were sorted using BD FACSAria Fusion. For γH2AX, WKM cells were fixed in 4% PFA for 10 min at RT, permeabilised in 0.1% Triton for 5 min at RT, blocked in 4% FBS for 20 min at RT and stained with H2A.XS139ph (GeneTex, GTX127342) (diluted 1:1000 in 3% FCS/PBS) followed by Alexa Fluor 594 goat anti-rabbit (Invitrogen, A11012), diluted 1:1000 in 3% FCS/PBS.

Single cell RNA sequencing (scRNAseq)

3′-scRNAseq was completed using Chromium Next GEM single cell 3′ GEM library and Gel bead Kit v3.1 (10x Genomics) and sequenced using a NextSeq 500 (Illumina) at Genomics Birmingham (University of Birmingham). Analysis was completed using CellRanger (10x Genomics: v3.1.0), Loupe (10x Genomics: v5.1) and R (v4.1): Using Seurat (v4.0.4).[66–68] For the scRNAseq experiment in Figure 2, Loupe browser was used to label cell types based on expression of key markers (Figure S1A and Table S5 for barcode cluster assignment). Data was then analyzed in Seurat, where the following QC metrics were used: nFeature_RNA >100 & nFeature_RNA <4000 & nCount_RNA >100 & nCount_RNA <30000 (although nCount_RNA <40000 was used for gata2a$^{-/-}$ sample) & percent.mito <20%. All data related to scRNAseq experiment in Figure 2 was analyzed using SCTransform(), unless indicated (as in Figure 2C where data from RNA assay is shown calculated using LogNormalize(), followed by ScaleData() and RunPCA()). Differential expression was calculated using FindMarkers() in Seurat using the wilcox statistical test. Seurat was used to plot heatmaps and violin plots, EnhancedVolcano for Volcano Pots (where log2FC was plotted vs. -Log10p value). Volcano plots and heatmaps in Figure S5 were plotted using Graphpad Prism (v9). GO term enrichment analysis was completed for biological function using Metascape.[69] Pseudotime analysis was completed using Monocle3 using a UMAP projection plotted in Loupe (Table S6). Cells that passed the QC metrics above where then analyzed using Monocle3.[70–74] Mapping of GFP was completed following this protocol: https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/tutorial_mr#panel. For the scRNAseq experiment in Figure 6, analysis was carried out as outlined above, except for the QC, where cells with at least 1000 UMI counts were retained (the following cells were analyzed: Table S7). The UMAP was imported from loupe (UMAP coordinates: Table S8).

Single cell assay for transposable chromatin followed by sequencing (scATAC-seq)

scATAC (assay for transposable accessible chromatin assay)-seq was completed using Chromium Next GEM Single Cell ATAC Library & Gel Bead Kit v1.1, Chromium Next GEM Chip H Single Cell Kit and Chromium i7 Multiplex Kit N, Set A (10x Genomics). Libraries were sequenced using NextSeq 500 (Illumina) at Genomics Birmingham (University of Birmingham). Peaks were called using MACS3 using the following commands: macs3 callpeak -t ~/MUT1/outs/possorted_bam.bam ~/MUT2/outs/possorted_bam.bam ~//MUT3/outs/possorted_bam.bam -n MUT -g 1.3E9 and callpeak -t ~/WT1/outs/possorted_bam.bam ~/WT2/outs/possorted_bam.bam ~//WT3/outs/possorted_bam.bam -n WT -g 1.3E9. Peaks were filtered for qValue <0.01 and signalValue >2. Analysis was completed using CellRanger ATAC (10xGenomics: v1.2.0) and peaks were specified using the output from MACS3 during the cellranger-atac reanalyze step. An overarching WT peak file was use for all WT samples and an overarching MUT peak file was use for all MUT samples. We also used R (v4.1): using Seurat (as above) and Signac (v1.4.0).[75] Cells with <6000 peak region fragments, >40 pct reads in peaks, <1.5 nucleosome signal, and >2 TSS enrichment were used for analysis. Results were presented using Seurat and Signac. Motif scanning of chosen regions (open peaks: FC > 0.2, closed peaks FC < −0.2 when comparing *gata2a*$^{-/-}$ and WT) was completed using findMotifsGenome.pl in HOMER (v4.4).[76] Peak heatmaps were constructed by running FindMarkers() in R with the following parameters to obtain a Log2FC for every peak in the assay: output <- FindMarkers(Seurat_obj, ident.1 = "1″, ident.2 = "2″, test.use = 'LR', latent.vars = 'nCount_peaks', logfc.threshold = -Inf, min.pct = -Inf, only.pos = FALSE). Peaks were then ordered by Log2FC, converted to FC and peaks where FC = 1 were removed. Peaks were annotated using HOMER and the following commands: annotatePeaks.pl peaksBED.txt danRer11-size 2000 -hist 10 -ghist -bedGraph peaksBDG.txt > peaks.cdt. Peaks were annotated with motif sites using: annotatePeaks.pl peaksBED.txt danRer11 -size 2000 -hist 10 -ghist -m motif.motif > peaks_motif.cdt. Peaks heatmaps were constructed using Java TreeView.[77]

Immunofluorescence

WKM cells were cytospun onto Superfrost plus slides (Thermo Scientific, cat:12372098) and left to air dry for 15 min at RT. Slides were pre-extracted (with pre-extraction buffer: 20mM NcCl, 3mM MgCl2, 300mM sucrose, 10mM PIPES, 0.5% Triton X-100) on ice for 5 min and then fixed in 4% PFA for 10 min at RT. Slides were washed in PBS 3 times and then blocked in 100ul blocking buffer (10% FCS/PBS) for 1 h at RT. H2A.XS139ph (GeneTex, GTX127342), 1:500 (diluted in 3% FCS/PBS), was added to the slide for 1 h at RT, washed in PBS and then incubated with AlexaFluro 594 goat anti-rabbit (Invitrogen, A11012), 1:1000 (diluted in 3% FCS/PBS). Slides were then washed in PBS and mounted using DAPI mounting medium (abcam, ab104139).

Western blot

HPC7 cells were harvested by centrifugation and whole cell extracts were obtained by lysis in UTB buffer (8 M Urea, 50 mM Tris, 150 mM β-mercaptoethanol). Lysates were separated and analyzed by SDS-PAGE following standard procedures using a Tris-Bicine running buffer. The following antibodies were used: γH2AX (Millipore; cat#05–636; RRID:AB_309864), H2A (Millipore; cat#07–146; RRID:AB_11212920), Anti-rabbit HRP (Agilent; cat#P0399; RRID: AB_2617141), Anti-goat HRP (Agilent; Cat# P0449, RRID: AB_2617143), GATA2 (R&D; MAB2046-SP). Immunoblots were scanned, and band intensities quantified using the 'Gels' option in Fiji.[78] Data was processed using GraphPad Prism 9.0.

Cell proliferation assay

HPC7 cells were plated at a concentration of $1x10^6$ cells per well in a 24 well plate. The pan-GATA inhibitors and NPM1 inhibitors (as above) were added in the same concentrations as previously. Cells were counted manually using cell counting slides at 24, 48 and 72 h after adding the inhibitors. Data was plotted and analyzed for statistical significance using GraphPad Prism 9.0.

Whole-mount *in situ* hybridisation

Whole-mount *in situ* hybridization (ISH) was carried out as described previously,[82] using probes for *runx1*,[79] *mpx*[80] and *cebpa* (a kind gift from Elspeth Payne, UCL).[81] Embryos at the required stages were imaged with a Moticam S6 camera mounted on a Nikon E800N stereomicroscope. Analysis and quantification of *runx1* and *cebpa* gene expression was performed as described.[83] Briefly, images were converted to grayscale and inverted using Fiji.[78] Next, we measured pixel intensity in an ROI containing the relevant staining against an unstained background ROI to obtain a pixel intensity value for each embryo. The intensity values were plotted, and statistical tests (One way ANOVA) performed using GrapPad 9.0 software.

CRISPR/Cas9 F0 injections and mRNA injections

The Alt-R CRISPR/Cas9 system used to target *gata2a* was purchased from Integrated DNA Technologies (IDT). Briefly, pre-designed crRNAs for *gata2a* (Dr.Cas9.GATA2A.AA, Dr.Cas9.GATA2A.AB and Dr.Cas9.GATA2A.AC, see Figure S3E) were individually annealed with universal tracrRNA to make gRNAs. These were then each assembled to Cas9 protein (Alt-R S.p. HiFi Cas9 Nuclease V3, 10 mg/ml, Catalog #1081060), generating Cas9/gRNA ribonucleoproteins (RNPs).[84] 1nL of pooled RNPs (concentration is 28.5μM or 9.5μM/RNP) were injected into one cell stage-embryos. Embryos were grown up to the required stages, manually dechorionated and fixed in 4% PFA at 4°C overnight before dehydration to 100% ethanol in preparation for ISH.

To prepare wildtype GATA2 mRNA for the rescue experiments, the plasmid containing the murine GATA2 ORF (Genscript, cat #OMu19368D) was linearised with DraIII (NEB) and used as template for *in vitro* transcription using the mMESSAGE mMACHINE T7 Transcription Kit (Thermofisher, AM1344) as per kit instructions. The mRNA was diluted to 100 ng/μl in TE buffer (mRNA only) or directly into the pooled RNP mix (sgRNA and mRNA condition). 1nL of each mix was injected into the single cell of one cell stage-embryos. All embryos were grown to 28hpf, manually dechorionated and fixed in 4% PFA at 4°C before dehydration to 100% ethanol in preparation for ISH.

Live imaging

To investigate whether inhibition of NPM1 affects HSPC development, transgenic Tg(runx1P2:Citrine)^{ox1Tg} [45] embryos were treated with DMSO or 30μM of the NPM1 inhibitor NSC348884[40] from 24 to 48hpf. To quantify the number of HSPCs post-treatment, transgenic embryos were anesthetised in tricaine (160 μg/mL), mounted in 3% methyl cellulose and imaged with an AxioCam MRm camera (Zeiss) and AxioVision software (Zeiss) in a Stereo Lumar V.12 stereomicroscope (Zeiss).

QUANTIFICATION AND STATISTICAL ANALYSIS

Graphpad Prism (v9) was used for statistical analysis in Figure 1B, 1D–1G, 5B, 5H, 6D, 6E, and S4F, S5G, S6A, S6C, S6D. Detail of statistical tests used are indicated in figure legends along with p values.

FindMarkers() and the Wilcox test from Seurat (v4.04) in R (v4.1) was used to calculate the p values in Figures 3A, 3B, 3D, 3E, 5D, and S3C, S3E, S5B.