Deletion of a conserved Gata2 enhancer impairs haemogenic endothelium programming and adult Zebrafish haematopoiesis

Tomasz Dobrzycki1, Christopher B. Mahony2, Monika Krecsmarik1,3, Cansu Koyunlar4, Rossella Rispoli1,5, Joke Peulen-Zink4, Kirsten Gussinklo4, Bakhta Fedlaoui2, Emma de Pater4, Roger Patient1,3 & Rui Monteiro1,2,3✉

Gata2 is a key transcription factor required to generate Haematopoietic Stem and Progenitor Cells (HSPCs) from haemogenic endothelium (HE); misexpression of Gata2 leads to haematopoietic disorders. Here we deleted a conserved enhancer (i4 enhancer) driving pan-endothelial expression of the zebrafish gata2a and showed that Gata2a is required for HE programming by regulating expression of runx1 and of the second Gata2 orthologue, gata2b. By 5 days, homozygous gata2aΔi4/Δi4 larvae showed normal numbers of HSPCs, a recovery mediated by Notch signalling driving gata2b and runx1 expression in HE. However, gata2aΔi4/Δi4 adults showed oedema, susceptibility to infections and marrow hypo-cellularity, consistent with bone marrow failure found in GATA2 deficiency syndromes. Thus, gata2a expression driven by the i4 enhancer is required for correct HE programming in embryos and maintenance of steady-state haematopoietic stem cell output in the adult. These enhancer mutants will be useful in exploring further the pathophysiology of GATA2-related deficiencies in vivo.
Hematopoietic stem cells (HSCs) are the source of all blood produced throughout the lifetime of an organism. They are capable of self-renewal and differentiation into progenitor cells that generate specialised blood cell types. DNA-binding transcription factors are fundamental players in the inception of the haematopoietic system as it develops in the embryo, but also play a crucial role in maintaining homeostasis of the haematopoietic system in the adult organism. They coordinate differentiation, proliferation and survival of haematopoietic cells and ensure their levels are appropriate at all times throughout life. Misexpression of key transcription factors may thus lead to a failure to produce HSCs or, alternatively, to haematopoietic disorders and eventually leukaemia. Therefore, understanding how transcription factors drive the haematopoietic process provides opportunities for intervention when haematopoiesis is dysregulated.

The development of blood occurs in distinct waves: primitive, pro-definitive and definitive, each of them characterised by the generation of blood progenitors in a specific location and restricted in time, where the definitive wave produces multi-lineage self-renewing HSCs. The specification of HSCs initiates in cells with arterial characteristics and proceeds through an endothelial intermediate, termed the haemogenic endothelium (HE)3. In zebrafish and other vertebrates, expression of runx1 defines the bona fide HE population. Haematopoietic stem and progenitor cells (HSPCs) emerge from the HE by endothelial-to-haematopoietic transition (EHT), both in zebrafish and in mice6–8. They arise between 28 and 48 h post fertilisation (hpf) from the HE in the ventral wall of the dorsal aorta (DA)9, the analogue of the mammalian aorta-gonad-mesonephros (AGM)10. After EHT, the HSCs enter the bloodstream through the posterior cardinal vein (PCV)9 to colonise the caudal haematopoietic tissue (CHT), the zebrafish equivalent of the mammalian foetal liver11,12. Afterwards the HSCs migrate again within the bloodstream to colonise the kidney marrow (WKM) and thymus9, the final niche for HSCs, equivalent to the bone marrow in mammals1.

Gata2 is a key haematopoietic transcription factor (TF) in development. In humans, GATA2 haploinsufficiency leads to blood disorders, including MonoMAC syndrome (Mono-cytopenia, Mycobacterium avium complex) and myelodysplastic syndrome (MDS)12,13. While its presentation is variable, Mono-MAC syndrome patients always show cytopenias, ranging from mild to severe, and hypocellular bone marrow13,14. These patients are susceptible to mycobacterial and viral infections, and have a propensity to develop MDS and Acute Myeloid Leukaemia (AML), with a 75% prevalence and relatively early onset at age 2013.

Gata2 knockout mice are embryonic lethal and die by E10.515. Conditional Gata2 knockout under the control of the endothelial VE-cad promoter abolished the generation of intra-aortic clusters16, suggesting that Gata2 is required for HSPC formation. Further studies in the mouse revealed a decrease in HSC numbers in Gata2 heterozygous mutants, but also a dose-dependency of adult HSCs on Gata217.

Gata2 expression in the endothelium is regulated by an intronic enhancer element termed the +9.5 enhancer18,19. Deletion of this enhancer results in the loss of HSPC emergence from HE, leading to lethality by E1419. The same element is also mutated in 10% of all the MonoMAC syndrome patients12.

Because of a partial genome duplication during the evolution of teleost fish, numerous zebrafish genes exist in the form of two paralogues, including gata220. This provides an opportunity to separately identify the temporally distinct contributions made by each Gata2 orthologue. Gata2a and gata2b are only 57% identical and are thought to have undergone evolutionary sub-functionalisation from the ancestral vertebrate Gata2 gene21,22.

Gata2b is expressed in HE from 18hpf and is thought to regulate runx1 expression in HE23. Lineage tracing experiments showed that gata2b-expressing HE cells gave rise to HSCs in the adult21. Similar to the mouse Gata2, gata2b expression depends on Notch signalling and is a bona fide marker of HE, currently regarded as the functional ‘haematopoietic homologue’ of Gata2 in zebrafish23. By contrast, gata2a is expressed in all endothelial cells and in the developing central nervous system12,23. Homozygous gata2aum27 mutants showed arteriovenous shunts in the dorsal aorta at 48hpf24. However, gata2a is expressed at 11hpf in the haemangioblast population in the posterior lateral mesoderm (PLM) that gives rise to the arterial endothelial cells in the trunk25, well before gata2b is expressed in HE. This suggests that gata2a might play a role in endothelial and HE programming and thus help to elucidate an earlier role for Gata2 in HSC development.

Here we show that the gata2a locus contains a conserved enhancer in its 4th intron, correspondig to the described +9.5 enhancer in the mouse Gata2 locus18,19. Using CRISPR/Cas9 genome editing, we demonstrated that this region, termed the i4 enhancer, is required for endothelial-specific gata2a expression26. Homozygous mutants (gata2aΔi4/Δi4 mutants) showed decreased expression of the HE-specific genes runx1 and gata2b. Thus, endothelial expression of gata2a, regulated by the i4 enhancer, is required for gata2b and runx1 expression in the HE. Strikingly, their expression recovers and by 48hpf, the expression of haematopoietic markers in gata2aΔi4/Δi4 mutants is indistinguishable from wild-type siblings. We have demonstrated that this recovery is mediated by an independent input from Notch signalling, sufficient to recover gata2b and runx1 expression in HE and thus HSPC emergence by 48hpf. We conclude that runx1 and gata2b are regulated by two different inputs, one Notch-independent input from Gata2a and a second from the Notch pathway, acting as a fail-safe mechanism for the initial specification of HSPCs in the absence of the input by Gata2a. Despite the early rescue, gata2aΔi4/Δi4 adults showed increased susceptibility to infections, oedema, a hypocellular WKM and neutropenia, a phenotype resembling key features of GATA2 deficiency syndromes in humans. We conclude that Gata2a is required for HE programming in the embryo and to maintain the steady-state haematopoietic output from adult HSPCs and that this function requires the activity of the i4 enhancer.

Results
Analysis of open chromatin regions in the gata2a locus. Because Gata2 genes are duplicated in zebrafish, we set out to unpick the different roles Gata2a and Gata2b play during HSC generation and homeostasis by identifying their regulatory regions. Analysis of sequence conservation revealed that one region within the fourth intron of the zebrafish gata2a locus was conserved in vertebrates, including mouse and human (Fig. 1a–c). This region, which we termed ‘i4 enhancer’, corresponds to the endothelial +9.5 Gata2 enhancer identified previously in the mouse18,19 and human26. Notably, the gata2b locus did not show broad conservation in non-coding regions (Supplementary Fig. 1a).

To investigate whether the i4 element was a potentially active enhancer, we first performed ATAC-seq to identify open chromatin regions in endothelial cells (ECs) in zebrafish. We used a Tg(kdrl:GFP) transgenic line that expresses GFP in all endothelium28 and isolated the higher GFP-expressing ECs (kdrl:GFPhigh, termed kdrl:GFP+ for simplicity) as this fraction was enriched for endothelial markers compared to the kdrl:GFP fraction (Supplementary Fig. 1b, c). Principal Component Analysis on the ATAC-seq data from 26hpf kdrl:GFP+ cells
**Fig. 1** The i4 enhancer in the gata2a locus is conserved and drives pan-endothelial expression of a GFP reporter in zebrafish. 

a) Kdrl:GFP+ (green) and kdr:GFP− (blue) cells were FACS-sorted from 26hpf embryos and used for preparation of ATAC-seq libraries. 

b) The image of the mapped reads represents stacked means of two biological ATAC-seq replicates. Differential peak analysis identified four chromatin regions (blue shading) in the locus of gata2a that are significantly more open in the kdr:GFP+ population (p < 0.0001). A region in the fourth intron (termed i4 enhancer) is conserved throughout vertebrates. 

Black and grey shading denotes regions of high conservation between the species analysed. 

c) The highly conserved 150 bp region (red) contains putative transcription factor binding sites, mapped computationally. Light blue: Ets binding sites; purple: E-box binding sites; green: GATA binding sites; asterisks: conserved residues. 

d) Widefield fluorescent image of a live Tg(gata2a-i4-1.1 kb:GFP) zebrafish embryo at 27hpf showing GFP fluorescence in the endothelial cells and in the heart (endocardium). 

e) Higher magnification image of the trunk of the embryo from panel d. f–f′ Confocal images of a trunk fragment of a Tg(gata2a-i4-1.1 kb:GFP) embryo immunostained with anti-GFP antibody (f) and probed for gfp mRNA (f′) at 25hpf. 

f′ Merged images from panels f–f′ with Hoechst nuclear staining in blue, showing complete overlap of GFP protein and mRNA. 

g–g″ Confocal images of the dorsal aorta (DA) and posterior cardinal vein (PCV) of a Tg(gata2a-i4-1.1 kb:GFP) embryo immunostained with anti-GFP antibody (g) and probed for runx1 mRNA (g′) at 25hpf. See panel e for approximate position within the embryo. 

h) Merged images from panels g–g′, also showing Hoechst nuclear staining in blue. 

h′ Counting of the runx1+ cells represented in panels g–g′ in 25 embryos shows that >90% of runx1+ cells are also GFP+. N = 3. Error bars: ± SD. See also Supplementary Fig. 1.

---

**Notes:**
- Fig. 1 illustrates the conservation and expression patterns of the i4 enhancer in the gata2a locus.
- The i4 enhancer is conserved throughout vertebrates and drives pan-endothelial expression in zebrafish.
- Differential peak analysis identified specific chromatin regions that are more open in GFP+ cells.
- The highly conserved 150 bp region contains putative transcription factor binding sites.
- GFP fluorescence is observed in endothelial cells and the heart at early developmental stages.
(n = 2) and kdrl:GFP^− cells (n = 4) revealed strong differences between the open chromatin regions in the two cell populations, further supported by a correlation analysis (Supplementary Fig. 1d–f). 78,026 peaks were found in common between replicates of the ATACseq in kdrl:GFP^+ cells (Supplementary Fig. 1g). 44,025 peaks were differentially expressed between the kdrl:GFP^+ and kdrl:GFP^− fractions (Supplementary Fig. 1h). An analysis of known motifs present in the kdrl:GFP^+ population revealed an enrichment for the ETS motif (Supplementary Fig. 1i). ETS factors are essential regulators of gene expression in endothelium.30 In addition, we performed gene ontology (GO) term analysis on the peaks showing >3-fold enrichment or depletion in ECs (Supplementary Fig. 1j–l). As expected, non-ECs showed a broad range of GO terms whereas EC-enriched peaks were associated with terms like angiogenesis or blood vessel development (Supplementary Fig. 1k, l).

Differential peak analysis in the gata2a locus identified four differentially open sites within a 20 kb genomic region (Fig. 1b), including one peak in intron 4 corresponding to the predicted i4 enhancer. It contained a core 150 bp-long element that included several binding motifs for the GATA, E-box and Ets transcription factors (Fig. 1b). Although the positioning of the E-box site relative to the adjacent GATA site differs in zebrafish and mammals (Fig. 1b, c), the necessary spacer distance of ~9 bp between the two sites was conserved. Thus, this site may be a target for TF complexes containing an E-box-binding factor and a GATA family TF.

Thus, the intronic enhancer (i4) identified in the zebrafish gata2a locus is accessible to transposase in endothelial cells and contains highly conserved binding sites for key haematopoietic transcription factors, suggesting that genetic regulation of gata2a expression in zebrafish HE is a conserved feature of vertebrate gata2 genes.

The gata2a-i4 enhancer drives GFP expression in endothelium. To investigate the activity of the gata2a-i4 enhancer in vivo, the conserved genomic 150 bp region (Fig. 1b, c), together with flanking ±500 bp (gata2a-i4-1.1 kb-GFP) or ±150 bp (gata2a-i4-450 bp-GFP) was cloned into a Tol2-based reporter Elb:GFP construct31 and used to generate stable transgenic lines (Supplementary Fig. 2). The earliest activity of the enhancer was observed at the 14-somite stage (14ss), when gfp mRNA was detected in the PLM (Supplementary Fig. 2a, b). After 22hpf, the reporter signal was pan-endothelial (Fig. 1d–e, Supplementary Fig. 2c–i). Around 27hpf, higher intensities of GFP fluorescence and correspondingly higher levels of gfp mRNA were visible in the floor of the DA (Fig. 1d–e, Supplementary Fig. 2e–h). While the GFP protein was still visible in the vasculature around 3dpf, it was likely carried over from earlier stages, since the gfp mRNA was not detectable any more (Supplementary Fig. 2i, j). We focussed our subsequent analysis on the gata2a-i4-1.1 kb-GFP transgenics as they showed stronger expression of the transgene. At 25hpf, the expression of GFP protein and gfp mRNA overlapped completely in the endothelial cells of the DA (Fig. 1f–g). Overall, these data confirm that the i4 enhancer is active in vivo in endothelial cells at the correct time to regulate definitive haematopoiesis. The endothelial activity of the corresponding +9.5 enhancer was also observed in mouse embryos,38 indicating functional conservation of the gata2a-i4 enhancer across vertebrates.

To further characterise the enhancer activity in vivo, Tg (gata2a-i4-1.1 kb-GFP) embryos were stained for gata2a mRNA and for GFP protein (Supplementary Fig. 2k–o). We found a large overlap between gata2a^+ and GFP^+ cells at 30hpf in the DA, with a small proportion of GFP^+ cells that did not express gata2a mRNA (<5%, Supplementary Fig. 2o). This could suggest that some cells require activity of other endothelial enhancers to trigger transcription of gata2a or that gfp mRNA has a longer half-life than gata2a mRNA. Importantly, the GFP signal was absent in gata2a-expressing neural cells (Supplementary Fig. 2l–n), indicating that the i4 enhancer is specifically active in (haemogenic) endothelial cells.

Next we examined the expression of the HE marker runx1 in gata2a-i4-1.1 kb-GFP embryos at 25hpf. At this stage, over 90% of runx1^+ cells were GFP^+ (Fig. 1g–h). We conclude that the GFP expression under the gata2a-i4 enhancer marks the majority of the HE population.

Endothelial gata2a required gata2a-i4 enhancer activity. To investigate whether endothelial-specific expression of gata2a is required for definitive haematopoiesis, we deleted the conserved gata2a-i4 enhancer using CRISPR/Cas9 genome editing.32 We generated a deletion mutant lacking 231 bp of the i4 enhancer (Supplementary Fig. 3a–c) and named it gata2aΔi4/i4. Homozygous gata2aΔi4/i4 mutants showed decreased levels of gata2a expression in endothelial cells when compared to wild-type embryos (Fig. 2a, b). By contrast, gata2a expression in the neural tube appeared unaffected in the gata2aΔi4/i4 mutants (Fig. 2a, b). At 28hpf, expression of the pan-endothelial marker kdrl was indistinguishable between wild-type and gata2aΔi4/i4 mutants (Fig. 2c, d). To verify these results, we crossed homozygous gata2aΔi4/i4 mutants to Tg(kdrl:GFP) transgenics and analysed vascular morphology. Gata2aΔi4/i4 embryos showed no gross vascular abnormalities at 48hpf as assessed by the expression of the Tg(kdrl:GFP) transgene (Fig. 2e, f).

Next, we isolated endothelial cells from Tg(kdrl:GFP) and Tg(kdrl:GFP); gata2aΔi4/i4 embryos by FACS (Fig. 2g) at 23hpf and 30hpf and confirmed by qRT-PCR that the endothelial markers kdrl, dld and dll4 were unaffected in gata2aΔi4/i4 embryos (Supplementary Fig. 3d–f). The arterial marker efnb2a33 was decreased at 23hpf in gata2aΔi4/i4 mutants but recovered by 30hpf (Supplementary Fig. 3g). In addition, gata2a was significantly decreased in the kdrl:GFP^+ ECs in 23hpf gata2aΔi4/i4 embryos compared to wild types (Fig. 2h). At 30hpf this decrease was not statistically significant. This was likely due to a decrease in expression of gata2a that appears to occur in wild-type ECs during development, whereas gata2a expression in mutants remained low (Fig. 2h). Importantly, there was no difference in gata2a expression in the non-endothelial population (kdrl:GFP^− cells) between wild-type and gata2aΔi4/i4 mutants at either 23 hpf or 30hpf (Fig. 2i). Altogether, these data suggest that genomic deletion of the gata2a-i4 enhancer is sufficient to reduce expression of gata2a specifically in endothelium.

Gata2a regulates runx1 and gata2b in haemogenic endothelium. To investigate a potential role of gata2a in HSC development, we compared the expression of runx1, the key marker of HE in zebrafish,4 in wild-type and gata2aΔi4/i4 embryos. Quantitative in situ hybridization (ISH) analysis34 showed that runx1 expression was decreased in gata2aΔi4/i4 embryos at 24hpf (Supplementary Fig. 4a–c) and 28hpf compared to wild-type siblings (Fig. 3a–c). Further analysis in kdrl:GFP^+ ECs showed that this decrease in runx1 expression was already detectable at 23hpf in gata2aΔi4/i4 mutants (Fig. 3d), at the onset of its expression in HE.33 Thus, deletion of the gata2a-i4 enhancer results in impaired runx1 expression in the early stages of HE programming. This correlates well with decreased runx1 expression levels in +9.5−/− mouse AGM explants,19 further supporting the critical evolutionary role of the intronic enhancer of Gata2 in HSC specification.
Next, we tested whether Gata2a could act upstream of gata2b by measuring its expression in gata2a Δi4/Δi4 embryos. Quantitation of the ISH signal showed that gata2b expression was decreased in gata2a Δi4/Δi4 embryos compared to wild-type siblings at 26hpf (Supplementary Fig. 4d) and 28hpf (Fig.3e–g), but recovered to wild-type levels by 30hpf (Supplementary Fig. 4e). Accordingly, kdrl:GFP+; gata2a Δi4/Δi4 cells express significantly lower levels of gata2b mRNA than the wild-type kdrl:GFP+ endothelial population at 23hpf, but not at 30hpf (Fig. 3h). These data suggest that endothelial expression of gata2a
is required upstream of gata2b and runx1 for the proper specification of HE, uncovering a previously unrecognized role for Gata2a in definitive haematopoiesis.

Recovery of embryonic HSC activity in gata2aΔi4/Δi4 mutants. The qPCR analysis in sorted kdrl:GFP+; gata2aΔi4/Δi4 and wild-type kdrl:GFP+ ECs (Fig. 3d) already suggested a recovery of runx1 expression from 30hpf. Of note, the kdrl:GFP+ population likely includes the kdrl+, runx1-expressing erythromyeloid progenitors (EMP) located in the caudal region. This region was not included in the quantification of ISH but cannot be separated by sorting for kdrl:GFP+ and could thus explain the discrepancy between image...
hybridization (ventral view) for hybridization for cmyb
Whitney test). Error bars: median ± SD.

...of wild-type (homozygous mutant embryos at 4dpf. This effect is absent at 5dpf. (i)
wild-type and signi...he thymus37 showed that around half of the gata2aΔi4/Δi4 mutants had reduced gata2a expression at 4dpf compared to wild-type (Fig.4).

At 48hpf, the expression of runx1 in the DA showed no significant difference between gata2aΔi4/Δi4 mutants and wild-type controls (Fig. 4a, b). These data suggest that the decrease of runx1 expression at early stages of HE programming in gata2aΔi4/Δi4 mutants is transient and recovers by 2dpf. Indeed, analysis of the HSPC marker cmyb11 in the CHT at 4dpf showed no differences between gata2aΔi4/Δi4 and wild-type larvae (Fig. 4c, d). Expression of the T-cell progenitor marker rag1 in the thymus37 showed that around half of the gata2aΔi4/Δi4 larvae had reduced rag1 expression at 4dpf compared to wild-type (Fig. 4e, f). This effect was absent at 5dpf (Fig. 4g, h), suggesting that HSPC activity was normal in gata2aΔi4/Δi4 mutants from 4dpf onwards. Next, we crossed the gata2aΔi4/Δi4 mutants to Tg (itga2b-GFP) transgenics, where itga2b-GFPhigh and itga2b-GFP low cells in the CHT mark thrombocytes and HSPCs, respectively9,38. Our analysis revealed no difference in itga2b-GFPhigh HSPC or itga2b-GFP high thromocyte numbers in the CHT region at 5dpf between wild-type and gata2aΔi4/Δi4 mutants (Fig.4i–l). Taken together, our data suggest that endothelial gata2a expression mediated by the i4 enhancer is required for the initial expression of gata2b and runx1 in the HE but largely dispensable after 2dpf.

**Notch recovers haematopoiesis in gata2aΔi4/Δi4 mutants.** The recovery of gata2b expression by 30hpf (Fig. 3h, Supplementary Fig. 4e) coincides temporally with the observed decrease in gata2a in wild-type endothelial cells (Fig. 2h). Thus, we reasoned that other regulators of gata2b might compensate for the lack of endothelial gata2a in gata2aΔi4/Δi4 mutants and thus lead to a recovery of the initial haematopoietic phenotype. Therefore, we...
investigated whether the loss of *gata2b* in *gata2a*Δi4/Δi4 background resulted in a more severe haematopoietic phenotype than observed in the *gata2a*Δi4/Δi4 mutants. For this, we injected *gata2a*Δi4/Δi4 and wild-type controls with a suboptimal amount (7.5 ng) of a *gata2b* morpholino oligonucleotide (MO). Quantitative ISH analysis confirmed that this amount of *gata2b* MO had no effect on *runx1* expression at 32hpf (Fig. 5a, b). As expected, *runx1* expression in *gata2a*Δi4/Δi4 embryos was significantly reduced compared to wild-type siblings (Fig. 5a, b). *Gata2b* knockdown in *gata2a*Δi4/Δi4 embryos further reduced *runx1* expression (Fig. 5a, b). To test whether this stronger reduction of *runx1* at 32hpf affected later stages of embryonic haematopoiesis, we assessed *cmyb* expression in the CHT at 4dpf (Fig. 5c). We scored *cmyb* expression levels as ‘wild-type’ or ‘reduced’ and found that the ‘reduced’ embryos were substantially overrepresented in the *gata2a*Δi4/Δi4 mutants injected with the *gata2b* MO, compared to wild-type fish and non-injected *gata2a*Δi4/Δi4 siblings (Fig. 5c).

To verify whether Gata2b is required for definitive haematopoiesis downstream of Gata2a, we generated a frameshift truncating...
mutant for Gata2b and increased gata2aΔi4/+; gata2b Δi4/Δi4 adults to investigate cmyb expression at 33hpf in their progeny. Gata2bΔi4/Δi4 mutants showed a more severe decrease in cmyb expression than gata2aΔi4/Δi4 mutants (Supplementary Fig. 4f–i). Double gata2bΔi4/Δi4; gata2aΔi4/Δi4 mutants showed no further reduction in cmyb expression compared to gata2bΔi4/Δi4 mutants, suggesting that Gata2a was not sufficient to drive cmyb expression in HE in the absence of Gata2b (Supplementary Fig. 4f–i). Taken together, we conclude that Gata2b is regulated by Gata2a and is required for definitive haematopoiesis.

Next, we tested whether forced ectopic expression of gata2b was sufficient to speed up the haematopoietic recovery of gata2aΔi4/Δi4 embryos. Thus, we overexpressed gata2b under the control of the gata2a-i4-450bp enhancer in wild-type and gata2aΔi4/Δi4 mutant embryos and measured runx1 expression at 28hpf in the DA. Gata2aΔi4/Δi4 embryos showed a significant decrease in runx1 expression in comparison to wild-type (Fig. 5d). Ectopic expression of gata2b under the gata2a-i4 enhancer significantly increased runx1 expression in wild-type and mutants (Fig. 5d). Importantly, it was sufficient to bring the runx1 expression levels in the mutants up to the levels detected in uninjected wild-type embryos (Fig. 5d), demonstrating that gata2b alone was sufficient to drive runx1 expression and drive the haematopoietic recovery in gata2aΔi4/Δi4 mutants. Thus, gata2b can recover the definitive haematopoietic programme in the absence of endothelial gata2a.

Because the expression of gata2b is regulated by Notch signalling, we investigated whether inhibition of Notch would also prevent the haematopoietic recovery of gata2aΔi4/Δi4 embryos. For this, we used the Notch inhibitor DAPM and titrated it down to a suboptimal dose (25 μM) that did not significantly affect runx1 expression at 30hpf in wild-type embryos (Supplementary Fig. 5a). This dose induced a small but measurable decrease in gata2b expression in DAPM-treated embryos while higher doses had a more robust effect (Supplementary Fig. 5b). Next, we treated wild-type and gata2aΔi4/Δi4 mutant embryos with DAPM and measured runx1 expression in the DA at 36hpf (Fig. 5e, f). Suboptimal DAPM treatment did not affect runx1 expression in wild-type embryos (Fig. 5e, f), but gata2aΔi4/Δi4 mutants showed lower runx1 levels and DAPM treatment further reduced runx1 expression (Fig. 5e, f). Treatment with 25 μM DAPM did not significantly affect gata2b expression at 36hpf in either wild-type or gata2aΔi4/Δi4 mutant embryos (Supplementary Fig. 5c). These experiments suggested that Notch signalling alone may be sufficient to rescue expression of HE markers in gata2aΔi4/Δi4 mutants. Indeed, ectopic activation of Notch signalling in endothelium using a filia-NICD:GFP construct40 led to increased runx1 and gata2b expression in wild-type embryos (Supplementary Fig. 5d, e). When overexpressed in gata2aΔi4/Δi4 mutants, filia-NICD:GFP rescued runx1 expression to near wild-type levels at 26hpf, whereas gata2b was increased beyond normal levels independently of the genotype (Supplementary Fig. 5f, g). Taken together, these experiments confirm that Notch activity regulates runx1 and gata2b in HE and is sufficient to drive haematopoietic recovery in gata2aΔi4/Δi4 mutants. Thus, we conclude that HE programming requires two independent inputs on runx1 and gata2b expression: one from Gata2a, driven in ECs by the i4 enhancer, and the other from Notch signalling, necessary and sufficient to drive HE programming even in the absence of gata2a.

Impaired haematopoiesis in adult gata2aΔi4/Δi4 mutants. To investigate whether Gata2a plays a role in adult haematopoiesis, we first asked whether the gata2a-i4-1.1 kbGFP reporter was active in haematopoietic cells in the adult. Whole kidney marrow (WKM) cells isolated from the transgenic fish showed that the i4 enhancer is active in haematopoietic cells previously defined by flow cytometry41 as progenitors, lymphoid + HSPC (containing the HSPC) and myeloid cells (Supplementary Fig. 6a–c). Accordingly, single cell transcriptional profiling showed higher levels of gata2a in HSPCs, progenitors, neutrophils and thrombocytes (Supplementary Fig. 6d–f). Consistent with this notion, we observed a high incidence of infections and heart oedemas in gata2aΔi4/Δi4 adult fish, with over 25% suffering from one of these defects by 6 months of age, compared to <1% of wild-type fish (Fig. 6a–c). The heart oedemas and the infections are suggestive of lymphatic defects and immune deficiency as observed in human patients bearing genetic GATA2 haploinsufficiency syndromes such as MonoMAC syndrome13. Notably, around 10% of MonoMAC syndrome patients show mutations in the homologous enhancer region of GATA212.14.

Next, we counted the total number of haematopoietic cells in wild-type and gata2aΔi4/Δi4 mutant WKM (Fig. 6d–f). To avoid any confounding effects in our analysis, we compared wild-type to gata2aΔi4/Δi4 mutants without overt signs of infection. The gata2aΔi4/Δi4 mutants showed a ~2-fold decrease in the total number of haematopoietic cells in the WKM (Fig. 6d–f). In

Fig. 5 Gata2b and Notch signalling are sufficient to recover haematopoietic markers in gata2aΔi4/Δi4 mutants. a Expression of runx1 in HE at 23hpf in wild-type (wt), gata2a MO-injected (7.5 ng) wt embryos, gata2aΔi4/Δi4 mutants and gata2b MO-injected (7.5 ng) gata2aΔi4/Δi4 mutants. b Quantification of the runx1 signal in situ hybridization (ISH) signal in wt, gata2b morphants, gata2aΔi4/Δi4 mutants and gata2aΔi4/Δi4 mutants injected with gata2b MO. runx1 expression is decreased in gata2aΔi4/Δi4 mutants (μmut = 41.9, μwt = 28.9, F = 44.641, d.f. = 3, 62.3; p < 0.001). Gata2b MO knockdown significantly decreases runx1 in the DA of gata2aΔi4/Δi4 mutants (μmut = 28.9, μmut+MO = 17.2), but not wt embryos at 33hpf (μwt = 41.9, μMO = 40.1; p = 0.89, Games-Howell post-hoc test, Welch’s ANOVA). n = 27, wt; n = 27, gata2aΔi4/Δi4; n = 33, wt + gata2b MO; n = 32, gata2aΔi4/Δi4 + gata2b MO. c Scoring cmyb expression at 4dpf in wt, gata2aΔi4/Δi4 mutants and gata2aΔi4/Δi4 mutant embryos injected with gata2b MO as wt (blue) or reduced (red). Gata2b MO knockdown (7.5 ng) inhibits the haematopoietic recovery of gata2aΔi4/Δi4 mutants. (X2 = 18.784, d.f. = 2, p < 0.001). d Quantification of the runx1 ISH signal, from 28hpf wt embryos (blue), gata2aΔi4/Δi4 mutants (red) and their siblings injected with a gata2a-i4-450bp:gata2b construct (shaded blue and red). Ectopic expression of gata2b increases runx1 expression in the HE of wt embryos (μmut = 38.8, μmut-gata2b = 53.4; p < 0.01) and rescues runx1 expression in the DA of gata2aΔi4/Δi4 mutants to wt levels (μmut = 17.9, μmut-gata2b = 33.2; p < 0.001, μmut-gata2b = 33.2; p < 0.01, Tukey HSD post-hoc test). n = 25, wt; n = 33, gata2aΔi4/Δi4; n = 18, wt + gata2a-i4-450bp:gata2b construct; n = 17, gata2aΔi4/Δi4 + gata2b-i4-450bp:gata2b construct. e Quantification of the runx1 ISH signal at 36hpf in embryos treated with a suboptimal dose (25 μM) of the Notch inhibitor DAPM. 25 μM DAPM showed no effect on runx1 expression in wt compared to DMSO-treated embryos (μDMSO = 40.5, μDAPM = 38; p = 0.735, Tukey HSD post-hoc test). DMSO-treated gata2aΔi4/Δi4 mutants show a decrease in runx1 expression (μmut+DMSO = 31.5; F = 25.774, d.f. = 3, 91; ANOVA). DAPM treatment significantly reduced runx1 expression in the DA of gata2aΔi4/Δi4 mutants (μmut+DMSO = 31.5, μmut+DAPM = 19.4). n = 27, wt + DMSO; n = 20, gata2aΔi4/Δi4 + DMSO; n = 30, wt + DAPM; n = 20, gata2aΔi4/Δi4 + DAPM. f Representative images of the average runx1 expression at 36hpf in wt and gata2aΔi4/Δi4 mutants treated with 25 μM DAPM. Error bars: mean ± SD. **p < 0.01; ***p < 0.001. See also Supplementary Fig. 5.
addition, neutrophils were similarly reduced (Fig. 6g), another characteristic in common with MonoMAC syndrome patients. Lastly, kidney marrow smears of ten 9-month old gata2aΔi4/Δi4 mutants were assessed. One of the ten mutants showed an excess of immature myeloid blast cells in the WKM (>98%) and only minor erythrocyte differentiation (Fig. 6h, i). The presence of excess blasts is usually an indication of AML in humans. Together these data strongly suggest that the i4 enhancer is a critical driver of gata2a in adult haematopoietic cells. The enhancer deletion in gata2aΔi4/Δi4 mutants leads to a hypocellular WKM characteristic in common with MonoMAC syndrome patients and neutropenia, strongly suggestive of marrow failure, a hallmark of disease progression in Gata2 deficiency syndromes.

**Discussion**

The sub-functionalisation of the Gata2 paralogues in zebrafish provided an opportunity to unpick the different roles of Gata2 in the multi-step process of definitive haematopoiesis. Here we have investigated the conservation of the Gata2 +9.5 enhancer and identified a homologous region in intron 4 of the zebrafish gata2a locus (gata2a-i4) that is not present in the gata2b locus. The zebrafish gata2a-i4 enhancer, like the mouse enhancer, is sufficient to drive pan-endothelial expression of GFP and necessary for endothelial expression of gata2a (Figs. 1 and 2). We traced the activity of the i4 enhancer back to the PLM, the source of precursors of endothelium and HSCs. This degree of sequence and functional conservation of the i4 enhancer led us to hypothesize that Gata2a might play a role in definitive haematopoiesis. Indeed, homozygous deletion of the i4 enhancer (gata2aΔi4/Δi4) allowed us to uncover a previously unknown function of Gata2a in regulating the initial expression of *runx1* and *gata2b* in HE. Although *cmyb* expression in HE was decreased in gata2aΔi4/Δi4 mutants, it was more severely reduced in gata2b−/− mutants, suggesting that Gata2b is more important for cmyb regulation than Gata2a. However, both Gata2 orthologues regulate gene expression in the HE before the first reported EHT events at 34 hpf.

Gata2 binds to the +9.5 enhancer to maintain its own expression in endothelial and haematopoietic cells. In zebrafish, it is likely that Gata2a binds the GATA motifs in the i4 enhancer and loss of gata2a in the endothelium of gata2aΔi4/Δi4 mutants (Fig. 2) seems to support this view. Interestingly, we detected a small region in intron 4 of the gata2b locus that was
not identified as a peak in our ATACSeq experiment but is conserved in some fish species (Supplementary Fig. 1a) and thus could potentially represent a divergent gata2a intronic enhancer. We speculate that the positive autoregulation of Gata2 was likely retained by both gata2 orthologues in zebrafish, but this possibility remains to be investigated.

The gata2aΔ4/Δ4 mutants recovered from the early defects in HE programming and displayed normal expression levels of cmyb in the CHT at 4dpf and rga1 in the thymus at 5dpf, used as indicators of the definitive haematopoietic programme11. We hypothesized that this could be due to the presence of the two homologues of Gata2 in zebrafish28,30, despite Gata2a and Gata2b proteins being only 50% identical11. Indeed, forced expression of gata2b under the gata2aΔ4 enhancer rescued DA expression of runx1 in the gata2aΔ4/Δ4 mutants to wild-type levels and suboptimal depletion of gata2b in the gata2aΔ4/Δ4 mutants resulted in more severe reduction in cmyb expression in the CHT by 4dpf (Fig. 4). In addition, we demonstrated that Notch signalling, a known regulator of gata2b expression31, is sufficient to rescue the initial HE programming defect induced by deletion of the gata2aΔ−/+ or gata2aΔ−/− alleles. Thus, gata2aΔ−/− and gata2aΔ−/+ cells are expected to be more competent for further populations upstream of runx1 and gata2b independently of Notch to initiate HE programming. The regulation of gata2b by Gata2a is transient, and the timing largely coincides with the natural decrease in endothelial expression of gata2a by 30hpf. After this stage, endothelial Notch signalling takes over the regulation of runx1 and gata2b expression, acting as a fail-safe mechanism that buffers against fluctuations in the system caused by loss of one or more of the initial inputs (in this case, Gata2a).

Despite the apparent haematopoietic recovery, we observed a high incidence of infections and oedema in gata2aΔ4/Δ4 adults, and a striking decrease in the number of haematopoietic cells in the WK. The decrease in haematopoietic cells in particular is reminiscent of the loss of proliferative potential of haematopoietic Gata2+/− heterozygous cells in the mouse17,45. This raises the possibility that in zebrafish the gata2a and gata2b paralogues may function as two Gata2 ‘alleles’ that together regulate the haematopoietic output of the WK. This will be addressed by comparing the adult phenotypes of gata2aΔ1/Δ4 and gata2bΔ1/− mutants.

Taken together, our initial characterization of WK shows that gata2aΔ1/Δ4 mutants present a phenotype consistent with Gata2 deficiency syndromes in humans brought about by GATA2 haploinsufficiency12,26. Strikingly, about 10% of all MonoMAC patients show mutations in the conserved +9.5 enhancer12,14, the corresponding regulatory element to the i4 enhancer. The i4 enhancer is active in the lymphoid + HSPC fraction that contains the HSC activity66, in the progenitor cells and in the myeloid fraction that contains eosinophils, previously identified as expressing high levels of a gata2a-GFP BAC transgenic reporter41. Thus, it is likely that gata2aΔ1/Δ4 adult fish show lineage-specific differentiation defects. Further characterization of the gata2aΔ1/Δ4 mutants will uncover which haematopoietic cells are most affected by the loss of i4 enhancer activity and how Gata2a regulates haematopoietic output, thus establishing a zebrafish animal model for human diseases linked to Gata2a haploinsufficiency.

Methods

Maintenance of zebrafish. Zebrafish (Danio rerio) were maintained in flowing system water at 28.5 °C, conductivity 450–550 μs and pH 7.0 ± 0.5 as described46. Fish suffering from infections or heart oedemas were culled according to Schedule 1 of the Animals (Scientific Procedures) Act 1986. Eggs were collected by natural mating. Embryos were grown at 24–32 °C in E3 medium with methylene blue and staged according to morphological features corresponding to respective age in hours or days post-fertilization (hpf or dpf, respectively). Published strains used in this work were wild-type (wtCC1), Tg(−6.0gata2a:EGFP)44,48 and Tg(kdr:GFP)78,43,48 animals were used at embryonic and larval stages and as adults (male and female) as specified in the figures. All animal experiments were approved by the relevant University of Oxford, University of Birmingham and Erasmus University ethics committees.

ATAC-seq. Tg(kdr:GFP)78,43 embryoids were dissociated for FACs at 26–27hpf to collect kdr+ and kdr− cell populations (40,000–50,000 cells each). They were processed for ATAC library preparation using optimised standard protocol27. Briefly, after sorting into Hanks’ solution (1XHBS, 0.25% BSA, 10 mM HEPES pH 7.4), the cells were spun down at 500 × g at 4 °C, washed with 50 μl cold Lysis Buffer (10 mM Tris-HCl, 10 mM NaCl, 3 mM MgCl2, 0.1% IGEPEAL, pH 7.4). The nuclei were pelleted for 10 min. at 500 × g at 4 °C and resuspended in the TD Buffer with Tn5 Transposase (Illumina), scaling the amounts of reagents accordingly to the number of sorted cells. The transposition reaction lasted 30 min. at 37 °C. The DNA was purified with PCR Purification MinElute Kit (QIAGEN). In parallel, transposase-unreated genomic DNA from kdr+ cells was purified with the DNeasy® Blood & Tissue Kit (QIAGEN).

The samples were amplified with appropriate Customized Nextera primers27 in NEB Next-High Fidelity 2x PCR Master Mix (NEB). The libraries were purified with the Agencourt AMPure XP purification MinElute Kit (QIAGEN). The quality of each library was verified using D1000 ScreenTape System (Agilent). Four biological replicas of the libraries were quantified with the KAPA Library Quantiﬁcation Kit for Illumina® platforms (KAPA Biosystems). The libraries were pooled (including the Tn5-untreated control), fragmented and endoPurify (Illumina) and the size distribution was measured with the HighSlex 4000 (Wellcome Trust Centre for Human Genetics). Raw sequenced reads were checked for base qualities, trimmed where 20% of the bases were below quality score 20, and filtered to exclude adapters using Trimmomatic (Version 0.32) and mapped to Zv9 reference genome (comprising 14,612 genes)49 using BWA with default parameters. The reads were visualised using UCSC Genome Browser (http://genome-euro.ucsc.edu)50. The eight data sets were analysed with Principal Component Analysis (PCA) to identify outliers. Correlation among kdr+GFP and kdr−GFP samples was assessed with a tree map. The peaks were called for each sample using the Tn5-untreated control as input. We identified the common peaks between replicates and then used DistBed (edgeR method) to identify differential peaks between kdr+GFP and kdr−GFP samples (Supplementary Table 2). The threshold for differential peaks was p < 0.05.

Generation of transgenic and mutant zebrafish lines. Genomic regions containing the identified 150bp-long gata2aΔ−/+ enhancer flanked by ±500 bp (i4−1.1 kb) ±150 bp (i4−450bp) were amplified from wild-type zebrafish genomic DNA with NEB Phusion® polymerase (see Supplementary Table 1 for primer sequences) and inserted upstream of Elb minimal promoter (Elb:GFP)19 into Tol2 transposon (Addgene plasmid #3784551) with Gateway® cloning technology (Life Technologies) following the manufacturer’s protocol. One-cell zebrafish embryos were injected with 1 nl of an injection mix, containing 50 pg gata2aΔ−/+ Elb-GFP: Tol2 construct DNA + 30 pg tol2 transposable mRNAs,18 Transfected founders (Tg(gata2aΔ−/+ or gata2bΔ−/+ or gata2bΔ−/−)/+ Elb-GFP) were selected by a wide-field fluorescent microscope and outbred to wild-type fish. Carriers of monoallelic insertions were detected by the Mendelian distribution of 50% fluorescent offspring coming from wild-type outcrosses. These transgenics were then inbred to homozygosity.

To generate the i4 deletion mutant, we identified potential sgRNA target sites flanking the 150 bp conserved region within intron 4 of the gata2a locus (see Fig. 1a, Supplementary Fig 3a). sgRNAs were designed with the CRISPR design tool (http://crispr.mit.edu/; see Supplementary Table 1 for sequences) and prepared as described12. To reduce potential off-target effects of CRISPR/Cas9, we utilized the D10A ‘nickase’ version of Cas9 nuclease25, together with two pairs of sgRNAs flanking the enhancer (Supplementary Table 1, Supplementary Fig. 3a, b). We isolated two mutant alleles with deletions of 215 bp (A78−292) and 231 bp (A73–303) (Supplementary Fig. 3b). Both deletions included the highly conserved E-box, Ets and GATA transcription factor binding sites (Supplementary Fig. 3b). A further embryo was selected for further experiments. The adult zebrafish were viable and fertile as heterozygous (gata2aΔ1/4+) or homozygous (gata2aΔ1/Δ4) mutants. To unambiguously genotype wild types, heterozygotes and homozygous mutants, we designed a strategy consisting of two PCR primer pairs (Supplementary Fig. 3a, c). One primer pair flanked the whole region, producing a 600 bp wild-type band and 369 bp mutant band. In the second primer pair, one of the primers was designed to bind within the deleted region, only giving a 367 bp band in the presence of the wild-type allele (Supplementary Fig. 3c).

To generate the gata2bΔ mutant we designed a CRISPR/Cas9 strategy for a frameshift truncating mutation in exon 3 deleting both zinc fingers. sgRNAs were designed as described above and guides were prepared according to Gagnon et al.54 with minor adjustments. Guide RNAs were generated using the Agilent SureGuide gRNA Synthesis Kit, Cat# 5190–7706. Cas9 protein (IDT) and guide were allowed to form ribonucleaseprotein structures (RNPs) at RT and injected in 1 cell stage oocytes. 8 embryos were selected at 24 hpf and lysed for DNA isolation.

Heteroduplex PCR analysis was performed to test guide functionality and the other embryos from the injection were allowed to grow up. To aid future genotyping we selected mutants by screening F1 for a PCR detectable integration or deletion in
exon 3. Sequence verification showed that founder 3 had a 28 nt integration resulting in a frameshift truncating mutation leading to 3 new STOP codons in the third exon of gata2a. This was confirmed by Sanger sequencing performed on non-transgenic embryos. The i4 mutants resulted in a frameshift truncating mutation leading to 3 new STOP codons in the gata2a expression levels of the technical triplicates for each gene of interest. This value (ΔCt) was then converted to a ratio relative to the housekeeping genes with the formula 2−ΔΔCt.

Fluidigm Biomark qRT-PCR. To quantify the differences in gata2a expression between wild-type and mutant ECs, we crossed homozygous gata2aΔ/Δ mice with Tg(kdrGFP) transgensics to generate Tg(kdrGFP); gata2aΔ/Δ embryos. These fish, along with non-mutant Tg(kdrGFP), were used for FACS-mediated isolation of kdrGFP+ and kdrGFP− cells to quantitatively compare mRNA expression levels of the endothelial and non-endothelial cells of wild-type and gata2aΔ/Δ embryos, using the Fluidigm Biomark® qRT-PCR platform. Briefly, 1 ng RNA from FACS-sorted cells was used for Specific Target Amplification in a 10 µl reaction with the following reagents: 5 µl 2xBuffer and 1.2 µl enzyme mix (Fluidigm SuperScript III One-Step Kit with RNase-Free DNase, 0.1 µl SUPERase In RNase Inhibitor, 0.1 µl TE buffer (Invitrogen), 2.5 µl 0.2x TaqMan® assay mix (see Supplementary Table 2 for the details of TaqMan® assays). The reactions were run on 7500 Fast Real-Time PCR System (Applied Biosystems) and the results were analysed with the accompanying software. No-template controls were run on each plate for each primer pair. Each reaction was validated with the melt curve analysis. The baseline values were calculated automatically for each reaction. The threshold values were manually set to be equal for all the reactions run on one plate, within the linear phase of exponential amplification. The relative mRNA levels in each sample were calculated by subtracting the geometric mean of Ct values for housekeeping genes eef1a1l1 and tbc1d1 from the Ct values for the gene of interest. This value (ΔCt) was then converted to a ratio relative to the housekeeping genes with the formula 2−ΔΔCt.

Fluorescence-activated cell sorting (FACS). Approximately 100 embryos at the required stage were collected in Low Binding 8xSeal® Microcentrifuge Tubes (Sorenson) and pre-homogenized by pipetting up and down in 500 µl Deyolking Buffer (116 mM NaCl, 2.9 mM KCl, 5 mM HEPES, 1 mM EDTA). They were spun down for 1 min. at 500 g and incubated for 15 min. at 30 °C in Trypsin + Collagenase Solution (1xHBSS, 0.05% Gibco® Trypsin + EDTA (Life Technologies®), 20 mg/ml collagenase (Sigma)). During that time, they were homogenized by pipetting up and down every 3 min. The lysis was stopped by adding 50 µl folic acid serum and 650 µl filter-sterilized Hanks’ solution (1xHBSS, 0.25% BSA, 10 mM HEPES pH8). The cells were rinsed with 1 ml Hanks’ solution and passed through a 40 µm cell strainer (Falcon®). They were resuspended in ~400 µl Hanks’ solution with 1:10,000 Hoechst 33258 (Molecular Probes) and transferred to a 5 ml polystyrene round bottom tube for FACS sorting. The cells were sorted on FACSAria Fusion sorter by Kevin Clark (MRC UK MIMM FACs Facility). The gates of GFP (488–530) and DiD (561–582) channels were set with reference to samples derived from non-transgenic embryos. The fluorescence readouts were compensated when necessary. For ATAC-seq library preparation, the cells were sorted into Hank’s solution. For RNA isolation, the cells were sorted directly into RLT Plus lysis buffer (Qiagen). The cells were provided with the mildly alkaline Phosphatase and the Micro Plus kit (QIAGEN), according to the accompanying protocol. The RNA was quantified and its quality assessed with the use of Agilent RNA 6000 Pico kit. All RNA samples were stored at −80 °C.

SYBR® Green qRT-PCR. 3 µl of the cDNA diluted in H2O were used for technical triplicate qRT-PCR reactions of 20 µl containing the Fast SYBR® Green Master Mix (Thermo Fisher Scientific) and probe and appropriate primer pair (see Supplementary Table 1). The reactions were run on 7500 Fast Real-Time PCR System (Applied Biosystems) and the results were analysed with the accompanying software. No-template controls were run on each plate for each primer pair. Each reaction was validated with the melt curve analysis. The baseline values were calculated automatically for each reaction. The threshold values were manually set to be equal for all the reactions run on one plate, within the linear phase of exponential amplification. The relative mRNA levels in each sample were calculated by subtracting the geometric mean of Ct values for housekeeping genes eef1a1l1 and tbc1d1 from the Ct values of the technical triplicates for each gene of interest. This value (ΔCt) was then converted to a ratio relative to the housekeeping genes with the formula 2−ΔΔCt.

May-Grunwald and Wright-Giemsa staining. Cell staining with May-Grunwald (MG) stain (Sigma MG500) and Giemsa (GIEMSA STAIN, FLUKA 48990) was performed on haematoxytoxic cell samples. After cytospin, slides are allowed to air-dry and were stained with May-Grunwald for 5 min. at room temperature with a 1:11 dilution of MG-distilled water. Next, slides were drained and stained with a 1:9 dilution of Giemsa-distilled water solution for 30 min at room temperature. Excess solution was drained and removed by further washes in distilled water. Finally, the slides were air-dried and mounted in DPX (65522, Sigma) for imaging.

Whole-mount in situ hybridization and immunohistochemistry. Whole-mount in situ hybridization (ISH) was carried out as described previously, using probes for kdr, runx1, cmrb, gata2a, gata2b, ragg,37,69,90 and gf (Supplementary Table 1). For conventional ISH embryos were processed, imaged and the ISH signal quantified as described. Briefly, the pixel intensity values were assessed for normal distribution with a Q-Q plot and transformed when necessary. Mean values (µ) and experimental group were compared with 2-tailed independent-samples t-tests or with ANOVA with 95% confidence levels, testing for the equality of variances with a Levene’s or Brown-Forsythe test and applying the Welch correction when necessary. For ANOVA, differences between each two groups were assessed with either Tukey’s post-hoc test (for equal variances) or with Games-Howell test (for unequal variances). For gene expression levels, the IBM® SPSS® Statistics (version 22) or GraphPad Prism 8.0 package were used.

For the analysis of cmrb expression in the CHT at 4dpf, the embryos scored as ‘high’ or ‘low’ were tested for equal distribution between morphants and un injected controls or among wild-type, heterozygous and mutant genotypes with contingency Chi-squared tests, applying Continuity Correction for 2 × 2 tables, using IBM® SPSS® Statistics (version 22).

For fluorescence ISH (FISH) combined with immunohistochemistry, ISH was performed first following the general whole-mount in situ hybridisation protocol. The signal was developed with SIGMAFAST Fast Red TR/Naphthol, the embryos rinsed in phosphate-buffered saline with tween20 (PBST) and directly processed for immunohistochemistry. Embryos were blocked in blocking buffer (5% goat serum/0.3% Triton X-100 in PBST) for 1 h at RT before incubated with primary antibody against GFP (rabbit, 1:500, Molecular Probes), diluted in blocking buffer overnight with either Tukey’s post-hoc test (for equal variances) or with Games-Howell test (for unequal variances). For gene expression levels, the IBM® SPSS® Statistics (version 22) or GraphPad Prism 8.0 package were used.

Fluorescence microscopy and cell counting. For widefield fluorescence microscopy, live embryos were anaesthetised with 150 mg/ml MS222 and mounted in 3% methylcellulose and imaged on an AxioImager M12 stereomicroscope (Zeiss) equipped with a Zeiss AxioCam MrM. To count itg2b-GFP+ and itg2b-GFP− cells in the CHT, Tg(itg2b:GFPLcb (kdr:Cherry)); gata2aΔ/Δ+ animals were incased and grown in E3 medium supplemented with PTU to prevent pigment formation. At 5dpf, the larvae were anaesthetised with MS222 and the tail was cut off and fixed for 1 h at room temperature in 4% PFA. Next, the tails were mounted on 35 mm glass bottomed dishes (MARIETK) in 1% low melt agarose and imaged using a ×40 oil objective on an LSM880 confocal microscope (Zeiss). Cells in the CHT region were counted manually on Z-stacks as itg2b:GFPLow (HSpc) or itg2b:GFPHigh (thrombocytes). Genomic DNA from the heads was extracted and used for quantification, as described for 2-tailed paired-tailed-samples t-tests with 95% confidence levels, using a Mann-Whitney test for non-parametric distribution. The graphs were generated using GraphPad Prism 8.3.0 and show medians ± SD.

Statistics and reproducibility. Data were analysed using either IBM® SPSS® Statistics (version 22) or GraphPad Prism software (v8.02). In situ quantification data was analysed with 2-tailed independent-samples t-tests or with ANOVA with 95% confidence levels, testing for the equality of variances with a Levene’s or Brown-Forsythe test and applying the Welch correction when necessary. For ANOVA, differences between each two groups were assessed with either Tukey’s post-hoc test (for equal variances) or with Games-Howell test (for unequal variances). For gene expression levels, the IBM® SPSS® Statistics (version 22) or GraphPad Prism 8.0 package were used.

For fluorescence ISH (FISH) combined with immunohistochemistry, ISH was performed first following the general whole-mount in situ hybridisation protocol. The signal was developed with SIGMAFAST Fast Red TR/Naphthol, the embryos rinsed in phosphate-buffered saline with tween20 (PBST) and directly processed for immunohistochemistry. Embryos were blocked in blocking buffer (5% goat serum/0.3% Triton X-100 in PBST) for 1 h at RT before incubated with primary antibody against GFP (rabbit, 1:500, Molecular Probes), diluted in blocking buffer overnight with 4% PFA. Next, the tails were mounted on 35 mm glass bottomed dishes (MARIETK) in 1% low melt agarose and imaged using a ×40 oil objective on an LSM880 confocal microscope (Zeiss). Cells in the CHT region were counted manually on Z-stacks as itg2b:GFPLow (HSpc) or itg2b:GFPHigh (thrombocytes). Genomic DNA from the heads was extracted and used for quantification, as described for 2-tailed paired-tailed-samples t-tests with 95% confidence levels, using a Mann-Whitney test for non-parametric distribution. The graphs were generated using GraphPad Prism 8.3.0 and show medians ± SD.
samples t-tests with 95% confidence levels, using a Mann-Whitney test for non-parametric distribution. Gene expression data were analysed with 2-tailed paired-sample t-tests with 95% confidence levels. All experiments were repeated at least three times with similar results obtained; sample sizes are shown in the respective figure legends.

### Reporting summary
Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability
All data generated or analysed during this study (images, quantitation data in the form of graphs and ATACseq data) are included in this published article and its supplementary information files. The source data are available in Supplementary Data 1 and the list of the called ATACseq peaks is available in Supplementary Data 2. The ATACseq data was deposited in GEO (Accession number GSE143763).

Received: 25 April 2019; Accepted: 28 January 2020; Published online: 13 February 2020

### References
1. Ciau-Uitz, A., Monteiro, R., Kirmizitas, A. & Patient, R. Developmental hematopoiesis: ontology, genetic programming and conservation. *Exp. Hematol.* 42, 669–683 (2014).
2. Bonkhofer, F. et al. Blood stem cell-forming haemogenic endothelium in zebrafish derives from the arterial endothelium. *Nat. Commun.* 10, 3577 (2019).
3. Gritz, E. & Hirschi, K. K. Specification and function of hemogenic endothelium during embryogenesis. *Cell Mol. Life Sci.* 73, 1547–1567 (2016).
4. Swiers, G. et al. Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. *Nat. Commun.* 4, 2924 (2013).
5. Kissa, K. & Herbold, P. Blood stem cells emerge from arterial endothelium by a novel type of cell transition. *Nature* 464, 112–115 (2010).
6. Bertrand, J. Y. et al. Haematopoietic stem cells emerge directly from arterial endothelium during development. *Nature* 464, 108–111 (2010).
7. Casto, J. C. et al. In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature* 464, 116–120 (2010).
8. Kissa, K. et al. Live imaging of emerging hematopoietic stem cells and early thymus colonization. *Blood* 111, 1147–1156 (2008).
9. Hsu, A. P. et al. GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. *Blood* 121, 3830–3837 (2013). S3831-3837.
10. Sluder, G. et al. Transcription factor GATA-2. *Nat. Rev. Mol. Cell Biol.* 1397 (2016). quiz 1518.
11. Willett, C. E., Zapata, A. G., Hopkins, N. & Steiner, L. A. Expression of zebrafish rag genes during early development identifies the thymus. *Dev. Biol.* 182, 331–341 (1997).
12. Liu, H.-F. et al. Analysis of thrombocytosis development in CD41-GFP transgenic zebrafish. *Blood* 106, 3803–3810 (2005).
13. Walsh, D. M. et al. Naturally secreted oligomers of amyloid beta protein potently inhibit hippocampal long-term potentiation in vivo. *Nature* 416, 535–539 (2002).
14. Liu, Z. et al. Primary cilia regulate hematopoietic stem and progenitor cell specification through Notch signaling in zebrafish. *Nat. Commun.* 10, 1839 (2019).
15. Marion, A. et al. Transplantation and in vivo imaging of multilineage engraftment in zebrafish bloodless mutants. *Nat. Immunol.* 4, 1238–1246 (2003).
16. Athanasiadis, E. I. et al. Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. *Nat. Commun.* 8, 2045 (2017).
17. Macaulay, I. C. et al. Single-cell RNA-sequencing reveals a continuous spectrum of HSC differentiation in Hematopoietic Cells. *Cell Rep.* 14, 966–977 (2016).
18. Sanalkumar, R. et al. Mechanism governing a stem cell generating cis-regulatory element. *Proc. Natl Acad. Sci. USA* 111, E1091–1110 (2014).
19. Rodrigues, N. P. et al. Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. *Blood* 110, 477–484 (2005).
20. 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 (2011).
21. Westerfield, M. The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio). 5th edn, (Univ of Oregon Press, 2007).
22. Chi, N. C. et al. Foxn1 directly regulates tbx2b expression and aortoventricular canal formation. *Genes Dev.* 22, 734–739 (2008).
23. Howe, K. et al. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 496, 498–503 (2013).
24. Kent, W. J. et al. The human genome browser at UCSC. *Genome Res.* 12, 996–1006 (2002).
25. Birnbaum, R. Y. et al. Coding exons function as tissue-specific enhancers of nearby genes. *Genome Res.* 22, 1059–1068 (2012).
26. Ran, F. A. et al. Double nicking by RNA-guided CRISPR/Cas9 for enhanced genome editing specificity. *Cell* 154, 1380–1389 (2013).
27. Mali, P. et al. CAS9 transcriptional activators for target specificiency and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* 31, 833–838 (2013).
28. McGovern, J. A. et al. Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *Plos ONE* 9, e98186 (2014).
55. Xu, H. et al. Genome-wide identification of suitable zebrafish Danio rerio reference genes for normalization of gene expression data by RT-qPCR. J. Fish. Biol. 88, 2095–2110 (2016).

56. Hu, Y., Xie, S. & Yao, J. Identification of novel reference genes suitable for qRT-PCR normalization with respect to the Zebrafish developmental stage. PLoS ONE 11, e0149277 (2016).

57. Stachura, D. L. & Traver, D. Cellular dissection of zebrafish hematopoiesis. Methods Cell Biol. 133, 11–53 (2016).

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

59. Thompson, M. A. et al. The cloche and spadetail genes differentially affect hematopoiesis and vasculogenesis. Dev. Biol. 197, 248–269 (1998).

60. Monteiro, R. et al. Transforming growth factor beta drives hemogenic endothelium programming and the transition to hematopoietic stem cells. Dev. Cell 38, 358–370 (2016).

Acknowledgements
We thank the staff of the Biomedical Services Units (Oxford, Birmingham and Rotterdam) for fish husbandry. We thank Kevin Clark and Sally-Ann Clark from the WIMM flow cytometry facility for cell sorting. The flow cytometry facility is supported by the MRC HIU, MRC MHU (MC_UU_12009), NIHR Oxford BRC and John Fell Fund (131/030 and 101/517), the EPA fund (CF182 and CF170). We thank the Wolfson Imaging Centre Oxford for imaging. The Centre is supported by a Wolfson Foundation (grant 18272), and an MRC/BBSRC/EP/RC grant (MR/K015777X/1) to MICA – Nanoscopy Oxford. Both facilities were supported by the WIMM Strategic Alliance awards G0902418 and MC_UU_12035. We thank Fatma Kok and Douglas Vermijnen for critical reading of the manuscript. We thank Feng Liu for the generous gift of the fi1a-NICD/GFP construct. This research was supported by the British Heart Foundation (BHF Oxford CoRE and BHF IBF Fellowship FS13/50/30436 to R.M. and M.K.), by a Wellcome Trust Chromosome and Developmental Biology PhD Scholarship (WT102345/Z/13/Z to T.D.) and by MRC MHU programme (MC_UU_12009/8 to R.P.).

Author contributions
T.D., M.K., C.K., E.P. and R.M. designed the study. T.D., M.K., C.K., J.P-Z., K.G., C.B.M. and R.M. performed experiments and analyzed the data. J.P-Z., B.F. and K.G. performed experiments. R.R. performed the bioinformatics analyses. T.D. and R.M. wrote the paper and R.P., E.P. and R.M. edited the paper. R.P., E.P. and R.M. secured funding.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s42003-020-0798-3.

Correspondence and requests for materials should be addressed to R.M.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020