bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A zebrafish model for MonoMAC syndrome identifies an earlier role for gata2 in haemogenic endothelium programming and generation of haematopoietic stem cells

Tomasz Dobrzycki[1], Monika Krecsmarik[1,2], Cansu Koyunlar[4], Rossella Rispoli[1], Joke Peulen-Zink[4], Kirsten Gussinklo[4], Emma de Pater[4], Roger Patient[1,2] and Rui Monteiro[1,2,3,*]

1. MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford, OX3 9DS, United Kingdom
2. BHF Centre of Research Excellence, Oxford, United Kingdom
3. Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham, UK
4. Department of Hematology, Erasmus MC, Rotterdam, The Netherlands

*corresponding author:

Email: r.monteiro@bham.ac.uk

Keywords: gata2a, gata2b, haemogenic endothelium, haematopoietic stem cells, haploinsufficiency, bone marrow failure, acute myeloid leukemia

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Haematopoietic stem and progenitor cells (HSPCs) maintain the vertebrate blood system throughout life and their emergence from haemogenic endothelium (HE) is tightly regulated by transcription factors such as Gata2. Zebrafish have two orthologues of Gata2, *gata2a* and *gata2b*, the latter required for HSPC emergence. Here we deleted a conserved enhancer driving *gata2a* expression in endothelium (i4 enhancer) and showed that Gata2a is required for HE programming by regulating expression of *gata2b* and *runx1*. By 5 days, homozygous *gata2a*$^{\Delta i4/\Delta i4}$ larvae showed normal numbers of HSPCs, a recovery mediated by Notch signalling driving *gata2b* expression in HE. However, *gata2a*$^{\Delta i4/\Delta i4}$ adults showed lymphoedema, susceptibility to infections and marrow hypocellularity, consistent with bone marrow failure of MonoMAC syndrome patients. Thus, Gata2a is required for HE programming and haematopoiesis in the adult. Like MonoMAC syndrome patients, *gata2a*$^{\Delta i4/\Delta i4}$ mutants developed acute myeloid leukemia. These mutants will be invaluable to explore the pathophysiology of MonoMAC syndrome *in vivo*.

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Haematopoietic 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 are part of elaborate gene regulatory networks that coordinate differentiation, proliferation and survival of haematopoietic cells and ensure their levels are appropriate at all times throughout life. The only time during ontogenesis when HSCs are generated *de novo* is during embryonic development and misexpression of key transcription factors may lead to a catastrophic failure to produce HSCs or, alternatively, to haematopoietic disorders and eventually leukaemia. Therefore, understanding how transcription factors drive the haematopoietic process holds great clinical promise as it 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 (Ciau-Uitz et al., 2014). The specification of HSCs initiates in cells with arterial characteristics and proceeds through an endothelial intermediate, termed the haemogenic endothelium (HE) (Gritz and Hirschi, 2016). In zebrafish and other vertebrates, expression of *runx1* in the floor of the dorsal aorta defines the *bona fide* HE population (Burns et al., 2002; Gering and Patient, 2005; Kalev-Zylinska et al., 2002; Swiers et al., 2013). Haematopoietic stem and progenitor cells (HSPCs) emerge from the HE by a process termed endothelial-to-haematopoietic transition (EHT), both in zebrafish and in mice (Bertrand et al., 2010; Boisset et al., 2010; Eilken et al., 2009; Kissa and Herbomel, 2010; Lam et al., 2010). They first arise at around 34 hours post fertilisation (hpf) from the HE in the ventral wall of the DA (Kissa et al., 2008), the analogue of the mammalian AGM (Davidson and Zon, 2004). After EHT, the HSCs enter the bloodstream through the posterior cardinal vein (PCV) (Kissa et al., 2008) to colonise the caudal haematopoietic tissue (CHT), the zebrafish equivalent of the mammalian foetal liver (Murayama et al., 2006). Afterwards the HSCs migrate again within the bloodstream to colonise the kidney marrow (WKM) and thymus (Kissa et al., 2008), the final niche for HSCs, equivalent to the bone marrow in mammals (Ciau-Uitz et al., 2014).

Gata2 is a key haematopoietic transcription factor (TF) in development. *Gata2* knockout mice are embryonic lethal and die by E10.5 (Tsai et al., 1994). Conditional *Gata2* knockout under the control of the endothelial *VE-cad* promoter abolished the generation of intra-aortic clusters (de Pater et al.,

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2013), suggesting that Gata2 is required for HSPC formation. In addition, conditional deletion of Gata2 mediated by a *Vav*-Cre transgene demonstrated that Gata2 is also required for the maintenance and survival of HSCs in the foetal liver, after HE specification and HSC emergence (de Pater et al., 2013). Further studies in the mouse revealed a decrease in HSC numbers in *Gata2* heterozygous mutants, but also a dose-dependency of adult HSCs on Gata2 (Ling et al., 2004). In humans, *GATA2* haploinsufficiency leads to blood disorders, including MonoMAC syndrome (Monocytopenia, Mycobacterium avium complex) and myeloid dysplastic syndrome (MDS) (Hsu et al., 2013; Spinner et al., 2014). While its presentation is variable, MonoMAC syndrome patients always show cytopenias, ranging from mild to severe, and hypocellular bone marrow (Wlodarski et al., 2017; Wlodarski et al., 2016), highlighting the importance of maintaining Gata2 expression in adult haematopoietic cells. These patients are susceptible to mycobacterial and viral infections (Hsu et al., 2015) and have a propensity to develop myelodisplastic syndrome (MDS) and Acute Myeloid Leukaemia (AML), with a 75% prevalence and relatively early onset at age 20 (Wlodarski et al., 2017).

Gata2 expression in the endothelium is regulated by an enhancer element located in its intron 4, termed the +9.5 enhancer (Gao et al., 2013; Khandekar et al., 2007). Deletion of this enhancer results in the loss of HSPC emergence from HE, leading to lethality by E14 (Gao et al., 2013). The same element is also mutated in 10% of all the MonoMAC syndrome patients (Hsu et al., 2013).

Because of a whole genome duplication during the evolution of teleost fish, numerous zebrafish genes exist in the form of two paralogues, including *gata2* (Gillis et al., 2009; Taylor et al., 2003; Taylor et al., 2001). *Gata2a* and *gata2b* are only 57% identical and are thought to have undergone evolutionary sub-functionalisation from the ancestral vertebrate *Gata2* gene (Butko et al., 2015; Liu et al., 2016). This provides an opportunity to separately identify the temporally distinct contributions made by each Gata2 orthologue. *gata2b* is expressed in HE in the floor of the dorsal aorta from 18hpf and is required for *runx1* expression in HE (Butko et al., 2015). In addition, lineage tracing experiments showed that *gata2b*-expressing HE cells gave rise to HSCs in the adult (Butko et al., 2015). Similar to the mouse Gata2, *gata2b* expression depends on Notch signalling (Butko et al., 2015). Thus, *gata2b* is a *bona fide* marker of HE (Gao et al., 2016) and is currently regarded as the functional 'haematopoietic homologue' of Gata2 in zebrafish.

By contrast, *gata2a* is expressed in all endothelial cells and in the developing central nervous system (Butko et al., 2015; Yang et al., 2010) and *gata2a^{um27}* mutants showed arteriovenous shunts in the dorsal aorta at 48hpf (Zhu et al., 2011). However, *gata2a* is expressed already at 11hpf in the haemangioblast population in the posterior lateral mesoderm (PLM) that gives rise to the arterial endothelial cells in the trunk (Patterson et al., 2007). This is well before *gata2b* is expressed in HE,

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

suggesting 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, corresponding to the described +9.5Kb enhancer in the mouse Gata2 locus (Gao et al., 2013; Johnson et al., 2012; Khandekar et al., 2007), a feature that was not found in the *gata2b* locus. Using CRISPR/Cas9 genome editing, we demonstrated that this region, termed the i4 enhancer, is required for endothelial-specific *gata2a* expression. Analysis of homozygous mutants (*gata2a*$^{\Delta i4/\Delta i4}$ mutants) showed decreased expression of the HE-specific gene *runx1* and its upstream regulator, *gata2b.* Thus, endothelial expression of *gata2a*, regulated by the i4 enhancer, is required for *gata2b* expression in the HE. Strikingly, *gata2b* expression recovers and by 48hpf, the expression of haematopoietic markers in *gata2a*$^{\Delta i4/\Delta i4}$ 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 *gata2b* is 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. Strikingly, *gata2a*$^{\Delta i4/\Delta i4}$ adults had decreased numbers of haematopoietic cells in the kidney marrow (WKM), suggesting that Gata2a is required in haematopoietic cells in the adult and its function cannot be rescued by Gata2b. In addition, *gata2a*$^{\Delta i4/\Delta i4}$ mutants showed neutropenia and a propensity to develop AML, a phenotype strikingly similar to the human GATA2 haploinsufficiency syndrome, MonoMAC.

Results

A region of open chromatin in intron 4 of the zebrafish *gata2a* locus is specific to endothelium and highly conserved

Analysis of the zebrafish *gata2a* locus revealed that one region within the fourth intron was conserved across other vertebrates, including mouse and human (Fig. 1A). This region, which we termed 'i4 enhancer', corresponds to the endothelial +9.5 Gata2 enhancer identified previously in the mouse (Gao et al., 2013; Khandekar et al., 2007) and human (Johnson et al., 2012). Notably, the *gata2b* locus did not show broad conservation in non-coding regions (Fig. S1A).

To investigate whether the i4 element was an active enhancer, we performed ATAC-seq (Buenrostro et al., 2013) to identify open chromatin regions in endothelial cells (ECs) in zebrafish. We used the

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tg(*kdrl*:GFP) transgenic line, in which GFP marks all endothelium (Jin et al., 2005). We used only the higher GFP-expressing ECs (*kdrl:GFP^{high}*, termed *kdrl:GFP^{+}* for simplicity) as this fraction was enriched for endothelial markers compared to the *kdrl:GFP^{low}* fraction (Fig. S1B,C). Principal Component Analysis on the ATAC-seq data from 26hpf *kdrl*:GFP^{+} cells (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 (Fig. S1D-F). 78,026 peaks were found in common between replicates of the ATACseq in *kdrl*:GFP^{+} cells (Fig. S1G). 44,025 peaks were differentially expressed between the *kdrl:GFP^{+}* and *kdrl*:GFP^{-} fractions (Fig. S1H). An analysis of known motifs at transcription start site (TSS)-specific peaks (2% of the total peaks) present in the *kdrl:GFP^{+}* population revealed an enrichment for the ETS motif (Fig. S1H,I). ETS factors are essential regulators of gene expression in endothelium (Meadows et al., 2011). In addition, we performed gene ontology (GO) term analysis on the peaks showing >3-fold enrichment or depletion in ECs (Fig S1J-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 (Fig. S1K,L).

Differential peak analysis in the *gata2a* locus identified four differentially open sites within a 20kb genomic region (Fig. 1A), including one peak in intron 4 corresponding to the predicted i4 enhancer. It contained a core 150bp-long element that included several binding motifs for the Gata, E-box and Ets transcription factor families (Fig. 1A). Although the positioning of the E-box site relative to the adjacent GATA site differs in zebrafish and mammals (Fig. 1A), the necessary spacer distance of ~9bp between the two sites (Wozniak et al., 2008) is conserved. Thus, this site may be a target for transcriptional complexes containing an E-box-binding factor and a GATA family TF.

In summary, 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 the endothelium, including the haemogenic endothelium

To investigate the activity of the *gata2a*-i4 enhancer *in vivo*, the conserved genomic 150bp region (Fig. 1A), together with flanking ±500bp (*gata2a*-i4-1.1kb:GFP) or ±150bp (*gata2a*-i4-450bp:GFP) was cloned into a Tol2-based reporter E1b:GFP construct (Kawakami et al., 2004) and used to generate stable transgenic lines (Fig. S2). The earliest activity of the enhancer was observed at the 14-somite stage (14ss), when *gfp* mRNA was detected in the PLM (Fig. S2A,B). After 22hpf, the reporter signal

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

was pan-endothelial (Fig. 1B-C, S2C-I). Around 27hpf, higher intensities of GFP fluorescence and corresponding higher levels of *gfp* mRNA were visible in the floor of the DA (Fig. 1B-C, S2E-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 anymore (Fig. S2I-J). As both reporters showed a similar pattern, we focussed our subsequent analysis on the *gata2a*-i4-1.1kb:GFP transgenics as it showed stronger expression of the transgene (Fig. S2K). At 25hpf, the expression of GFP protein and *gfp* mRNA overlapped completely in the endothelial cells of the DA (Fig. 1D-D''). Overall, these data confirm that the i4 enhancer is active *in vivo* at the correct time window and tissue to regulate definitive haematopoiesis. Importantly, an analogous pattern was observed in mouse embryos driving LacZ under the +9.5 enhancer (Khandekar et al., 2007), suggesting functional conservation of the *gata2a*-i4 enhancer across vertebrates.

To further characterise the enhancer activity *in vivo*, the Tg(*gata2a*-i4-1.1kb:GFP) embryos were stained for *gata2a* mRNA and for GFP protein (Fig. S2K-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%, Fig. S2O). 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 (Fig. S2L-N), indicating that the i4 enhancer is specifically active in (haemogenic) endothelial cells.

Next we examined the expression of the HE marker *runx1* (Gering and Patient, 2005) in *gata2a*-i4-1.1kb:GFP embryos at 25hpf. At this stage, over 90% of *runx1*$^+$ cells were GFP$^+$ (Fig. 1E-F). We conclude that the GFP expression under the *gata2a*-i4 enhancer marks the majority of the HE population.

Deletion of the *gata2a*-i4 enhancer results in decreased expression of *gata2a* in the embryonic endothelium without affecting endothelial development

Gata2a is expressed in early development in various tissues, including ectoderm, primitive erythrocytes, neural tissue and endothelial cells (Meng et al., 1997; Patterson et al., 2007; Yang et al., 2007). To investigate whether the endothelial-specific expression of *gata2a* is required for definitive haematopoiesis, we deleted the conserved *gata2a*-i4 enhancer using CRISPR/Cas9 genome editing (Bassett et al., 2013). We generated a deletion mutant lacking 231bp of the i4 enhancer (Fig. S3A-C, see Material and Methods) and named it *gata2a*$^{\Delta i4/\Delta i4}$. Homozygous *gata2a*$^{\Delta i4/\Delta i4}$ mutants showed decreased levels of *gata2a* mRNA expression in endothelial cells when compared to wild type embryos at 28hpf (X^2=10.720, d.f.=1, p<0.01) (Fig. 2A,B). By contrast, *gata2a* expression in the neural tube appeared unaffected in the *gata2a*$^{\Delta i4/\Delta i4}$ mutants (Fig. 2A,B). At this stage, expression of the pan-

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

endothelial marker *kdrl* was indistinguishable between wild type and *gata2a*$^{\Delta i4/\Delta i4}$ mutants (Fig. 2C,D). To verify these results, we crossed homozygous *gata2a*$^{\Delta i4/\Delta i4}$ mutants to Tg(*kdrl*:GFP) transgenics and analysed vascular morphology. *Gata2a*$^{\Delta i4/\Delta i4}$ 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); *gata2a*$^{\Delta i4/\Delta i4}$ embryos by FACS (Fig. 2G) at 23hpf and 30hpf and confirmed that *kdrl* expression was unaffected in *gata2a*$^{\Delta i4/\Delta i4}$ embryos (Fig. S4A). In addition, the qPCR analysis revealed significantly decreased levels of *gata2a* mRNA in the *kdrl*:GFP^{+} ECs in 23hpf *gata2a*$^{\Delta i4/\Delta i4}$ embryos compared to wild types (t=20.026, d.f.=5, p<0.001) (Fig. 2H). At 30hpf this decrease was not statistically significant. However, this was likely due to a decrease in expression of *gata2a* in wild type ECs that appears to occur normally during development, while the mutant expression levels remained low (Fig. 2H). Importantly, there was no difference in *gata2a* expression in the non-endothelial population (*kdrl*:GFP^{-} cells) between wild type and *gata2a*$^{\Delta i4/\Delta i4}$ mutants at either 23hpf 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 the endothelial tissue. This is in agreement with the reported reduction in *Gata2* expression in the endothelium of mouse *Gata2* +9.5 enhancer deletion mutants (Gao et al., 2013; Johnson et al., 2012).

The previously described *gata2a*$^{um27/um27}$ mutant showed arteriovenous shunts and reduced perfusion of the DA, detectable around 48hpf (Zhu et al., 2011). However, *gata2a*$^{\Delta i4/\Delta i4}$ embryos showed no gross vascular abnormalities at 48hpf. In addition, *gata2a*$^{\Delta i4/\Delta i4}$ mutants and wild type embryos expressed similar levels of the endothelial markers *kdrl*, *dld* and *dll4*. (Fig. S4A-C). By contrast, *gata2a*$^{um27/um27}$ mutants showed reduced *kdrl* expression (Zhu et al., 2011). Moreover, because the DA endothelium is specified normally, the reduction in *gata2a* expression in *gata2a*$^{\Delta i4/\Delta i4}$ mutants likely results from reduced transcription rather than reflecting a secondary phenotype due to an earlier loss of endothelial cells. We conclude that the morphological defects observed in *gata2a*$^{um27/um27}$ mutants were likely a result of an earlier loss of *gata2a* in the haemangioblast, prior to the 14-somite stage where i4 enhancer activity was first detected (Fig. S2A,B).

Deletion of the *gata2a*-i4 enhancer results in reduced expression of *runx1* and *gata2b* at early stages of HE specification

To investigate a potential role of *gata2a* in HSC development, we compared the expression of *runx1*, the key marker of HE in zebrafish (Gering and Patient, 2005; Kalev-Zylinska et al., 2002), in wild type and *gata2a*$^{\Delta i4/\Delta i4}$ embryos. Quantitative *in situ* hybridization analysis (Dobrzycki et al., 2018) revealed

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

decreased expression of *runx1* in *gata2a*$^{\Delta i4/\Delta i4}$ embryos compared to wild type siblings at 28hpf (μ_{wt}=34.8, μ_{mut}=25.3; F=4.956, d.f.=2, 58; p=0.01; ANOVA) (Fig. 3A-C). Further qPCR analysis in *kdrl*:GFP^{+} cells showed that this decrease in *runx1* expression was already detectable at 23hpf (t=2.585, d.f.=5, p<0.05) in *gata2a*$^{\Delta i4/\Delta i4}$ mutants (Fig. 3D), at the onset of its expression in HE (Wilkinson et al., 2009). Thus, deletion of the *gata2a*-i4 enhancer results in impaired *runx1* expression in the early stages of HE programming, suggesting a previously unreported role of *gata2a* in definitive haematopoiesis. This correlates well with published data, where +9.5$^{-/-}$ mouse AGM explants showed lower levels of *Runx1* expression compared to +9.5$^{+/+}$ controls (Gao et al., 2013), further supporting the critical evolutionary role of the intronic enhancer of *Gata2* in HSC specification.

Loss of function experiments showed that Gata2b, another Gata2 paralogue in zebrafish, is an upstream regulator of *runx1* expression in the HE (Butko et al., 2015). Thus, we tested whether *gata2a* could act upstream of *gata2b* by measuring *gata2b* expression in *gata2a*$^{\Delta i4/\Delta i4}$ embryos (Fig.3 E-G). Similarly to *runx1*, quantitation of the *in situ* hybridization signal showed that *gata2b* expression was decreased in *gata2a*$^{\Delta i4/\Delta i4}$ embryos compared to wild type siblings at 28hpf (μ_{wt}=39, μ_{mut}=30.1; F=5.05, d.f.=2, 54; p=0.01; ANOVA) (Fig. 3E-G). In addition, *kdrl*:GFP^{+}; *gata2a*$^{\Delta i4/\Delta i4}$ cells express significantly lower levels of *gata2b* mRNA than the wild type *kdrl*:GFP^{+} endothelial population in 23hpf embryos (t=3.334, d.f.=5, p<0.05) (Fig. 3H). Taken together, these data suggest that endothelial expression of *gata2a* is required upstream of *gata2b* for the proper specification of HE, uncovering a previously unrecognized role for Gata2a in definitive haematopoiesis.

Expression of late markers of embryonic HSC activity is unaffected in *gata2a*$^{\Delta i4/\Delta i4}$ mutants

Strikingly, the qPCR analysis (Fig. 3D,H) showed that by 30hpf the expression of both *runx1* and *gata2b* in *kdrl*:GFP^{+}; *gata2a*$^{\Delta i4/\Delta i4}$ ECs was indistinguishable from wild type *kdrl*:GFP^{+} ECs. Of note, the *kdrl*:GFP^{+} population likely includes the *kdrl*$^{+}$, *runx1*-expressing EMPs located in the caudal region (Bertrand et al., 2007). This region was not included in the quantification of *in situ* hybridization but cannot be separated by sorting for *kdrl*:GFP^{+} only and could thus explain the discrepancy between image quantification and qPCR. Alternatively, the recovery of *gata2b* expression by 30hpf detected by qPCR suggested that other regulators of *gata2b* might compensate for the lack of endothelial *gata2a* in *gata2a*$^{\Delta i4/\Delta i4}$ mutants and thus lead to a recovery of the initial haematopoietic phenotype. Thus, to further characterize the haematopoietic phenotype in the *gata2a*$^{\Delta i4/\Delta i4}$ mutants, we tested whether expression of markers of haematopoietic activity in the embryo was affected from 48hpf onwards (Fig. 4).

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

At 48hpf, the expression of *runx1* in the DA showed no significant difference between *gata2a*$^{\Delta i4/\Delta i4}$ mutants and wild type controls (μ_{wt}=33.1, μ_{mut}=37.5) (Fig. 4A-B). These data suggest that the decrease of *runx1* expression at early stages of HE programming in *gata2a*$^{\Delta i4/\Delta i4}$ mutants is transient and recovers by 2dpf. Indeed, analysis of the HSPC marker *cmyb* (Kissa et al., 2008; Murayama et al., 2006) in the CHT at 4dpf showed no differences between *gata2a*$^{\Delta i4/\Delta i4}$ and wild type larvae (Fig. 4C-D). Expression of the T-cell progenitor marker *rag1* in the thymus (Willett et al., 1997) can be used to assess the output of HSPCs from the DA . While at 4dpf around half of the *gata2a*$^{\Delta i4/\Delta i4}$ larvae showed reduced *rag1* expression compared to wild type (Fig. 4E,F), this effect was absent at 5dpf (Fig. 4G,H). These data suggested that HSPC activity was normal in *gata2a*$^{\Delta i4/\Delta i4}$ mutants from 4dpf onwards. To further characterize this, we crossed the *gata2a*$^{\Delta i4/\Delta i4}$ mutants to the Tg(*itga2b*:GFP;*kdrl*:mCherry) transgenics, where *itga2b*-GFP^{high} and *itga2b*-GFP^{low} cells in the CHT mark thrombocytes and HSPCs, respectively (Kissa et al., 2008; Lin et al., 2005). Our analysis revealed no difference in *itga2b*-GFP^{low} HSPC or *itga2b*-GFP^{high} thrombocyte numbers in the CHT region at 5dpf between wild type and *gata2a*$^{\Delta i4/\Delta i4}$ mutants (Fig.4I-K). 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.

Haematopoietic recovery of *gata2a*$^{\Delta i4/\Delta i4}$ mutants is mediated by the Notch-*gata2b* pathway

We hypothesized that the recovery, not observed in +9.5$^{-/-}$ mouse mutants (Gao et al., 2013), could be due to the presence of two *Gata2* orthologues in zebrafish, where *gata2a* is an upstream regulator of *gata2b* expression. *Gata2b* is crucial for HE specification and HSPC emergence (Butko et al., 2015) and our qRT-PCR data from sorted *kdrl*:GFP^{+} cells suggested that its expression in *gata2a*$^{\Delta i4/\Delta i4}$ mutants recovered by 30hpf (see Fig. 3H). Therefore, we investigated whether the loss of *gata2b* in *gata2a*$^{\Delta i4/\Delta i4}$ background resulted in a more severe haematopoietic phenotype than observed in the *gata2a*$^{\Delta i4/\Delta i4}$ mutants. For this, we injected *gata2a*$^{\Delta i4/\Delta i4}$ and wild type controls with a sub-optimal amount (7.5ng) of a *gata2b* morpholino oligonucleotide (MO) (Butko et al., 2015). Quantitative *in situ* hybridization analysis (Dobrzycki et al., 2018) showed that this amount of *gata2b* MO had no effect on *runx1* expression at 32hpf (μ_{wt}=41.9, μ_{MO}=40.1; n.s., Welch's ANOVA) (Fig. 5A,B). In agreement with the previously observed effect on *runx1* expression in *gata2a*$^{\Delta i4/\Delta i4}$ embryos, we observed a significant reduction compared to wild type siblings (μ_{wt}=41.9, μ_{mut}=28.9; F=44.641, d.f.=3, 62.3; p<0.001) (Fig. 5A,B). However, sub-optimal *gata2b* knockdown in *gata2a*$^{\Delta i4/\Delta i4}$ embryos further significantly reduced levels of *runx1* expression (μ_{mut}=28.9, μ_{mut+MO}=17.2; p<0.001) (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*

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

expression levels as 'wild type' or 'reduced' and found that the 'reduced' embryos were largely overrepresented in the *gata2a*$^{\Delta i4/\Delta i4}$ mutants injected with the *gata2b* MO, compared to wild type fish and non-injected *gata2a*$^{\Delta i4/\Delta i4}$ siblings (X^2=18.784, d.f.=2, p<0.001) (Fig. 5C).

To verify that Gata2b is required for definitive haematopoiesis downstream of Gata2a, we generated a frameshift truncating mutant for Gata2b. Next we incrossed *gata2a*$^{\Delta i4/+}$; *gata2b*$^{+/-}$ adults and investigated *cmyb* expression at 33hpf in their progeny. Both single mutants showed reduced levels of *cmyb* expression in the trunk (Fig S5A-D), suggesting that all *gata2* alleles are important for HSPC generation at this stage. In the absence of Gata2b, Gata2a was not sufficient to drive *cmyb* expression in HE, because *gata2a*$^{+/+}$; *gata2b*$^{-/-}$*;* embryos showed no further reduction in c*myb* expression when compared with *gata2b*$^{-/-}$*; gata2a*$^{\Delta i4/\Delta i4}$ (Fig. S5A-D). Thus, we conclude that Gata2b is required for definitive haematopoiesis and its expression is regulated by Gata2a.

Next, we tested whether forced ectopic expression of *gata2b* was sufficient to speed up the haematopoietic recovery of *gata2a*$^{\Delta i4/\Delta i4}$ embryos. Thus, we overexpressed *gata2b* under the control of the *gata2a*-i4-450bp enhancer in wild type and *gata2a*$^{\Delta i4/\Delta i4}$ mutant embryos and measured *runx1* expression at 28hpf in the DA. As expected, we observed a significant decrease in *runx1* expression in *gata2a*$^{\Delta i4/\Delta i4}$ embryos compared to wild type (μ_{wt}=38.8, μ_{mut}=17.9; F=47.373, d.f.=3, 89; p<0.001; ANOVA) (Fig. 5D). Ectopic expression of *gata2b* under the *gata2a*-i4 enhancer significantly increased the *runx1* pixel intensity levels in wild type (μ_{wt}=38.8, $\mu_{wt+gata2b}$=53.4; p<0.01) and mutants (μ_{mut}=17.9, $\mu_{mut+gata2b}$=33.2; p<0.001) (Fig. 5D). Importantly, it was sufficient to bring the *runx1* expression levels in the mutants up to the levels detected in uninjected wild type fish (μ_{wt}=38.8, $\mu_{mut+gata2b}$=33.2; n.s.) (Fig. 5D), suggesting that *gata2b* alone was sufficient to drive *runx1* expression in ventral wall of the DA and drive the haematopoietic recovery in *gata2a*$^{\Delta i4/\Delta i4}$ mutants. Thus, pan-endothelial *gata2b* can recover the definitive haematopoietic programme even in the absence of endothelial *gata2a*.

Because the expression of *gata2b* is regulated by Notch signalling (Butko et al., 2015), we investigated whether inhibition of Notch would also prevent the haematopoietic recovery of *gata2a*$^{\Delta i4/\Delta i4}$ embryos. For this, we used the Notch inhibitor DAPM (Walsh et al., 2002), and titrated it down to a sub-optimal dose (25μM) that did not affect *runx1* expression in treated embryos (Fig. 5E). Next, we treated wild type and *gata2a*$^{\Delta i4/\Delta i4}$ mutant embryos with the sub-optimal dose of DAPM and measured *runx1* expression in the DA at 36hpf (Fig. 5E,F). As expected, the DAPM treatment did not affect the expression of *runx1* in the wild type embryos (μ_{DMSO}=40.5, μ_{DAPM}=38; n.s.) (Fig. 5E,F). While the *gata2a*$^{\Delta i4/\Delta i4}$ mutants showed lower levels of *runx1* compared to the wild type controls (μ_{DMSO}=40.5, $\mu_{mut+DMSO}$=31.5; F=25.774, d.f.=3, 91; p<0.01; ANOVA) (Fig. 5E,F), DAPM treatment of the mutants further significantly reduced the expression levels of *runx1* ($\mu_{mut+DMSO}$=31.5, $\mu_{mut+DAPM}$=19.4;

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

p<0.001) (Fig. 5E,F). Taken together, these data suggest that activity of Notch is sufficient to drive *gata2b*-mediated recovery in *gata2a*$^{\Delta i4/\Delta i4}$ mutants, as inhibition of either Notch or *gata2b* results in more severe loss of *runx1* expression compared to untreated *gata2a*$^{\Delta i4/\Delta i4}$ embryos. Thus, we conclude that HE programming requires two independent inputs on *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* (Fig. 5E).

Adult *gata2a*$^{\Delta i4/\Delta i4}$ mutants suffer from defects resembling GATA2 haploinsufficiency

Although the expression of haematopoietic markers recovers in the *gata2a*$^{\Delta i4/\Delta i4}$ embryos due to Notch-*gata2b* activity, these mutants may have long term effects, particularly if Gata2a and Gata2b could not fully compensate for each other in adult haematopoiesis. Consistent with this notion, we observed a high incidence of infections and heart oedemas in *gata2a*$^{\Delta i4/\Delta i4}$ 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 point to an additional function of the *gata2a*-i4 enhancer in the heart or the lymphatic system, while the infections may suggest immune defects, as observed in human patients that bear genetic GATA2 haploinsufficiency syndromes like Emberger's or MonoMAC syndrome (Spinner et al., 2014). Around 10% of MonoMAC syndrome patients show mutations in the homologous enhancer region of *GATA2* (Hsu et al., 2013; Wlodarski et al., 2016). Therefore, to investigate whether the phenotype in adult *gata2a*$^{\Delta i4/\Delta i4}$ mutants resembled human Gata2 haploinsufficiency in humans, we first asked whether the *gata2a*-i4-1.1kb:GFP 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 cytometry (Traver et al., 2003) as progenitors, lymphoid+HSPC (containing the HSPCs) and myeloid cells (Fig S6A-C)). Accordingly, single cell transcriptional profiling showed higher levels of *gata2a* in HSPCs, progenitors, neutrophils and thrombocytes (Fig. S6D-F) (Athanasiadis et al., 2017; Macaulay et al., 2016).

Next, we counted the total number of haematopoietic cells in the WKM of wild types and *gata2a*$^{\Delta i4/\Delta i4}$ mutants (Fig 6D-F). To avoid any confounding effects in our analysis, we compared WKM of wt to *gata2a*$^{\Delta i4/\Delta i4}$ mutants without overt signs of infection. The *gata2a*$^{\Delta i4/\Delta i4}$ mutants showed a ~2-fold decrease in the total number of haematopoietic cells in the WKM ($\mu_{wt}=4.37x10^5$;$\mu_{mut}=2.37x10^5$, p=0.0185, Mann-Whitney test) (Fig. 6D-F). In addition, the numbers of neutrophils were similarly reduced ($\mu_{wt}=2.17x10^5$;$\mu_{mut}=1.03x10^5$, p=0.0269, Mann-Whitney test) (Fig. 6G), another characteristic phenotype of MonoMAC syndrome patients (Wlodarski et al., 2016). Lastly, cytology of kidney

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

marrow smears of three 9-month old *gata2a*$^{\Delta i4/\Delta i4}$ mutants were assessed. A clear AML was detected in one of the three mutants, with 98% blasts and only minor erythrocyte differentiation in the WKM (Fig. 6H,I). Together these data strongly suggest that the i4 enhancer is a critical driver of gata2a expression in adult haematopoietic cells. Loss of *gata2a* expression in *gata2a*$^{\Delta i4/\Delta i4}$ mutants leads to a hypocellular WKM and neutropenia, strongly suggestive of marrow failure that can progress to AML, a hallmark of disease progression in MonoMAC syndrome patients.

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

Endothelial expression of *gata2a* through a conserved intronic enhancer is required for HE programming

Gata2 is critically required for the generation and maintenance of embryonic HSCs (de Pater et al., 2013; Tsai et al., 1994). Expression of *Gata2* in the haemogenic endothelium (HE) and its subsequent role in the specification of the HSCs depends on its intronic enhancer, termed the VE enhancer (Khandekar et al., 2007) or +9.5 enhancer (Gao et al., 2013). The homologous region in humans was also correlated with susceptibility to blood-related disorders (Johnson et al., 2012). Here we have investigated the conservation of this 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 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. The zebrafish *gata2a*-i4 enhancer contains multiple conserved transcription factor binding sites and, like the mouse VE enhancer (Khandekar et al., 2007), is sufficient to drive pan-endothelial expression of GFP and necessary for endothelial expression of *gata2a* (Fig. 1,2). We traced the activity of the i4 enhancer back to the PLM, the source of precursors of endothelium and HSCs (Davidson and Zon, 2004; Medvinsky et al., 2011). 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 (*gata2a*$^{\Delta i4/\Delta i4}$) allowed us to uncover a previously unknown function of Gata2a in regulating expression of its paralogue *gata2b* in HE. *Runx1*, a target of Gata2b (Butko et al., 2015) that is critical for the EHT process (Kissa and Herbomel, 2010), was also reduced in *gata2a*$^{\Delta i4/\Delta i4}$ mutants, suggesting that a novel Gata2a/Gata2b/Runx1 axis regulates gene expression in the HE, before the first reported EHT events at 34hpf (Kissa and Herbomel, 2010).

Regulation of *gata2a* and *gata2b*

Gata2 is thought to bind to the +9.5 enhancer to maintain its own expression in endothelial and haematopoietic cells (Johnson et al., 2012; Sanalkumar et al., 2014). In zebrafish, it is likely that Gata2a binds the GATA motifs in the i4 enhancer and loss of *gata2a* in the endothelium of *gata2a*$^{\Delta i4/\Delta i4}$ 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 (Fig. S1A) and thus could potentially represent a divergent gata2b intronic enhancer. We speculate that the positive autoregulation of Gata2 was likely retained by both *gata2* orthologues in zebrafish. In this case, Gata2a would bind to the *gata2b* intronic enhancer in HE

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

only until enough Gata2b is present, at which point its intronic enhancer would 'switch' Gata2a for Gata2b to maintain *gata2b* expression. This 'switch', however, would replace one activator for another, rather than replacing an activating Gata factor (Gata2) for a repressive Gata factor as described previously for the Gata2/Gata1 switch in erythroid cells (Grass et al., 2006). Alternatively, the positive auto-regulation function might have been modified in HE so that Gata2a binds to both *gata2a* and *gata2b* intronic enhancers to regulate their expression. These possibilities remain to be investigated.

Rescue of the haematopoietic defects in *gata2a*$^{\Delta i4/\Delta i4}$ mutants

The gata2a$^{\Delta i4/\Delta i4}$ mutants recovered from the early defects in HE programming and displayed normal expression levels of *cmyb* in the CHT at 4dpf and *rag1* in the thymus at 5dpf, used as indicators of the definitive haematopoietic programme (Murayama et al., 2006). We hypothesized that this could be due to the presence of the two homologues of *Gata2* in zebrafish (Gillis et al., 2009), despite Gata2a and Gata2b proteins being only 50% identical (Butko et al., 2015). Indeed, forced expression of *gata2b* under the *gata2a*-i4 enhancer rescued the expression of *runx1* in the *gata2a*$^{\Delta i4/\Delta i4}$ mutants to wild type levels and sub-optimal depletion of *gata2b* in the *gata2a*$^{\Delta i4/\Delta i4}$ mutants resulted in more severe reduction in *runx1* and *cmyb* expression in the CHT by 4dpf (Fig.4). In addition, we demonstrated that Notch signalling, a known regulator of *gata2b* expression (Butko et al., 2015), is sufficient to rescue the initial HE programming defect induced by deletion of the *gata2a*-i4 enhancer. We propose a model in which *gata2a* acts upstream of *gata2b* and independently of Notch to initiate HE programming. In this context, Notch signalling functions 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).

Gata3 is also expressed in ECs prior to HSC emergence (Fitch et al., 2012) and could contribute towards the haematopoietic recovery in *gata2a*$^{\Delta i4/\Delta i4}$ mutants. However, distinguishing its non-cell autonomous role in the production of catecholamines (Fitch et al., 2012) from a putative role in endothelium would require tissue-specific inactivation of Gata3 using a Cre-lox system or by deleting endothelial-specific enhancers as we did for the *gata2a*-i4 enhancer. There are very few examples reported to date and, to our knowledge, only one in the zebrafish model (Chiang et al., 2017).

In the mouse, two different studies showed that mouse +9.5$^{-/-}$ enhancer mutants exhibit decreased haematopoietic output from cultured AGM explants (Gao et al., 2013) or foetal liver HSCs (Johnson et al., 2012). However, in both cases the experiments were performed at E11.5, well after the HE had been established and given rise to HSCs (Swiers et al., 2013). Furthermore, at E11.5 the numbers of haematopoietic cells in the AGM were roughly comparable to their wild type siblings (Gao et al.,

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2013; Johnson et al., 2012), raising the possibility that an initial HE defect might have gone unnoticed in those studies. Thus, by analysing the contribution of two zebrafish Gata2 paralogues, we uncovered a previously unappreciated contribution by a Gata2 gene in the programming of HE prior to HSC specification.

Loss of i4 enhancer activity leads to a Gata2 haploinsufficiency phenotype similar to the human MonoMAC syndrome

Although the *gata2a*$^{\Delta i4/\Delta i4}$ embryos showed an apparent haematopoietic recovery by 5dpf, we observed a high incidence of infections and oedema in *gata2a*$^{\Delta i4/\Delta i4}$ adults, and a striking decrease in the number of haematopoietic cells in the WKM. The decrease in haematopoietic cells in particular is reminiscent of the loss of proliferative potential of haematopoietic Gata2$^{+/-}$ heterozygous cells in the mouse (Ling et al., 2004; Rodrigues et al., 2005). 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 WKM. This will be addressed by comparing the adult phenotypes of *gata2a*$^{\Delta i4/\Delta i4}$ and gata2b$^{-/-}$ mutants.

Taken together, our initial characterization of WKM suggests that *gata2a*$^{\Delta i4/\Delta i4}$ mutants present a phenotype consistent with that of MonoMAC syndrome patients (Hsu et al., 2013; Johnson et al., 2012). Strikingly, about 10% of all MonoMAC patients show mutations in the conserved +9.5 enhancer (Hsu et al., 2013; Wlodarski et al., 2016), the corresponding regulatory element to the i4 enhancer. The i4 enhancer is active in the lymphoid+HSPC fraction that contains the HSC activity (Ma et al., 2011), in the progenitor cells and in the myeloid fraction that contains monocytes and other myeloid cells including the eosinophils, previously identified as expressing high levels of a *gata2a*-GFP BAC transgenic reporter (Traver et al., 2003). Similarly, *runx1*$^{W84X/W84X}$ mutants that recovered definitive haematopoiesis despite the lack of *runx1*, showed a decrease in the total number of blood cells in the WKM, especially myeloid and progenitor cells (Sood et al., 2010). This suggests lineage-specific differentiation defects, which are likely present in the *gata2a*$^{\Delta i4/\Delta i4}$ adult fish. In MonoMAC patients, the mononuclear cells (dendritic cells, monocytes, B and Natural Killer lymphocytes) are frequently affected (Collin et al., 2015). Further characterization of the *gata2a*$^{\Delta i4/\Delta i4}$ mutants will uncover which haematopoietic cells are most affected by the loss of i4 enhancer activity and how the mutants progress to AML, thus establishing a novel animal model for human diseases linked to *Gata2* haploinsufficiency.

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements

We thank Fatma Kok for critical reading of the manuscript. We thank the staff of the Biomedical Services Units (Oxford, Birmingham and Rotterdam) for fish husbandry. We thank Kevin Clark and Sally-Ann Clarke 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), and WIMM Strategic Alliance awards G0902418 and MC_UU_12025. We thank the Wolfson Imaging Centre Oxford imaging. The Wolfson Imaging Centre Oxford is supported by the MRC via the WIMM Strategic Alliance (G0902418), the Molecular Haematology Unit (MC_UU_12009), the Human Immunology Unit (MC_UU_12010), the Wolfson Foundation (grant 18272), and an MRC/BBSRC/EPSRC grant (MR/K015777X/1) to MICA – Nanoscopy Oxford (NanO): Novel Super-resolution Imaging Applied to Biomedical Sciences, Micron (107457/Z/15Z). The facility was supported by WIMM Strategic Alliance awards G0902418 and MC_UU_12025. This research was supported by the British Heart Foundation (BHF Oxford CoRE Transition Fellowship to R.M., BHF IBSR Fellowship FS/13/50/30436 to R.M. and M. K.) and by the MRC MHU programme number MC_UU_12009/8.

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. and R.M. performed experiments and analyzed the data. J.P-Z. 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.

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure legends

Figure 1. The i4 enhancer present in intron 4 of the *gata2a* locus is conserved and drives pan-endothelial expression of a GFP reporter in zebrafish. (A) *Kdrl*:GFP^{+} (green) and *kdrl*:GFP^{-} (blue) cells were FACS-sorted from 26hpf embryos and used for preparation of ATAC-seq libraries. The image of the mapped reads presents stacked means of two biological replicates. Differential peak analysis identified four chromatin regions (purple shading) in the locus of *gata2a* that are significantly more open in the *kdrl*:GFP^{+} population (p<0.0001). A region in the fourth intron (termed i4 enhancer) is conserved throughout vertebrates. The highly conserved 150bp region (red) contains putative transcription factor binding sites, mapped computationally. Light blue: Ets binding sites; purple: E-box binding sites; green: GATA binding sites; grey shading: regions of high conservaton; asterisks: conserved residues. (B) Widefield fluorescent image of a live Tg(*gata2a*-i4-1.1kb:GFP) zebrafish embryo at 27hpf showing GFP fluorescence in the endothelial cells and in the heart (endocardium). (C) Higher magnification image of the trunk of the embryo from panel B. (D-D') Confocal images of a trunk fragment of a Tg(*gata2a*-i4-1.1kb:GFP) embryo immunostained with anti-GFP antibody (D) and probed for *gfp* mRNA (D') at 25hpf. (D") Merged images from panels D-D' with Hoechst nuclear staining in blue, showing complete overlap of GFP protein and mRNA. (E-E') Confocal images of the dorsal aorta (DA) and posterior cardinal vein (PCV) of a Tg(*gata2a*-i4-1.1kb:GFP) embryo immunostained with anti-GFP antibody (E) and probed for *runx1* mRNA (E') at 25hpf. See panel B for approximate position within the embryo. (E") Merged images from panels E-E', also showing Hoechst nuclear staining in blue. (F) Counting of the *runx1*$^{+}$ cells represented in panels E'-E" in 25 embryos shows that >90% of *runx1*$^{+}$ cells are also GFP^{+}. N=3. Error bars: ±SD.

Figure 2. Deletion of the i4 enhancer in *gata2a*$^{\Delta i4/\Delta i4}$ mutants leads to reduced levels of *gata2a* mRNA in the endothelium. (A-B) A significant majority of *gata2a*$^{\Delta i4/\Delta i4}$ mutants have reduced levels of *gata2a* mRNA in the dorsal aorta (arrows) at 28hpf, compared to wild type siblings, as detected with *in situ* hybridization. The expression in the neural tube appears unaffected. **p<0.01. (C-D) *In situ* hybridization for the endothelial marker *kdrl* at 28hpf reveals no difference between *gata2a*$^{\Delta i4/\Delta i4}$ mutants and wt siblings. The dorsal aorta (arrows) appears unaffected. (E-F) Live images of the trunks of 48hpf Tg(*kdrl*:GFP) and Tg(*kdrl*:GFP); *gata2a*$^{\Delta i4/\Delta i4}$ embryos show normal vascular morphology in the mutants. The endothelium of the dorsal aorta (arrows) appears normal in the *gata2a*$^{\Delta i4/\Delta i4}$ embryos. (G) *Kdrl*:GFP^{high} and *kdrl*:GFP^{-} cells were sorted from non-mutant (WT, blue) and *gata2a*$^{\Delta i4/\Delta i4}$ (red) embryos carrying the Tg(*kdrl*:GFP) transgene. (H-I) qRT-PCR on RNA isolated from the sorted *kdrl*:GFP^{high} or *kdrl*:GFP^{-} cells (panel G) shows decreased levels of *gata2a* mRNA in the endothelium of *gata2a*$^{\Delta i4/\Delta i4}$ mutants at 23hpf and at 30hpf, compared to wild type. There is no difference in *gata2a* mRNA levels in non-endothelial cells between wt and *gata2a*$^{\Delta i4/\Delta i4}$ mutants. N=4

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

for *gata2a*$^{\Delta i4/\Delta i4}$ at 23hpf, N=3 for other samples. Note different scales of expression levels. ***$p<0.001$.

Figure 3. Loss of gata2a expression in the endothelium of *gata2a*$^{\Delta i4/\Delta i4}$ mutants leads to decreased levels of *runx1* and *gata2b* in the HE. (A-B) *In situ* hybridization for *runx1* expression in the HE of wild type and *gata2a*$^{\Delta i4/\Delta i4}$ embryos at 28hpf (arrows). (C) Quantification of the *runx1 in situ* hybridization signal from wild type (blue), heterozygous *gata2a*$^{+/\Delta i4}$ (het, yellow) and *gata2a*$^{\Delta i4/\Delta i4}$ (red) siblings at 28hpf shows significant decrease in *runx1* pixel intensity in the DA in the homozygous mutants compared to wild type (n=14, wild type; n=25, het; n=23, *gata2a*$^{\Delta i4/\Delta i4}$). Error bars: mean±SD. **$p<0.01$. (D) qRT-PCR on RNA isolated from the sorted *kdrl*:GFP^{+} cells shows decreased levels of *runx1* mRNA in the endothelium of *gata2a*$^{\Delta i4/\Delta i4}$ mutants at 23hpf but not at 30hpf, compared to wild type. N=4 for *gata2a*$^{\Delta i4/\Delta i4}$ at 23hpf, N=3 for other samples. Note different scales of expression levels. *$p<0.05$. (E-F) *Gata2b* expression in the HE of wild type and *gata2a*$^{\Delta i4/\Delta i4}$ embryos at 28hpf (arrows). (G) Quantification of the *gata2b* mRNA signal, detected by *in situ* hybridization, from wild type (blue), heterozygous *gata2a*$^{+/\Delta i4}$ (het; yellow) and *gata2a*$^{\Delta i4/\Delta i4}$ (red) siblings at 28hpf shows significant decrease in *gata2b* pixel intensity in the DA in the homozygous mutants compared to wild type (n=22, wild type; n=24, het; n=11, *gata2a*$^{\Delta i4/\Delta i4}$). Error bars: mean±SD. *$p<0.05$. (H) qRT-PCR in sorted *kdrl*:GFP^{+} cells showed decreased levels of *gata2b* mRNA in the endothelium of *gata2a*$^{\Delta i4/\Delta i4}$ mutants at 23hpf but not at 30hpf, compared to wild type. N=4 for *gata2a*$^{\Delta i4/\Delta i4}$ at 23hpf, N=3 for other samples. *$p<0.05$.

Figure 4. *Gata2a*$^{\Delta i4/\Delta i4}$ mutants display a recovery of the initial haematopoietic defects from 48hpf. (A) Representative image of *runx1* expression in the trunk of a wild type embryo at 48hpf showing *runx1* mRNA in the dorsal aorta (arrow). (B) Quantification of the *runx1 in situ* hybridization signal in wild type (blue) and *gata2a*$^{\Delta i4/\Delta i4}$ mutants (red) siblings at 48hpf. There is no significant difference in *runx1* pixel intensity in the DA between the homozygous mutants and wild type (n=19, wild type; n=27, *gata2a*$^{\Delta i4/\Delta i4}$). Error bars: mean±SD. (C-D) *In situ* hybridization for *cmyb* in the CHT. We detected no difference in expression between wild type and *gata2a*$^{\Delta i4/\Delta i4}$ siblings at 4dpf. (E-H) *In situ* hybridization (ventral image) for *rag1* in the thymii, showing a slight decrease (relative to wild type) in *rag1* (red arrows) in approximately half of the homozygous mutant embryos at 4dpf. This effect is absent at 5dpf. (I-J) Maximum projections of *itga2b*:GFP transgenic embryos in the CHT at 5dpf in (I) wild type and (J) *gata2a*$^{\Delta i4/\Delta i4}$ siblings. (K) HSPC (*itga2b*:GFP^{low}) counts in the CHT of wild type (n=10) and *gata2a*$^{\Delta i4/\Delta i4}$ mutants (n=12) at 5dpf. No difference was detected between genotypes ($\mu_{wt}=153.5$; $\mu_{mut}=145.5$; p=0.98, Mann-Whitney test). (L) Thrombocyte (*itga2b*:GFP^{high}) counts in the CHT of wild type (n=10) and *gata2a*$^{\Delta i4/\Delta i4}$ mutants (n=12) at 5dpf. No difference was detected between genotypes (($\mu_{wt}=13$; $\mu_{mut}=13$; p=0.71, Mann-Whitney test). The scatter plots show the median±SD.

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 5. *Gata2b* and Notch signalling are sufficient to recover the expression of *runx1* and *cmyb* in *gata2a*$^{\Delta i4/\Delta i4}$ mutants. (A) Expression of *runx1* in the HE at 32hpf in wild type, *gata2b MO*-injected (7.5ng) wild type embryos, *gata2a*$^{\Delta i4/\Delta i4}$ mutants and *gata2b MO*-injected (7.5ng) *gata2a*$^{\Delta i4/\Delta i4}$ mutants. (B) Quantification of the *runx1 in situ* hybridization signal in wild type (blue), *gata2b* morphants (red squares) *gata2a*$^{\Delta i4/\Delta i4}$ mutants (red triangles) and *gata2a*$^{\Delta i4/\Delta i4}$ mutants injected with *gata2b* MO (yellow). Sub-optimal knockdown of *gata2b* significantly decreases *runx1* in the DA of *gata2a*$^{\Delta i4/\Delta i4}$mutants, but not in wild type embryos, at 32hpf. (n=27, wild type; n=27, *gata2a*$^{\Delta i4/\Delta i4}$; n=33, wild type + *gata2b* MO; n=32, *gata2a*$^{\Delta i4/\Delta i4}$ + *gata2b* MO). Error bars: mean±SD. ***p<0.001. (C) Scoring of *cmyb* expression at 4dpf in wild type embryos, *gata2a*$^{\Delta i4/\Delta i4}$ mutants and *gata2a*$^{\Delta i4/\Delta i4}$ mutants injected with *gata2b* MO as wild type (blue) or reduced (red). Sub-optimal knockdown of *gata2b* (7.5ng of *gata2b* MO) inhibits the haematopoietic recovery of *gata2a*$^{\Delta i4/\Delta i4}$ mutants. (D) Quantification of the *runx1* mRNA signal, detected by *in situ* hybridization, from 28hpf wild type embryos (blue), *gata2a*$^{\Delta i4/\Delta i4}$ 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 wild type embryos and rescues *runx1* expression in the DA of *gata2a*$^{\Delta i4/\Delta i4}$ mutants to wild type levels (n=25, wild type; n=33, *gata2a*$^{\Delta i4/\Delta i4}$;n=18, wild type + *gata2a*-i4-450bp:*gata2b* construct; n=17, *gata2a*$^{\Delta i4/\Delta i4}$ + *gata2a*-i4-450bp:*gata2b* construct) Error bars: mean±SD. **p<0.01; ***p<0.001. (E) Quantification of the *runx1 in situ* hybridization signal at 36hpf in embryos treated with a suboptimal dose (25μM) of the Notch inhibitor DAPM. 25 μM DAPM showed no effect on *runx1* expression in wild type (red squares) compared to DMSO-treated embryos (blue circles). As expected, DMSO-treated *gata2a*$^{\Delta i4/\Delta i4}$ mutants (red triangles) show a decrease in *runx1* expression whereas the DAPM treatment (yellow squares) significantly reduced *runx1* expression in the DA *gata2a*$^{\Delta i4/\Delta i4}$ mutants. (n=27, wild type+ DMSO; n=20, *gata2a*$^{\Delta i4/\Delta i4}$+ DMSO; n=30, wild type+DAPM; n=20, *gata2a*$^{\Delta i4/\Delta i4}$+ DAPM). Error bars: mean±SD. **p<0.01; ***p<0.001. (F) Representative images of the average *runx1* expression at 36hpf upon treatment of wild type and *gata2a*$^{\Delta i4/\Delta i4}$ mutants with 25μM DAPM.

Figure 6. *Gata2a*$^{\Delta i4/\Delta i4}$ mutants show cardiac oedema and marrow failure ultimately leading to AML, consistent with the MonoMAC syndrome phenotype in humans. (A-B) General morphology of zebrafish adults: (A) wild type; (B) *gata2a*$^{\Delta i4/\Delta i4}$ mutant showing skin infection (blue arrowhead) and pericardial oedema (yellow arrowhead). (C) Over 25% (n=29/108) of *gata2a*$^{\Delta i4/\Delta i4}$ mutants (red) catch infections or suffer from heart oedemas by 6 months. Only around 65% (n=69/108) survive for more than 12 months without overt signs of infections. Fewer than 1% (n=2/500) of wild type fish (blue) exhibit such defects. The graph does not include deaths by other causes. (D-E) May-

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Grunwald/Wright-Giemsa staining in cytospins of haematopoietic cells isolated from the WKM of zebrafish adults: (D) wild type; (E) *gata2a*$^{\Delta i4/\Delta i4}$ mutant. Note the apparent decrease in cell numbers. (F) Cell counts of haematopoietic cells isolated from WKM of wild type (n=14) and *gata2a*$^{\Delta i4/\Delta i4}$ mutants (n=8). The *gata2a*$^{\Delta i4/\Delta i4}$ mutants show a ~2-fold decrease in haematopoietic cell numbers in the WKM (p=0.0185, Mann-Whitney test). (G) Number of neutrophils isolated from WKM of wild type (n=14) and *gata2a*$^{\Delta i4/\Delta i4}$ mutants (n=7). The *gata2a*$^{\Delta i4/\Delta i4}$ mutants show a ~2-fold decrease in neutrophil numbers in the WKM (p=0.0296, Mann-Whitney test). The scatter plots show the median cell number±SD. (H-I) Kidney smears from 9 months post-fertilization adult animals were assessed. (H) Wild type shows various stages of lineage differentiation. (I) WKM smear 1 of 3 *gata2a*$^{\Delta i4/\Delta i4}$ mutants progressed to AML with very little erythroid differentiation (98% blasts, > 200 cells assessed). Scalebars: 500μm (A,B) and 10μm (D,E, H, I)

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

Maintenance of zebrafish

Zebrafish (*Danio rerio*) were maintained in flowing system water at 28.5°C, conductance 450-550µS and pH 7.0±0.5 as described (Westerfield, 2007). 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 (Kimmel et al. 1995) corresponding to respective age in hours or days post fertilization (hpf or dpf, respectively). Published lines used in this work were wild type (wt^{KCL}) and Tg(*kdrl*:GFP)^{s843} (Jin et al. 2005). All animal experiments were approved by the relevant University of Oxford, University of Birmingham and Erasmus University ethics committees.

ATAC-seq

Tg(*kdrl*:GFP)^{s843} embryos were dissociated for FACS at 26-27hpf to collect *kdrl* $^{+}$ and *kdrl* $^{-}$ cell populations (40,000-50,000 cells each). They were processed for ATAC library preparation using optimised standard protocol (Buenrostro et al., 2013). Briefly, after sorting into Hanks' solution (1xHBSS, 0.25% BSA, 10mM HEPES pH8), the cells were spun down at 500g at 4°C, washed with ice-cold PBS and resuspended in 50µl cold Lysis Buffer (10mM Tris-HCl, 10mM NaCl, 3mM $MgCl_2$, 0.1% IGEPAL, pH 7.4). The nuclei were pelleted for 10min. at 500g at 4°C and ressuspended in the TD Buffer with Tn5 Transposase (Illumina), scaling the amounts of reagents accordingly to the number of sorted cells. The transposition reaction lasted 30min. at 37°C. The DNA was purified with PCR Purification MinElute Kit (QIAGEN). In parallel, transposase-untreated genomic DNA from *kdrl*$^{+}$ cells was purified with the DNeasy® Blood & Tissue Kit (QIAGEN). The samples were amplified with appropriate Customized Nextera primers (Buenrostro et al., 2013) in NEBNext High-Fidelity 2x PCR Master Mix (NEB). The libraries were purified with PCR Purification MinElute Kit (QIAGEN) and Agencourt AMPure XP beads (Beckmann Coulter). 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), diluted to 1ng/µl and sequenced using 75bp paired-end reads on Illumina HiSeq 4000 (Wellcome Trust Centre for Human Genetics, Oxford). 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) (Howe et al., 2013) using *STAR* with default parameters. The results were visualised using UCSC Genome Browser (http://genome-euro.ucsc.edu/) (Kent et al., 2002). The eight data sets were analysed with Principal Component Analysis (PCA) to identify outliers. Correlation

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

among *kdrl:GFP* $^{+}$ and *kdrl:GFP* $^{-}$ samples was assessed with a tree map. The peaks were called for each sample using the Tn5-untreated control as input. DiffBind (EdgeR method) was used to identify differential peaks between *kdrl* $^{+}$ and *kdrl* $^{-}$ samples (Table S2). The threshold for differential peaks was $p<0.05$.

Generation of transgenic Tg(*gata2a*-i4:GFP) and mutant *gata2a*$^{\Delta i4/\Delta i4}$ and gata2b$^{-/-}$ zebrafish lines

Genomic regions containing the identified 150bp-long *gata2a*-i4 enhancer flanked by ±500bp (i4-1.1kb) or ±150bp (i4-450bp) were amplified from wild type zebrafish genomic DNA with NEB Phusion® polymerase (see Table S1 for primer sequences) and cloned upstream of E1b minimal promoter and GFP into a Tol2 recombination vector (Addgene plasmid #37845, (Birnbaum et al., 2012)) with Gateway® cloning technology (Life Technologies™) following the manufacturer's protocol. One-cell zebrafish embryos were injected with 1nl of an injection mix, containing 50pg *gata2a*-i4-E1b-GFP-Tol2 construct DNA + 30pg *tol2* transposase mRNA (Kawakami et al., 2004). Transgenic founders (Tg(*gata2a*-i4:GFP)) were selected under a widefield fluorescent microscope and outbred to wt fish. Carriers of monoallelic insertions were detected by the Mendelian distribution of 50% fluorescent offspring coming from wt outcrosses. These transgenics were then inbred to homozygosity.

To generate the i4 deletion mutant, we identified potential sgRNA target sites flanking the 150bp conserved region within intron 4 of the *gata2a* locus (see Fig. 1A, Fig S3A). sgRNAs were designed with the CRISPR design tool (http:/crispr.mit.edu/, see Table S1 for sequences) and prepared as described (Bassett et al., 2013). To reduce potential off-target effects of CRISPR/Cas9, we utilized the D10A 'nickase' version of Cas9 nuclease (Mali et al., 2013; Ran et al., 2013), together with two pairs of sgRNAs flanking the enhancer (Table S1, Fig. S3A-B). We isolated two mutant alleles with deletions of 215bp (Δ78-292) and 231bp (Δ73-303) (Fig. S3B). Both deletions included the highly conserved E-box, Ets and GATA transcription factor binding sites (Fig. S3B). The Δ73-303 allele was selected for further experiments and named Δi4. Adult zebrafish were viable and fertile as heterozygous (*gata2a*$^{\Delta i4/+}$) or homozygous (*gata2a*$^{\Delta i4/\Delta i4}$). To unambiguously genotype wild types, heterozygotes and homozygous mutants, we designed a strategy consisting of two PCR primer pairs (Fig. S3A, C). One primer pair flanked the whole region, producing a 600bp wild type band and 369bp mutant band. In the second primer pair, one of the primers was designed to bind within the deleted region, only giving a 367bp band in the presence of the wild type allele (Fig. S3C).

To generate the gata2b mutant we designed a CRISPR/Cas9 strategy for a frameshift truncating mutant in exon 3 deleting both zinc fingers. sgRNAs were designed as described above and guides were prepared according to Gagnon et al. (Gagnon et al., 2014) with minor adjustments. Guide RNAs

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

were generated using the Agilent SureGuide gRNA Synthesis Kit, Cat# 5190-7706. Cas9 protein (IDT) and guide were allowed to form ribonucleoprotein 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. To get rid of additional mutations caused by potential off target effects, founder 3 was crossed to WT for at least 3 generations. All experiments were performed with offspring of founder 3.

Fluorescence-activated cell sorting (FACS)

~100 embryos at the required stage were collected in Low Binding® SafeSeal® Microcentrifuge Tubes (Sorenson) and pre-homogenized by pipetting up and down in 500µl Deyolking Buffer (116mM NaCl, 2.9mM KCl, 5mM HEPES, 1mM EDTA). They were spun down for 1min. at 500g and incubated for 15min. at 30°C in Trypsin + Collagenase Solution (1xHBSS, 0.05% Gibco® Trypsin+EDTA (Life Technologies™), 20mg/ml collagenase (Sigma)). During that time, they were homogenized by pipetting up and down every 3min. The lysis was stopped by adding 50µl foetal bovine serum and 650µl filter-sterilized Hanks' solution (1xHBSS, 0.25% BSA, 10mM HEPES pH8). The cells were rinsed with 1ml Hanks' solution and passed through a 40µm cell strainer (Falcon®). They were resuspended in ~400µl Hanks' solution with 1:10,000 Hoechst 33258 (Molecular Probes®) and transferred to a 5ml polystyrene round bottom tube for FACS sorting. The cells were sorted on FACSAria Fusion sorter by Kevin Clark (MRC WIMM FACS Facility). The gates of GFP (488-530) and DsRed (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 buffer (QIAGEN) + 1% β-mercaptoethanol and processed with the RNEasy® 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

3µl of the cDNA diluted in H_2O were used for technical triplicate qRT-PCR reactions of 20µl containing the Fast SYBR® Green Master Mix (Thermo Fisher Scientific) and appropriate primer pair (see Table

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

S1). 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 *ubc* from the average 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 $2^{-\Delta Ct}$.

Fluidigm Biomark qRT-PCR

To quantify the differences in *gata2a* expression between wild type and mutant ECs, we crossed homozygous *gata2a*$^{\Delta i4/\Delta i4}$ mutants to Tg(*kdrl*:GFP) transgenics to generate Tg(*kdrl*:GFP); *gata2a*$^{\Delta i4/\Delta i4}$ embryos. These fish, along with non-mutant Tg(*kdrl*:GFP), were used for FACS-mediated isolation of *kdrl*:GFP^{+} and *kdrl*:GFP^{-} cells to quantitatively compare mRNA expression levels of *gata2a* in the endothelial and non-endothelial cells of wild type and *gata2a*$^{\Delta i4/\Delta i4}$ embryos, using the Fluidigm Biomark™ qRT-PCR platform. Briefly, 1ng RNA from FACS-sorted cells was used for Specific Target Amplification in a 10µl reaction with the following reagents: 5µl 2xBuffer and 1.2µl enzyme mix from SuperScript III One-Step Kit (Thermo Fisher Scientific), 0.1µl SUPERase• In™ RNase Inhibitor (Ambion), 1.2µl TE buffer (Invitrogen), 2.5µl 0.2x TaqMan® assay mix (see Table S3 for the details of TaqMan® assays). The reaction was incubated for 15min. at 50°C, for 2min. at 95°C and amplified for 20 cycles of 15s at 95°C/4min. at 60°C. The cDNA was diluted 1:5 in TE buffer and stored at -20°C. Diluted cDNA was used for qRT-PCR according to the Fluidigm protocol for Gene Expression with the 48.48 IFC Using Standard TaqMan® Assays (Table S3). Each sample was run in 3-4 biological replicates. The collected data were analysed with Fluidigm Real-Time PCR Analysis software (version 4.1.3). The baseline was automatically corrected using the built-in Linear Baseline Correction. The thresholds were manually adjusted for each gene to fall within the linear phase of exponential amplification, after which they were set to equal values for the housekeeping genes: *rplp0*, *rpl13a*, *cops2* (Xu et al., 2016), *lsm12b* (Hu et al., 2016) and *eef1a1l1*. The relative mRNA levels for each sample were calculated by subtracting the geometric mean of Ct values for the housekeeping genes from the Ct value for each gene of interest. This value (ΔCt) was then converted to a ratio relative to the housekeeping genes with the formula $2^{-\Delta Ct}$. The ΔCt values were analysed with 2-tailed paired-samples *t*-tests with 95% confidence levels. For Fluidigm Biomark™ qRT-PCR, the ΔCt values were analysed with 2-tailed independent-samples *t*-tests with 95% confidence levels, using IBM® SPSS® Statistics (version 22) software.

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Flow cytometry and isolation of WKM haematopoietic cells

Single cell suspensions of WKM cells were prepared from adult zebrafish kidneys of the required genotypes as described (Stachura and Traver, 2016). Flow cytometry analysis was performed on a FACS Aria II (BD Biosciences) after exclusion of dead cells by uptake of Hoechst dye (Hoechst 33342, H3570, ThermoScientific), as described (Traver et al., 2003). WKM cell counts were performed on a PENTRA ES60 (Hariba Medical) following the manufacturer's instructions. Note that the cell counter does not recognize the zebrafish nucleated erythrocytes, so these were excluded from this analysis. Cell counts for each genotype were analysed with 2-tailed paired-samples *t*-tests with 95% confidence levels, using a Mann-Whitney test for non-parametric distribution. The scatter plots were generated using GraphPad Prism 7 and show medians±SD.

May-Grunwald and Wright-Giemsa staining

Cell staining with May-Grunwald (MG) stain (Sigma MG500) and Giemsa (GIEMSA STAIN, FLUKA 48900) was performed on haematopoietic cell samples. After cytospin, slides are allowed to air-dry and were stained for 5 min at room temperature with a 1:1 mix of MG:distilled water. Next, slides were drained and stained with a 1:9 dilution of Giemsa:distilled water solution for 30min 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 (06522, Sigma) for imaging.

Whole mount *in situ* hybridization (ISH) and immunohistochemistry

Whole-mount ISH was carried out as described previously (Jowett and Yan, 1996), using probes for *kdrl*, *runx1, cmyb, gata2a , gata2b, rag1* (Kalev-Zylinska et al., 2002; Monteiro et al., 2016; Thompson et al., 1998; Willett et al., 1997), and *gfp* (Table S1). For conventional ISH embryos were processed, imaged and the ISH signal quantified as described (Dobrzycki et al., 2018). Briefly, the pixel intensity values were assessed for normal distribution with a Q-Q plot and transformed when necessary. Mean values (μ) of each experimental group were 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 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

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

unequal variances). For all these analyses, the IBM® SPSS® Statistics (version 22) or GraphPad Prism 8.0 package were used.
For the analysis of *cmyb* expression in the CHT at 4dpf, the embryos scored as 'high' or 'low' were tested for equal distribution between morphants and uninjected controls or among wild type, heterozygous and mutant genotypes with contingency Chi-squared tests, applying Continuity Correction for 2x2 tables, using IBM® SPSS® Statistics (version 22).

For fluorescent 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 (PBT) and directly processed for immunohistochemistry. Embryos were blocked in blocking buffer (5 % goat serum/0.3 % Triton X-100 in PBT) for 1 hour at RT before incubated with primary antibody against GFP (rabbit, 1:500, Molecular Probes), diluted in blocking buffer overnight at 4°C. Secondary antibody raised in goat coupled to AlexaFluor488 (Invitrogen) was used in 1:500 dilutions for 3h at RT. Hoechst 33342 was used as a nuclear counterstain.
Fluorescent images were taken on a Zeiss LSM880 confocal microscope using 40x or 63x oil immersion objectives. Images were processed using the ZEN software (Zeiss).

Widefield fluorescence microscopy

For widefield fluorescence microscopy, live embryos were anaesthetised with 160 µg/ml MS222 and mounted in 3% methylcellulose and imaged on a AxioLumar V.12 stereomicroscope (Zeiss) equipped with a Zeiss AxioCam MrM.

Counting of *itga2b*-GFP^{high} and *itga2b*-GFP^{low} cells

To count *itga2b*-GFP^{high} and *itga2b*-GFP^{low} cells in the CHT, Tg(*itga2b*:GFP;*kdrl*:mCherry); *gata2a*$^{\Delta i4/+}$ animals were incrossed 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 and fixed for 1h at room temperature in 4% PFA. Next, the tails were mounted on 35 mm glass bottomed dishes (MAtTEK) in 1 % low melt agarose and imaged using a 40x oil objective on an LSM880 confocal microscope (Zeiss). Cells in the CHT region were counted manually on Z-stacks as '*itga2b*:GFP^{low}' (HSPCs) or '*itga2b*:GFP^{high}' (thrombocytes). Genomic DNA from the heads was extracted and used for genotyping as described above. Cell counts for each genotype were analysed with 2-tailed paired-

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

samples *t*-tests with 95% confidence levels, using a Mann-Whitney test for non-parametric distribution. The scatter plots were generated using GraphPad Prism 8.0 and show medians±SD.

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Athanasiadis, E.I., Botthof, J.G., Andres, H., Ferreira, L., Lio, P., and Cvejic, A. (2017). Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. Nature Communications *8*, 2045.
Bassett, A.R., Tibbit, C., Ponting, C.P., and Liu, J.L. (2013). Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. Cell Rep *4*, 220-228.
Bertrand, J.Y., Chi, N.C., Santoso, B., Teng, S., Stainier, D.Y., and Traver, D. (2010). Haematopoietic stem cells derive directly from aortic endothelium during development. Nature *464*, 108-111.
Bertrand, J.Y., Kim, A.D., Violette, E.P., Stachura, D.L., Cisson, J.L., and Traver, D. (2007). Definitive hematopoiesis initiates through a committed erythromyeloid progenitor in the zebrafish embryo. Development *134*, 4147-4156.
Birnbaum, R.Y., Clowney, E.J., Agamy, O., Kim, M.J., Zhao, J., Yamanaka, T., Pappalardo, Z., Clarke, S.L., Wenger, A.M., Nguyen, L.*, et al.* (2012). Coding exons function as tissue-specific enhancers of nearby genes. Genome Res *22*, 1059-1068.
Boisset, J.C., van Cappellen, W., Andrieu-Soler, C., Galjart, N., Dzierzak, E., and Robin, C. (2010). In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. Nature *464*, 116-120.
Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods *10*, 1213-1218.
Burns, C.E., DeBlasio, T., Zhou, Y., Zhang, J., Zon, L., and Nimer, S.D. (2002). Isolation and characterization of runxa and runxb, zebrafish members of the runt family of transcriptional regulators. Exp Hematol *30*, 1381-1389.
Butko, E., Distel, M., Pouget, C., Weijts, B., Kobayashi, I., Ng, K., Mosimann, C., Poulain, F.E., McPherson, A., Ni, C.W.*, et al.* (2015). Gata2b is a restricted early regulator of hemogenic endothelium in the zebrafish embryo. Development *142*, 1050-1061.
Chiang, I.K., Fritzsche, M., Pichol-Thievend, C., Neal, A., Holmes, K., Lagendijk, A., Overman, J., D'Angelo, D., Omini, A., Hermkens, D.*, et al.* (2017). SoxF factors induce Notch1 expression via direct transcriptional regulation during early arterial development. Development *144*, 2629-2639.
Ciau-Uitz, A., Monteiro, R., Kirmizitas, A., and Patient, R. (2014). Developmental hematopoiesis: ontogeny, genetic programming and conservation. Exp Hematol *42*, 669-683.
Collin, M., Dickinson, R., and Bigley, V. (2015). Haematopoietic and immune defects associated with GATA2 mutation. British Journal of Haematology *169*, 173-187.
Davidson, A.J., and Zon, L.I. (2004). The 'definitive' (and 'primitive') guide to zebrafish hematopoiesis. Oncogene *23*, 7233-7246.
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.
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.
Eilken, H.M., Nishikawa, S., and Schroeder, T. (2009). Continuous single-cell imaging of blood generation from haemogenic endothelium. Nature *457*, 896-900.
Fitch, S.R., Kimber, G.M., Wilson, N.K., Parker, A., Mirshekar-Syahkal, B., Gottgens, B., Medvinsky, A., Dzierzak, E., and Ottersbach, K. (2012). Signaling from the sympathetic nervous system regulates hematopoietic stem cell emergence during embryogenesis. Cell Stem Cell *11*, 554-566.
Gagnon, J.A., Valen, E., Thyme, S.B., Huang, P., Akhmetova, L., Pauli, A., Montague, T.G., Zimmerman, S., Richter, C., and Schier, A.F. (2014). Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. PLoS One *9*, e98186.

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.
Gao, X., Wu, T., Johnson, K.D., Lahvic, J.L., Ranheim, E.A., Zon, L.I., and Bresnick, E.H. (2016). GATA Factor-G-Protein-Coupled Receptor Circuit Suppresses Hematopoiesis. Stem Cell Reports *6*, 368-382.
Gering, M., and Patient, R. (2005). Hedgehog signaling is required for adult blood stem cell formation in zebrafish embryos. Dev Cell *8*, 389-400.
Gillis, W.Q., St John, J., Bowerman, B., and Schneider, S.Q. (2009). Whole genome duplications and expansion of the vertebrate GATA transcription factor gene family. BMC Evol Biol *9*, 207.
Grass, J.A., Jing, H., Kim, S.I., Martowicz, M.L., Pal, S., Blobel, G.A., and Bresnick, E.H. (2006). Distinct functions of dispersed GATA factor complexes at an endogenous gene locus. Mol Cell Biol *26*, 7056-7067.
Gritz, E., and Hirschi, K.K. (2016). Specification and function of hemogenic endothelium during embryogenesis. Cell Mol Life Sci *73*, 1547-1567.
Howe, K., Clark, M.D., Torroja, C.F., Torrance, J., Berthelot, C., Muffato, M., Collins, J.E., Humphray, S., McLaren, K., Matthews, L.*, et al.* (2013). The zebrafish reference genome sequence and its relationship to the human genome. Nature *496*, 498-503.
Hsu, A.P., Johnson, K.D., Falcone, E.L., Sanalkumar, R., Sanchez, L., Hickstein, D.D., Cuellar-Rodriguez, J., Lemieux, J.E., Zerbe, C.S., Bresnick, E.H.*, et al.* (2013). GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. Blood *121*, 3830-3837, S3831-3837.
Hsu, A.P., McReynolds, L.J., and Holland, S.M. (2015). GATA2 deficiency. Current Opinion in Allergy and Clinical Immunology *15*, 104.
Hu, Y., Xie, S., and Yao, J. (2016). Identification of Novel Reference Genes Suitable for qRT-PCR Normalization with Respect to the Zebrafish Developmental Stage. PLoS One *11*, e0149277.
Jin, S.W., Beis, D., Mitchell, T., Chen, J.N., and Stainier, D.Y. (2005). Cellular and molecular analyses of vascular tube and lumen formation in zebrafish. Development *132*, 5199-5209.
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.
Jowett, T., and Yan, Y.L. (1996). Double fluorescent in situ hybridization to zebrafish embryos. Trends Genet *12*, 387-389.
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.
Kawakami, K., Takeda, H., Kawakami, N., Kobayashi, M., Matsuda, N., and Mishina, M. (2004). A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. Dev Cell *7*, 133-144.
Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res *12*, 996-1006.
Khandekar, M., Brandt, W., Zhou, Y., Dagenais, S., Glover, T.W., Suzuki, N., Shimizu, R., Yamamoto, M., Lim, K.C., and Engel, J.D. (2007). A Gata2 intronic enhancer confers its pan-endothelia-specific regulation. Development *134*, 1703-1712.
Kissa, K., and Herbomel, P. (2010). Blood stem cells emerge from aortic endothelium by a novel type of cell transition. Nature *464*, 112-115.
Kissa, K., Murayama, E., Zapata, A., Cortes, A., Perret, E., Machu, C., and Herbomel, P. (2008). Live imaging of emerging hematopoietic stem cells and early thymus colonization. Blood *111*, 1147-1156.
Lam, E.Y., Hall, C.J., Crosier, P.S., Crosier, K.E., and Flores, M.V. (2010). Live imaging of Runx1 expression in the dorsal aorta tracks the emergence of blood progenitors from endothelial cells. Blood *116*, 909-914.

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Lin, H.-F.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.
Ling, K.W., Ottersbach, K., van Hamburg, J.P., Oziemlak, A., Tsai, F.Y., Orkin, S.H., Ploemacher, R., Hendriks, R.W., and Dzierzak, E. (2004). GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells. J Exp Med *200*, 871-882.
Liu, J., Jiang, J., Wang, Z., He, Y., and Zhang, Q. (2016). Origin and evolution of GATA2a and GATA2b in teleosts: insights from tongue sole, Cynoglossus semilaevis. PeerJ *4*, e1790.
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.
Macaulay, I.C., Svensson, V., Labalette, C., Ferreira, L., Hamey, F., Voet, T., Teichmann, S.A., and Cvejic, A. (2016). Single-Cell RNA-Sequencing Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells. Cell Rep *14*, 966-977.
Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol *31*, 833-838.
Meadows, S.M., Myers, C.T., and Krieg, P.A. (2011). Regulation of endothelial cell development by ETS transcription factors. Semin Cell Dev Biol *22*, 976-984.
Medvinsky, A., Rybtsov, S., and Taoudi, S. (2011). Embryonic origin of the adult hematopoietic system: advances and questions. Development *138*, 1017-1031.
Meng, A., Tang, H., Ong, B.A., Farrell, M.J., and Lin, S. (1997). Promoter analysis in living zebrafish embryos identifies a cis-acting motif required for neuronal expression of GATA-2. Proc Natl Acad Sci U S A *94*, 6267-6272.
Monteiro, R., Pinheiro, P., Joseph, N., Peterkin, T., Koth, J., Repapi, E., Bonkhofer, F., Kirmizitas, A., and Patient, R. (2016). Transforming Growth Factor beta Drives Hemogenic Endothelium Programming and the Transition to Hematopoietic Stem Cells. Dev Cell *38*, 358-370.
Murayama, E., Kissa, K., Zapata, A., Mordelet, E., Briolat, V., Lin, H.F., Handin, R.I., and Herbomel, P. (2006). Tracing hematopoietic precursor migration to successive hematopoietic organs during zebrafish development. Immunity *25*, 963-975.
Patterson, L.J., Gering, M., Eckfeldt, C.E., Green, A.R., Verfaillie, C.M., Ekker, S.C., and Patient, R. (2007). The transcription factors Scl and Lmo2 act together during development of the hemangioblast in zebrafish. Blood *109*, 2389-2398.
Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y.*, et al.* (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell *154*, 1380-1389.
Rodrigues, N.P., Janzen, V., Forkert, R., Dombkowski, D.M., Boyd, A.S., Orkin, S.H., Enver, T., Vyas, P., and Scadden, D.T. (2005). Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. Blood *106*, 477-484.
Sanalkumar, R., Johnson, K.D., Gao, X., Boyer, M.E., Chang, Y.I., Hewitt, K.J., Zhang, J., and Bresnick, E.H. (2014). Mechanism governing a stem cell-generating cis-regulatory element. Proc Natl Acad Sci U S A *111*, E1091-1100.
Sood, R., English, M.A., Belele, C.L., Jin, H., Bishop, K., Haskins, R., McKinney, M.C., Chahal, J., Weinstein, B.M., Wen, Z.*, et al.* (2010). Development of multilineage adult hematopoiesis in the zebrafish with a runx1 truncation mutation. Blood *115*, 2806-2809.
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.
Stachura, D.L., and Traver, D. (2016). Cellular dissection of zebrafish hematopoiesis. Methods Cell Biol *133*, 11-53.
Swiers, G., Baumann, C., O'Rourke, J., Giannoulatou, E., Taylor, S., Joshi, A., Moignard, V., Pina, C., Bee, T., Kokkaliaris, K.D.*, et al.* (2013). Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. Nat Commun *4*, 2924.

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Taylor, J.S., Braasch, I., Frickey, T., Meyer, A., and Van de Peer, Y. (2003). Genome duplication, a trait shared by 22000 species of ray-finned fish. Genome Res *13*, 382-390.
Taylor, J.S., Van de Peer, Y., Braasch, I., and Meyer, A. (2001). Comparative genomics provides evidence for an ancient genome duplication event in fish. Philos Trans R Soc Lond B Biol Sci *356*, 1661-1679.
Thompson, M.A., Ransom, D.G., Pratt, S.J., MacLennan, H., Kieran, M.W., Detrich, H.W., 3rd, Vail, B., Huber, T.L., Paw, B., Brownlie, A.J.*, et al.* (1998). The cloche and spadetail genes differentially affect hematopoiesis and vasculogenesis. Dev Biol *197*, 248-269.
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.
Tsai, F.Y., Keller, G., Kuo, F.C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F.W., and Orkin, S.H. (1994). An early haematopoietic defect in mice lacking the transcription factor GATA-2. Nature *371*, 221-226.
Walsh, D.M., Klyubin, I., Fadeeva, J.V., Cullen, W.K., Anwyl, R., Wolfe, M.S., Rowan, M.J., and Selkoe, D.J. (2002). Naturally secreted oligomers of amyloid beta protein potently inhibit hippocampal long-term potentiation in vivo. Nature *416*, 535-539.
Westerfield, M. (2007). The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio). 5th edn (Eugene, OR: Univ of Oregon Press).
Wilkinson, R.N., Pouget, C., Gering, M., Russell, A.J., Davies, S.G., Kimelman, D., and Patient, R. (2009). Hedgehog and Bmp polarize hematopoietic stem cell emergence in the zebrafish dorsal aorta. Dev Cell *16*, 909-916.
Willett, C.E., Zapata, A.G., Hopkins, N., and Steiner, L.A. (1997). Expression of zebrafish rag genes during early development identifies the thymus. Dev Biol *182*, 331-341.
Wlodarski, M.W., Collin, M., and Horwitz, M.S. (2017). GATA2 deficiency and related myeloid neoplasms. Semin Hematol *54*, 81-86.
Wlodarski, M.W., Hirabayashi, S., Pastor, V., Stary, J., Hasle, H., Masetti, R., Dworzak, M., Schmugge, M., van den Heuvel-Eibrink, M., Ussowicz, M.*, et al.* (2016). Prevalence, clinical characteristics, and prognosis of GATA2-related myelodysplastic syndromes in children and adolescents. Blood *127*, 1387-1397; quiz 1518.
Wozniak, R.J., Keles, S., Lugus, J.J., Young, K.H., Boyer, M.E., Tran, T.M., Choi, K., and Bresnick, E.H. (2008). Molecular hallmarks of endogenous chromatin complexes containing master regulators of hematopoiesis. Mol Cell Biol *28*, 6681-6694.
Xu, H., Li, C., Zeng, Q., Agrawal, I., Zhu, X., and Gong, Z. (2016). 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.
Yang, L., Rastegar, S., and Strahle, U. (2010). Regulatory interactions specifying Kolmer-Agduhr interneurons. Development *137*, 2713-2722.
Yang, Z., Jiang, H., Zhao, F., Shankar, D.B., Sakamoto, K.M., Zhang, M.Q., and Lin, S. (2007). A highly conserved regulatory element controls hematopoietic expression of GATA-2 in zebrafish. BMC Dev Biol *7*, 97.
Zhu, C., Smith, T., McNulty, J., Rayla, A.L., Lakshmanan, A., Siekmann, A.F., Buffardi, M., Meng, X., Shin, J., Padmanabhan, A.*, et al.* (2011). Evaluation and application of modularly assembled zinc-finger nucleases in zebrafish. Development *138*, 4555-4564.

bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A Tg(kdrl:GFP) ATAC-seq

FACS 26hpf

2%

94%

488-530/30-A

FSC-A (x1,000)

Chromosome 11

gata2a

i4

reads

kdrl:GFP+ (endothelium)

kdrl:GFP- (non-endothelial)

intron

UTR

Medaka
Stickleback
Tetraodon
Fugu
X. tropicalis
Mouse
Human

Ets E-box GATA

400bp

150bp

```
Human     -TATTATTTTTCCATGGAGTCAC----CTATACTGTGTATTTTCATTTGAGTGATTTT--AAAAAAATGCCCTTTCGGATCTCCTGC
Chicken   --TATTATTTTCCATGGAGTCAC----CTATACTTTGTATTTTCATTCGAGCGATTTA--AAAAAAATGTCCTTTCTGAGCTCCTGG
Mouse     -TATTATTTTTCCATGGAGTCAC----CTATACTGTGTATTTTCATTTGAGTGATTTTTTAAAAAAATGTCCTTTCGGATCTCCTGC
Opossum   TATTATTTTTTCCATGGAGTCAC----CTATACTGTGTATTTTCATTTGAGTGATTTTT--TTTAAATGCCCTTTCTAAGCTCCTGG
Xenopus   --ATTCATTTTCCATGGAGTCAC----CTACACTCTGTACTTTGGGGCCTTTTCTTTGT-TGTAAGTGGCCCCATGTGAGATCATGC
Zebrafish -ACAGGGTTTTCCATGGAGTCAGTGTGCTGGCCCATGCTTTTGCATTTTGGGTTGAAGTGTTTTCAGAGCGCTGTTCAAAGGCCATC

Human     CGG-AGTTTCCTATCCGGACATCTGCAGCCTGAAGATAAGGAAACTTCGTGTATCTGT--TTCCGGACTCTGCGAGTTTTT--AGAG
Chicken   TAGTAGTTTCCTATCCGGATATCTGTACGTTAAAGATAAGGAAACTTTGTGTATCTGT--TTCCTGACTCTAAAAGCCTTAG-AGAG
Mouse     CGG-AGTTTCCTATCCGGACATCTGCAGCCGGTAGATAAGGAAACTTCGTGTATCTGT--TTCCGGAC-CGGCAAGTTTTC--AGAG
Opossum   CCAGAGTTTCCTATCCGGACATCTGCAGGCTGAAGATAAGGAAACTTAGTGTATCTGT--ATCCGGACTCTGCGAGCTTTGAGAGTG
Xenopus   TAGTGGCTTCCTATCCGGATATGTGCATTGTGCTGATAAGGAAACTTTGTGTCTCTGT--TTCCGGACATTAAATAACTTTAGAGAG
Zebrafish CAGTTGTTTCCTATCCGGATATCTGCCCGGGCCCGATAAGGAAGCCAGGCTCCCGGGTGTTTCCGTCTTCAACAAACCCGGCTGAAT
```

gata2a i4-1.1kb:EGFP

i4 GFP

B live GFP 27hpf 300μm C

C live GFP 27hpf 100μm D/E

D GFP DA PCV 25hpf 10μm

D' *gfp* 25hpf 10μm

D" merged with Hoechst 25hpf 10μm

E GFP DA PCV 25hpf 10μm

E' *runx1* 25hpf 10μm

E" merged with Hoechst 25hpf 10μm

F

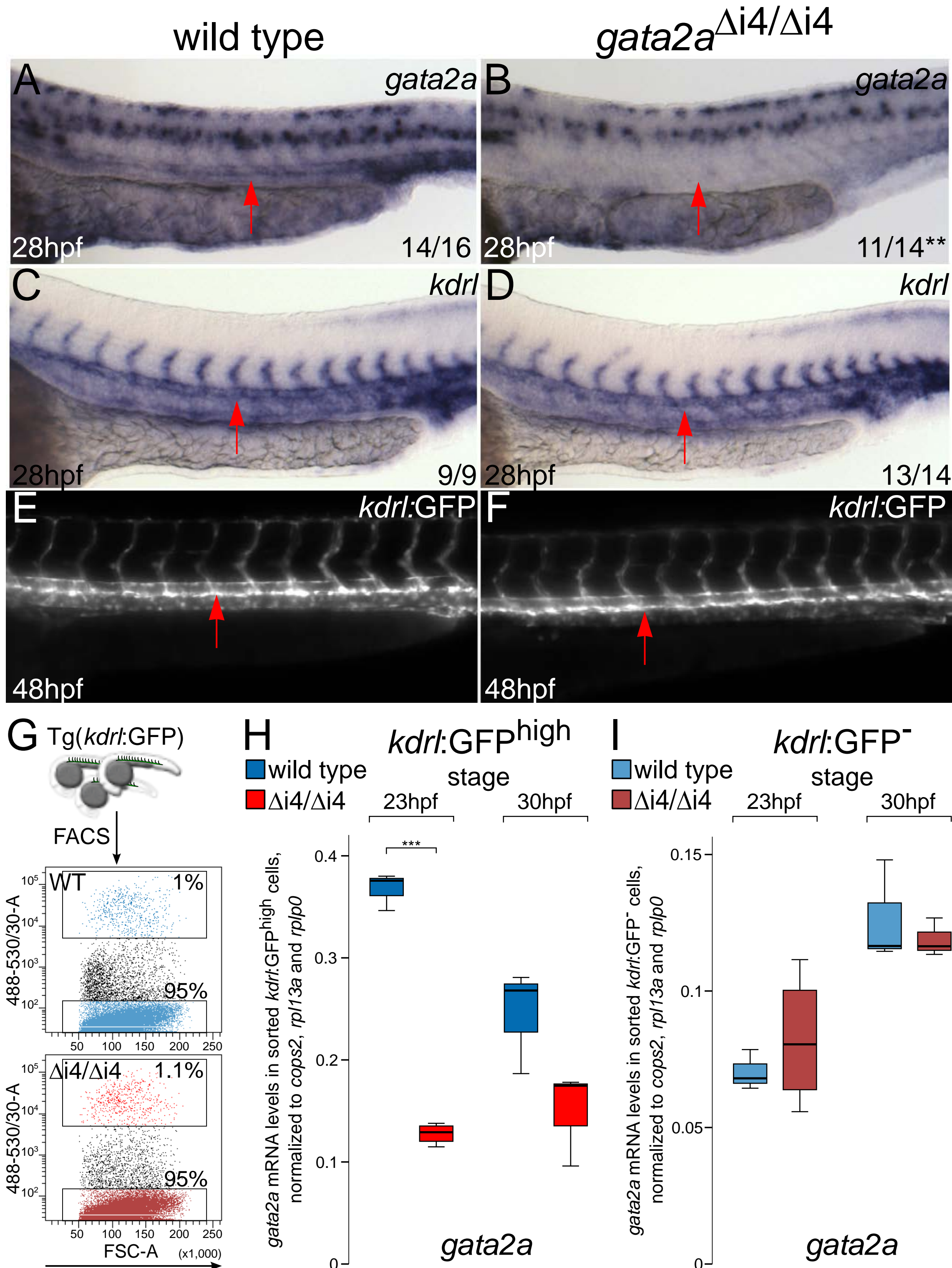
GFP+/- cells within *runx1*+ population in DA

100% 75% 50% 25% 0

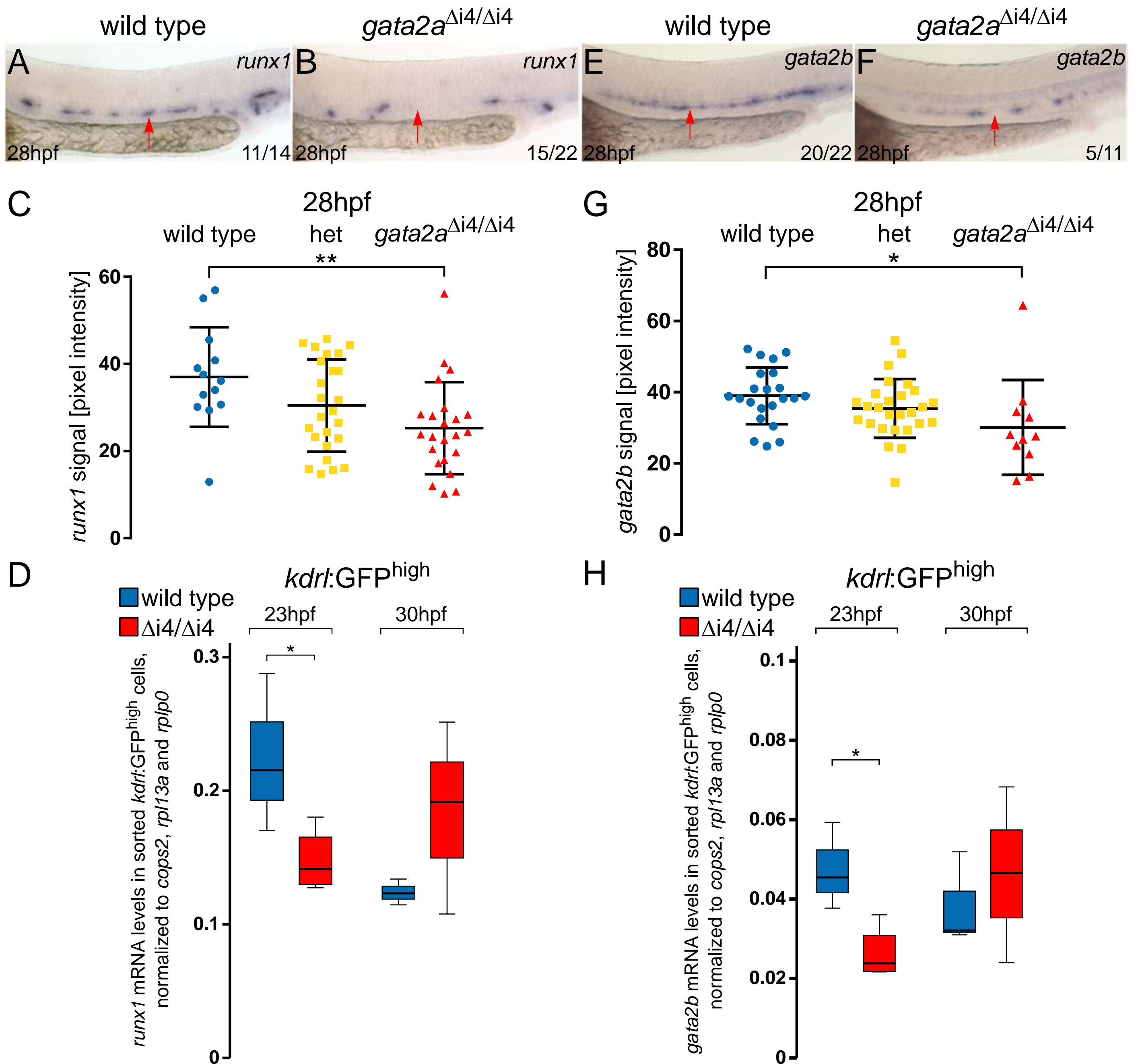
GFP-

GFP+

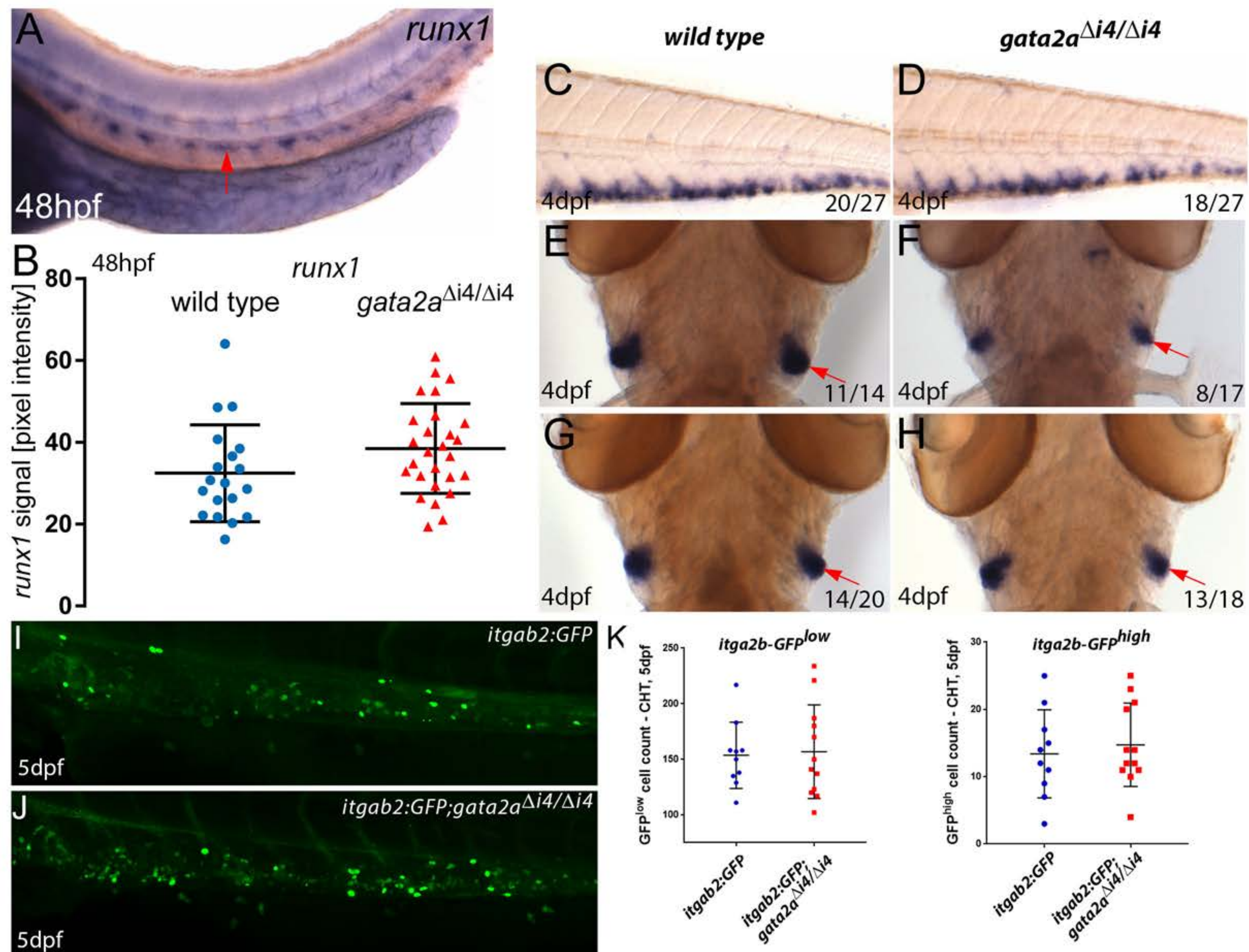
bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



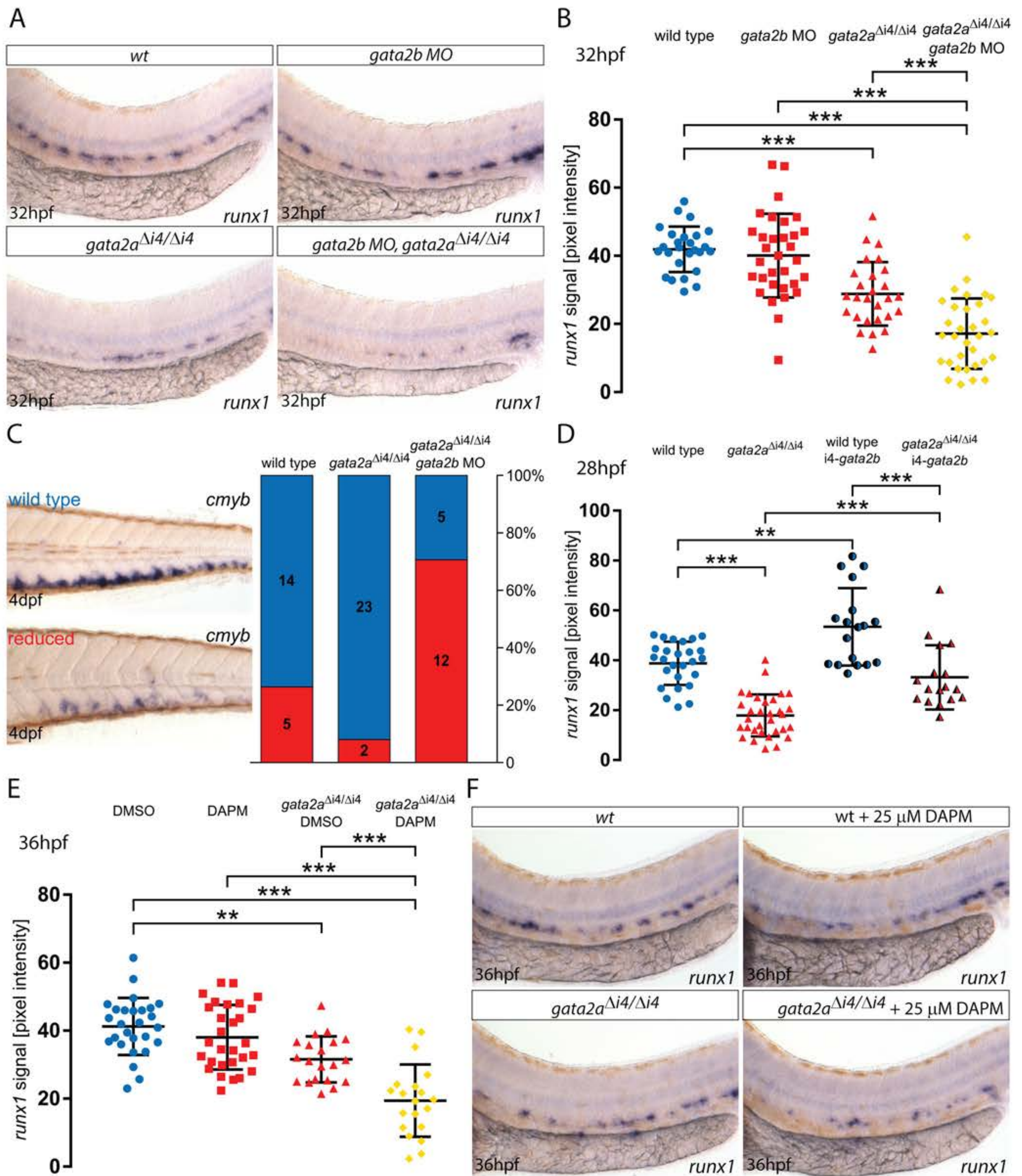
bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/516203; this version posted January 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.