UNIVERSITY OF CALIFORNIA, SAN DIEGO

Phylogenetic Comparative Analysis to Dissect the Hematopoietic Mechanisms of GATA2

A Thesis submitted in partial satisfaction of the requirements for the degree Master of Science

in

Biology

by

Kevin King-Yiu Ng

Committee in charge:

Professor David Traver; Chair
Professor Nicholas Spitzer
Professor Dong-Er Zhang

2015

The Thesis of Kevin King-Yiu Ng is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2015

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank Bart Weijts for his guidance and advice throughout this project; Emerald Butko for providing the basic knowledge and foundations of my research experience; and David Traver for the providing me the opportunity to learn from the research within his lab. I also thank the entire Traver Lab for making this a unique experience over the past three years. Thank you to my friends and family who have given me support and encouragement throughout my college career.

ABSTRACT OF THE THESIS

Phylogenetic Comparative Analysis to Dissect the Hematopoietic Function of GATA2

by

Kevin King-Yiu Ng

Master of Science in Biology

University of California: San Diego, 2015

David Traver, Chair

Treatment of hematological diseases such as leukemia and lymphoma rely on the transplantation of hematopoietic stem cells (HSCs) derived from limited numbers immune-compatible donors. Recent advances in stem-cell biology have enabled the generation of pluripotent stem cells (iPSCs) from terminally differentiated cells of healthy patient tissue. This provides the opportunity to generate *bona fide* HSCs, circumventing the limitations of donor-derived transplantation; however, inadequate knowledge of the innate molecular mechanisms hinders progress to this end.

A key hematopoietic regulator in mammals, GATA2, is indispensable for HSC formation and maintenance. Therefore, dissection of its molecular mechanism and function in HSC development would be a crucial step in *in vitro* synthesis of HSCs. Unfortunately, GATA2 also maintains functional roles in endothelial development making it difficult to isolate its specific role towards HSC specification. Fortunately, recent studies have identified two paralogs of GATA2 in zebrafish, *gata2b* and *gata2a,* in which the hematopoietic and endothelial functions seem to have segregated. Using a comparative approach of the similarities shared between GATA2 and its paralogs, we can provide further insight into GATA2 function in HSC development.

Our current studies have shown that *gata2b* and *gata2a* function in distinct manners. Furthermore, we have shown that mutations within the N-terminal zinc finger binding domain of GATA2 and its paralogs affect the function of the protein. These results indicate the *gata2b* zinc finger binding domain may result in its hematopoietic specific identity.

INTRODUCTION

Generation of patient specific hematopoietic stem cells has become a possibility due to recent advances in stem cell research. With the advent of induced pluripotent stem cells from terminally differentiated cells of healthy patient tissue (Takahashi and Yamanaka, 2006; Malik and Rao., 2013), the limitations of donor derived HSCs can be circumvented. Unfortunately, limited knowledge of the innate molecular mechanisms governing HSC formation prevents the formation of transplantable HSCs capable of repopulating the host blood system. Therefore, further investigation into the embryonic signaling pathways governing HSC development is necessary.

Significant progress has been made in utilizing transcription factors to generate HSCs *in vitro* (Aggarwal et al., 2010) given their potent ability in regulating cellular identity. However, these "synthetic" stem cells are unable to repopulate the blood system after transplantation into sublethal irradiated mice. This indicates that more potent transcription factors or a more specific driver of hematopoiesis is required. Studies into a variety of hematopoietic transcription factors and the regions in which they bind have implicated GATA2 is a key component in many of the transcriptional complexes formed (Wilson et al., 2010). GATA2 belongs to the GATA transcription factor family (Orkin et al., 1995), which have conserved dual zinc finger binding domains modulating interaction with a conserved WGATAR motif within the genome. Research into acute myeloid leukemia (AML) patients has identified mutations in GATA2 as a contributing factor to onset of the disease (Zhang et al., 2007).

Furthermore, specific mutations of the GATA2 zinc finger binding domain have been implicated to result in AML (Gao et al., 2014). These polymorphisms and deletions within the zinc finger have been shown to inhibit the activation of responsive enhancer elements. These studies indicate an importance of zinc finger structure in regulating the ability of transcription factors to bind and transactivate key gene pathways.

The ability of transcription factors to bind specific consensus sequences is modulated by the interactions it forms with other proteins. Together, the transcriptional complexes formed and availability of binding sites modulated by chromatin occupancy (Bresnick et al., 2012; Wilson et al., 2010) determines cellular fate. GATA2 transcriptional complexes have been shown to regulate a multitude of enhancer sequences resulting to functional roles in differentiation of hematopoietic stem and progenitor cells (Gottgens et al., 2002; Nottingham et al., 2007) as well as vascular specification (Linnemann et al., 2011). Furthermore, loss of function experiments have characterized and identified GATA2 as a key component in HSC development.

Amongst the GATA factors, *GATA2* is unique in its requirement for HSC formation and maintenance (Orkin et al., 1995). Complete knockout of GATA2 in mice has shown a loss of hematopoietic stem and progenitor cells (Tsai et al., 1994). Global knockout of a gene generally results in dramatic phenotypes due to functional roles in various aspects. Therefore, given the endothelial nature from which HSCs arise (Bertrand et al., 2010) targeted deletion within the endothelium was performed (Lim et al., 2012). Loss of HSCs as a result indicates that GATA2 functions within the

endothelium to specify HSC formation. Additionally, through the formation of different complexes, GATA2 regulates the distinction between an endothelial or hematopoietic cell fate (Shi et al., 2014). Knockout within HSCs results in increased apoptosis demonstrating that GATA2 is critical for the maintenance of HSCs (Lim et al., 2012). Together, these studies have shown that GATA2 is a necessary component in nearly all aspects of HSC development.

Interestingly, knockout of key enhancer elements within GATA2 results in hemorrhaging and anemia within the embryo (Johnson et al., 2012, Gao et al., 2013). This suggests a capacity of GATA2 in maintaining vascular integrity. Although GATA2 is critically important for hematopoiesis and a prime candidate for inducing HSC formation *in vitro,* dissecting its hematopoietic function has proven difficult. In addition, the mouse model has shown significant limitation in understanding the molecular mechanisms of such a key regulator in HSC development. However, recent research within zebrafish has provided the opportunity to dissect the endothelial and hematopoietic function of GATA2.

In the evolution of boney, jawed, ray-finned fish of the teleost lineage, a chromosomal duplication of *GATA2* has occurred (Gillis et al., 2009). Furthermore, because the process of hematopoiesis is conserved across vertebrate species (Bertrand et al., 2010) research within the zebrafish model has proven viable to the understanding of HSC development. Additionally, the optical transparency of zebrafish embryos enables *in vivo* imaging of this process in a non-invasive manner. Of the two zebrafish

paralogs, *gata2a* has been shown to be expressed throughout the vasculature (Brown et al., 2000; Detrich et al., 1995) and loss of h $gata2a$ resulted in defects in vascular integrity and circulation (Zhu et al., 2011) underlining its endothelial specific function. In contrast, *gata2b* expression is restricted to the hemogenic endothelium and knockdown of *gata2b* results in loss of hematopoietic stem cells (Butko et al., 2015). Together, these studies indicate the hematopoietic and endothelial functions of *GATA2* have been segregated between *gata2b* and *gata2a.*

Using the paralogs of *GATA2* within zebrafish, we can make a comparative analysis of the similarities shared between *GATA2* and either *gata2b* or *gata2a* in an effort to isolate the hematopoietic and endothelial components comprising *GATA2* function. To this extent, our current research has looked into multiple aspects of their function including, gene regulation, transactivation potential, protein-protein interaction, and protein stability. Here we show that *gata2a* and *gata2b* are indeed functionally unique. Moreover, the differences in their zinc finger binding domains are a key component in regulating their distinct functional roles.

MATERIALS AND METHODS

CRISPR Mutant Generation

Oligonucleotides used for gRNA generation was designed using the ZiFiT program. Oligos were selected based on lowest number of off-target results. Oligos were annealed and cloned into the T7cas9sgRNA2 vector (Addgene) following the protocol detailed by Chen and Wente. SgRNA was synthesized using the MEGAshortscript T7 kit (Ambion/Invitrogen) and purified using *mir*Vana miRNA Isolation Kit (Ambion/Invitrogen). Capped Cas9 RNA was synthesized from pCS2-nls-zCas9-nls(Addgene) plasmid using mMESSAGE mMACHINE SP6 (Invitrogen) and purified using RNeasy Mini Kit (Qiagen). A mix of 0.05% phenol red, 150 ng/ul Cas9 RNA and 100 ng/ul of gRNA was injected at 1 nl directly into the cell of 1-cell-stage embryos. Injected embryos were genotyped with gata2b Check primers designed using Primer3 to yield ~500bp fragments. See Table 1 for primer sequences.

Cell Culture

Human Embryonic Kidney 293 Cells (HEK) were cultured in DMEM (Corning) supplemented with 10% fetal bovine serume (FBS) (Corning) and 2% Penicillin/Streptomycin (10U/mL stock). Human Umbilical Vein Endothelial Cells (HUVEC) from Lonza were cultured using the EGM-2 BulletKit (Lonza). Both human cell culture line were incubated at 37^0C with 5% CO_2. Zebrafish Stromal Tissue

(ZEST) are cultured in DMEM supplemented with 500 mL Leibovitz's L-15 Medium (Life Technology), 150 mg sodium bicarbonate, 15mL 1M HEPES, 2% Penicillin/Streptomycin (5000U/mL stock), 10 mL L-glutamine (200mM), 10% FBS, and 2ml Gentamicin Sulfate (50mg/mL).

Plasmid Constructs

Full length ***homo sapien GATA2*** (Ensembl Ascension Number ENSG00000179348), *danio rerio gata2a* (Ensembl Ascension Number ENSDARG00000059327), and *danio rerio gata2b* (Ensembl Ascension Number ENSDARG00000009094) were cloned into PCS2+ vector. HA tag was cloned into PCS2+ gata2a plasmid and Flag tag was cloned into PCS2+ gata2b vector. GFP was fused to all construct along with a flex-linker using the Gibson Assembly Mastermix (NEB) following the primer design protocol as detailed by Gibson et al.

cDNA synthesis and quantitative PCR

RNA was isolated from cells using RNeasy Mini Kit (Qiagen) following manufacturer's protocol. cDNA was synthesized from isolated RNA using Superscript III 1st Strand (Life Technologies) using random hexamers following manufacturer's instructions. Quantitative PCR was run on a C1000 Thermal Cycler (BioRad) using SsoADV SYBR Green Mix (BioRad). Primers used shown in Table 1.

Immunoprecipitation

HEK293 cells were counted using a bright line hemacytometer (Hausser Scientific) and addition of dead cell discrimination dye Trypan Blue (Invitrogen). Cells were seeded at 1.2 x 10^6 cells per dish in 60mm tissue culture treated dishes (Corning) and cultured with 5 mL of media to 80% confluency. Transfection followed manufacturer's instruction using 5 uL Lipofectamine 2000 (Life Technologies) and a total 6.6 ug plasmid. Cells were harvested using 500 uL RiPA buffer with Protease Inhibitor and pre-cleared with 20 ul Protein A/G Beads (Milipore) for 30 min at 4^0C. Immunoprecipitation was performed using 1 µg antibody and incubated at 4^0C overnight with rotation. Precipitate was washed three times in RiPA buffer and resuspended 50 µL sample buffer.

Western Blot

Immunoprecipitated protein (as described under Immunoprecipitation) or cells harvested (as described under Cycloheximide) were loaded in NuPAGE 4-12% Bis-Tris Gel (Novax). Antibodies used are HA (Sigma, H3663), Flag (Sigma, F1804), GATA2 (Santa Cruz, sc-9008X), GAPDH (Cell Signal, 2118S), TAL1 (Santa Cruz, sc-12984X), RUNX1 (abcam, ab23980), FLI1 (abcam, ab15289), GFP (Aves Labs, GFP-1020), Chicken IgG HRP-linked (Sigma, A9792), Mouse IgG HRP-linked (BioLegend, 405308) and Rabbit IgG HRP-linked (Invitrogen, G21234)

Luciferase Reporter Assay

HEK293 cells (60%-70% confluent) were transfected using Lipofectamine 2000 following manufacturer's instructions. Per well (96-well), cells were transfected with 0.5 µL Lipofectamine 2000, 250 ng reporter plasmid, 25 ng Null renilla, and 50 ng expression or control plasmid. Null renilla was used for normalization of luciferase readout. Reporter activity was measured using Promega Dual Luciferase Reporter (Promega, E1910) and luminometer. Reporter plasmid was obtained from Hamish Scott (Center for Cancer Biology, Austrailia)

Cycloheximide

HEK293 cells (60%-70% confluent) were transfected using Lipofectamine 2000 following manufacturer's instructions. Per well (24-well), cells were transfected using 1.5 µL Lipofectamine 2000 and 500 ng expression plasmid. Cells were treated with 10 µM Cycloheximide (Sigma, C7698) 24 hours post transfection for 15, 30, and 60 min intervals. Cells were then lysed using RiPA lysis buffer and cell lysates western blotted. Protein quantification was done using ImageJ. Each condition was normalized to GAPDH protein levels and fold change was calculated relative to protein levels at time point 0.

Fluorescence Activated Cell Sorting

Transfected HUVEC was trypsinized, centrifuged and resuspended in DPBS w/ 2% FBS and sorted on a FACSAria (BD Biosciences) for GFP.

Table 1: Primers Used

Gata2b CRISPR	TAGGATGCCCCAGCGGA	AAACTCCGCTGGGGCAT
Gata2b Check	GCGAACTCCTGGTTTCACAT	TGCGTCTGGCTGTAAGACAT
GATA2 qPCR	CCACCCAAAGAAGTGTCTCC	GCCACTTTCCATCTTCATGC
Gata2a qPCR	TCTTGAATCACTTGGACTCG	GGACTGTGTATGAGGTGTGG
Gata2b qPCR	ACCACCACACTCTGGAGAC	CTGTTGCGTGTCTGAATACC
hs RUNX1 qPCR	GTCGGTCGAAGTGGAAGAGG	GATGGCTCTGTGGTAGGTGG
dr runx1 qPCR	ACGCTACCACACATACCTGC	CAGGTGATATGGGGACGAGC
hs RAB44 qPCR	ATGGACTGTGAGGAGGAACG	TTGCATCCTGAGTGACCTGG
dr rab44 qPCR	TGTCGTCATGATGTTGCTCG	GACACATTAGCCCCGTTTGC
hs TAL1 qPCR	CCTTCCCTATGTTCACCACC	CGGCTGTTGGTGAAGATACG
hs KDR qPCR	CATCGAAGTCTCATGCACGG	TCTGCGGATAGTGAGGTTCC
dr kdr qPCR	ACAGAGAAAGATGCTGGGAAT	GCTACTGCCGTACATGTGGA
dr kdrl qPCR	GATGATCACCCACAGCCTTT	ATCCGCAGAGTGGTTTCATC
hs Bactin qPCR	GATCGGCGGCTCCATCCTG	GACTCGTCATACTCCTGCTTGC
hs RSP18 qPCR	AGTTCCAGCATATTTTGCGAG	CTCTTGGTGAGGTCAATGTC
dr Bactin qPCR	CGTCTGGATCTAGCTGGTCGTGA	CAATTTCTCTTTCGGCTGTGGTG

List of primers used for CRISPR gRNA synthesis and qPCR

RESULTS

Figure 1: Zinc finger is highly conserved amongst GATA2 transcription factors

Amino acid sequence alignment of human and murine *GATA2* against zebrafish *gata2a* and *gata2b*. Conserved residues across all 4 transcription factors are highlighted in blue. Highlighted in red are residues conserved between *GATA2* and *gata2b*. Residues conserved between *GATA2* and *gata2a* are highlighted in green. Residues comprising N-terminal and C-terminal zinc fingers are denoted as Zinc Finger I and Zinc Finger II, respectively.

The C-terminal GATA2 Zinc Finger domain is highly conserved between species

Within the teleost lineage, a chromosomal duplication event of the GATA2 transcription factor has resulted in two paralogs, *gata2a* and *gata2b (*Gilles et al. 2009). These events typically results in paralogs with similar functions, however, the functions of the ancestral gene may also be divided between the two paralogs. The endothelial specific expression and function of *gata2a* (Brown et al., 2000; Detrich et al., 1995) and the hematopoietic expression and requirement of *gata2b* (Butko et al., 2015) suggests that the function of GATA2 has indeed been divided between the two paralogs. We hypothesize that, to exert a different function, *gata2a* and *gata2b* should possess different domains within their protein sequence. To gain insight into such a possibility, we aligned the amino acid sequence of human and murine GATA2 against the zebrafish orthologs *gata2a* and *gata2b*. Although, gata2a and gata2b share 73.33% and 58.03% sequence identity with GATA2 respectively (Figure 1), we could not identify unique protein domains in either *gata2a* or *gata2b*. However interestingly, we found a single amino acid substitution within the second zinc finger of *gata2a* and a two amino acid substitution for *gata2b*. This finding might indicate that possible difference in function between *gata2a* and *gata2b* may be due to these amino acid substitutions as transcriptional activity depends on sequence and conformation of these zinc fingers. Importantly, for both *gata2a* and *gata2b* these amino acid substitutions have resulted in the presence of a glycine. Glycines are the smallest amino acid building blocks of life and are used to create sharp turns within the protein conformation. Therefore, we hypothesize that the C-terminal Zinc finger domain of both gata2a and gata2b undergo a conformational change, thereby suggesting a high

probability for functional deviation. Our hypothesis is strengthened by the fact that mutations within the human GATA2 C-terminal zinc finger binding domain is often found in several blood borne diseases (Gao et al. 2014; Hahn et al. 2011).

Together, *in silico* analysis of the C-terminal zinc finger domain of the zebrafish gata2a and gata2b revealed an amino substitution that results with high probability in conformational, and thereby, functional changes.

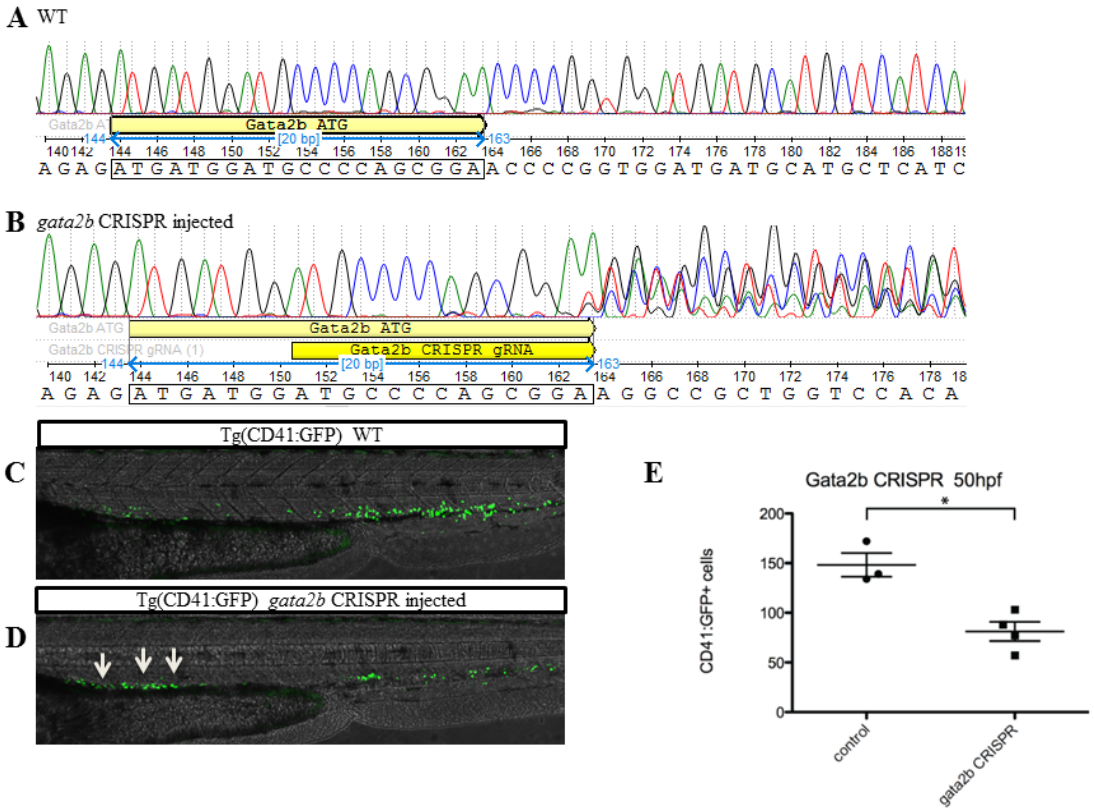


Figure 2:*gata2b* mutants result in a decrease of HSCs

Injection of *gata2b* gRNA with Cas9 mRNA into single cell embryos. (A) Sequencing result of wild-type uninjected embryo. (B) Sequencing result of embryo injected with *gata2b* gRNA and Cas9 mRNA mix. Imaging of CHT done through confocal microscopy of (C) Tg (*CD41:GFP)* uninjected embryo at 50 hpf and (D) Tg(*CD41:GFP)* injected with *gata2b* gRNA/Cas9 mRNA mix. (E) Graph of counted GFP positive cells in the CHT region of 3 uninjected and 4 injected embryos. * p=0.011 through Student's T-Test

***gata2b* mutants result in a decrease of HSCs**

Recently, it was shown that morpholino mediated knockdown of gata2b resulted in a markable decrease in HSC numbers (Butko et al. 2015), however, due to possible off-target effects of morpholinos (Janet Heasman, 2002), we generated mutants of *gata2b* by utilizing the CRISPR/Cas9 method (Jao et al., 2013; Hruscha et al., 2013). To ensure full truncation of the gata2b protein, the CRISPR gRNA was designed to target several base pairs downstream of ATG start site (Figure 2A). Sequencing of primary injected embryos (F_0) showed a sequence disruption exactly at the PAM site, the Cas9 recognition site, (indicated by red arrow) compared to wild-type (WT) uninjected embryos (Figure 2B and 2C). Sequence disruption was observed in ten injected embryos, showing the robustness and specificity of the CRISPR/Cas9 method.

Because both gRNA and Cas9 are injected in the zebrafish embryo at the single cell stage and a high degree of mutation was observed, we tested whether the F_0 embryos showed a similar phenotype to that of *gata2b* morphants. To this extent, we injected Tg(CD41:eGFP) (Lin et al., 2005) embryos and quantified the GFP positive cells within the dorsal aorta (DA) and caudal hematopoietic tissue (CHT) at 50 hpf. Within the un-injected embryos, a high number of GFP positive single cells are observed within the CHT, a typical behavior of HSCs prior to migration and seeding of the thymus and the kidney (Figure 2B). In contrast, *gata2b* gRNA injected embryos showed a noticeable decrease in GFP positive cells within the aorta and CHT (Figure 2C). We quantified the total number of GFP positive cells in three un-injected embryos and four injected embryos, observing an average of 150 cells and 75 cells, respectively (Figure 2E). Student's T-test was performed showing a significant ($p = 0.011$) decrease of GFP positive cells in *gata2b* gRNA injected embryos. Together, our results

demonstrate that mutating the *gata2b* gene results in a phenocopy of the *gata2b* morphants, thereby validating not only the *gata2b* morpholino, but also illustrating the usefulness of the CRISPR/Cas9 method in primary injected embryos. Once a stable line has been established, these mutants, in conjunction with the *Tg(gata2b:gal4)* (Butko et al., 2015), will be essential for performing rescue experiments with either human GATA2 or Gata2a.

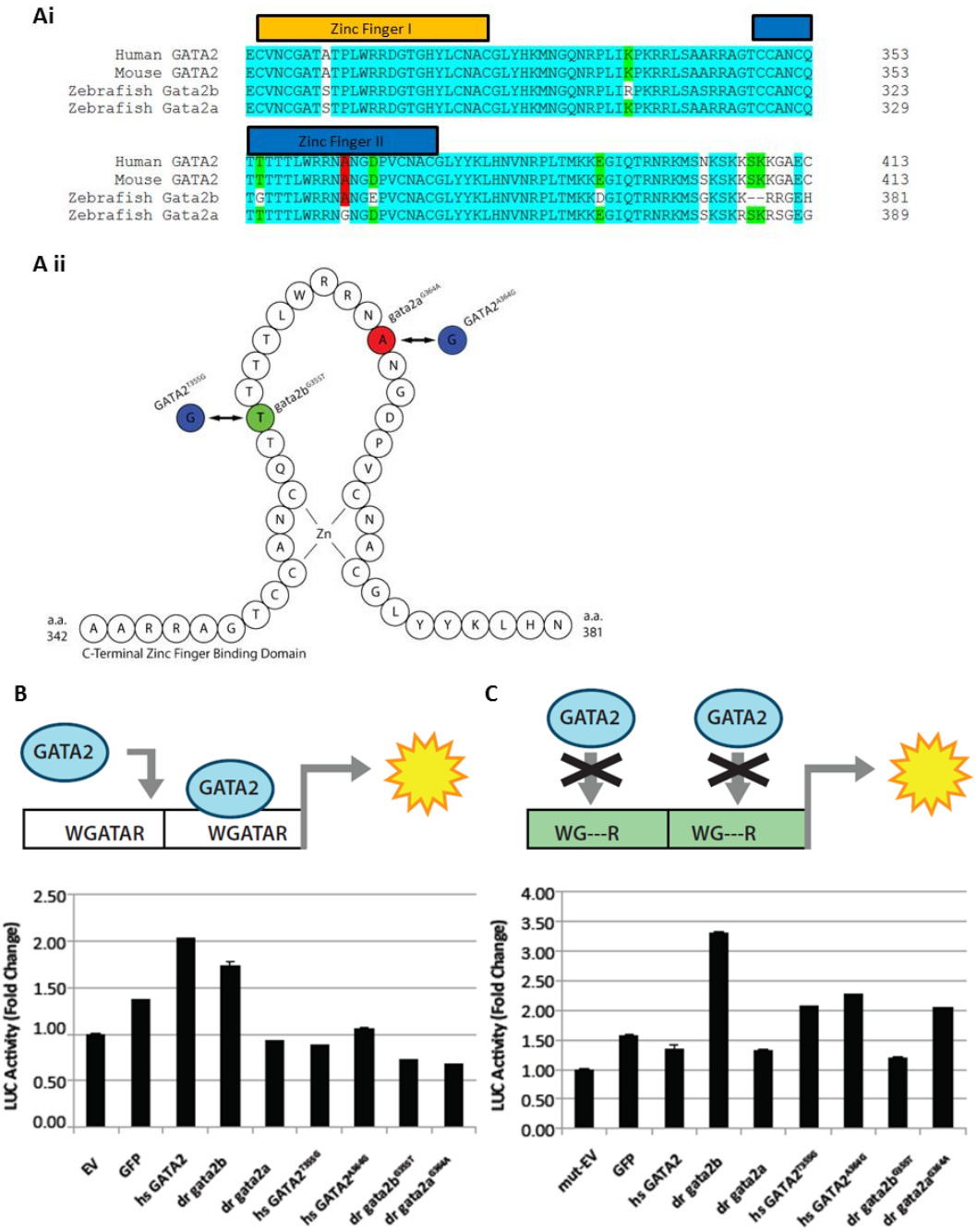


Figure 3: *gata2b* and *gata2a* display different transcriptional activation potential

(A) Amino acid sequence alignment of the Zinc Finger Binding Domains of human and murine *GATA2* against zebrafish *gata2b* and *gata2a.* Threonine to Glycine mutation of *gata2b* denoted by green arrow and Alanine to Glycine mutation of *gata2a* denoted by red arrow. (A ii) Image representing amino acid substitution within C-terminal zinc finger. (B) Luciferase-Renilla assay data using reporter with wild-type GATA2 binding consensus and (C) mutated GATA2 binding consensus.

gata2b and *gata2a* display different transcriptional activation potential

The general function of zinc finger binding domains is mainly to interact with a specific DNA motif and to mediate protein-protein interactions (Laity et al., 2001). As such, the single amino acid polymorphisms within the second zinc finger of both *gata2b* and *gata2a* (Figure 3A i) could drastically alter their binding affinity for the GATA binding consensus within the genome. To investigate this hypothesis we converted the second zinc finger domain of gata2b and gata2a to the human zinc finger domain, hereby noted as $gata2b^{G355T}$ and $gata2a^{G364A}$, respectively. Additionally, we mutated the second zinc finger domain of the human *GATA2* gene to that of the zebrafish *gata2b* or *gata2a* second zinc finger domain hereby noted as $GATA2^{T355G}$ and $GATA2^{A364G}$, respectively (Figure 3A ii).

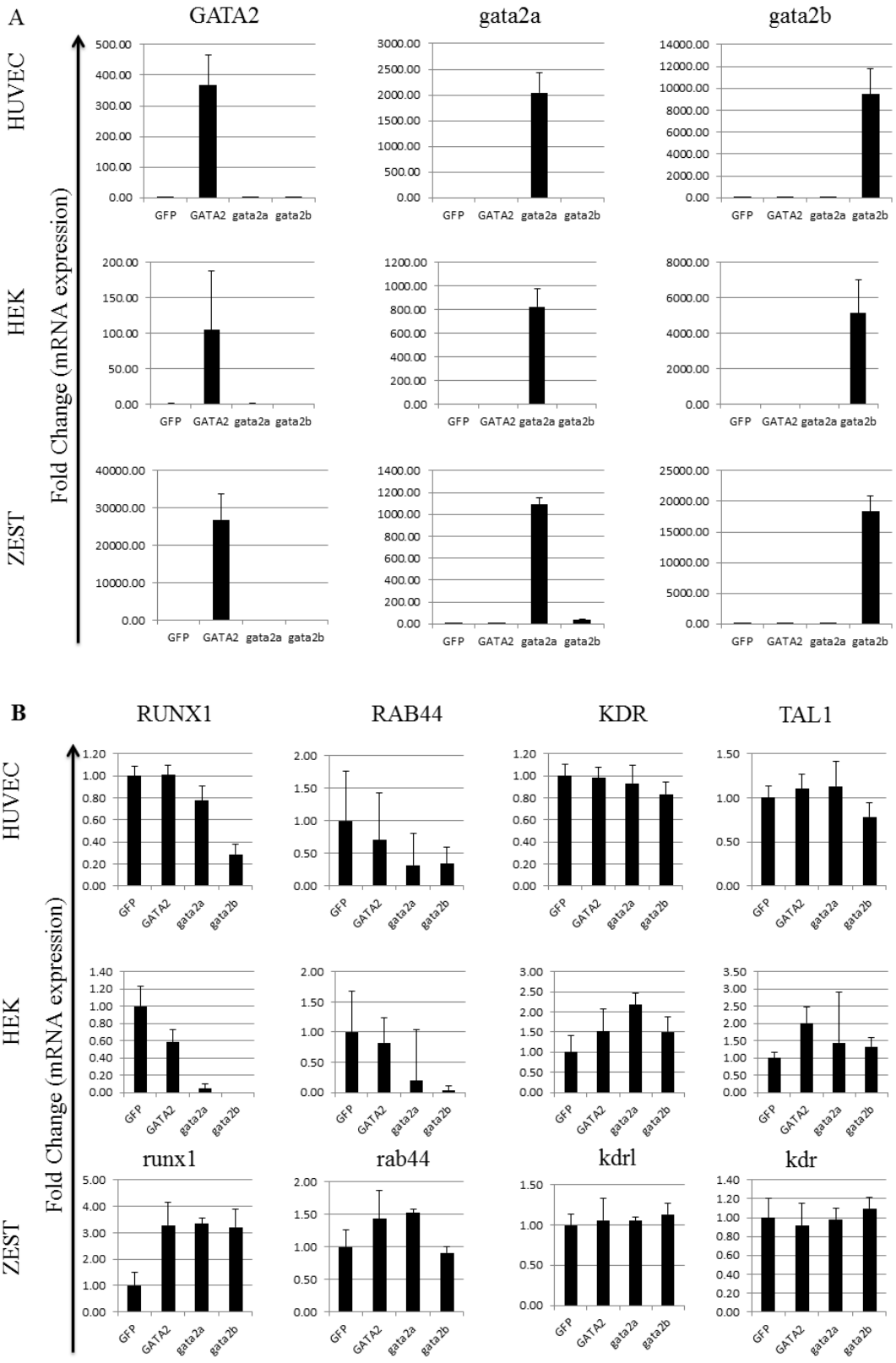
We tested the transactivation ability of these different constructs in a luciferase-renilla reporter assay To this extent, we co-transfected a reporter containing a duplicated enhancer sequence that contains the GATA2 binding consensus. Luciferase values were corrected with an internal Renilla control to account for variations in transfection efficiency. These normalized expression values were compared to an empty vector control and plotted as fold change in luciferase activity (Figure 3B). Moreover, a GFP expressing plasmid was used to account for possible non-specific binding of overexpressed proteins to the enhancer sequence. As expected, GATA2 robustly bound and activated the reporter construct with a 2-fold increase in luciferase activity as shown previously by Hahn et al., 2011. Remarkably, *gata2b* overexpression resulted in a 1.7-fold increase while *gata2a* overexpression was unable to activate the promoter. Surprisingly, none of the mutant constructs were able to activate the

luciferase reporter above background; instead, a slight repression of the reporter construct was observed in the zebrafish mutant constructs with a 0.73- and 0.68- fold change in $GATA2^{T355G}$ and $GATA2^{A364G}$ overexpression samples, respectively.

To show the specificity of the GATA2 zinc finger domain as well as the importance of the binding domain sequence, the same experiment was performed using a reporter construct with a mutation in the GATA motif of the enhancer sequence (Figure 3C). As expected, GATA2 was unable to activate the mutated reporter construct, as evidenced by the 1.3-fold increase in transactivation, against a 2 fold induction of the wild-type promoter. Unexpectedly however, *gata2b* induced the mutated construct by 3.3-fold, thereby suggesting that the mutated GATA2 consensus might be preferable for gata2b binding. The inability of *gata2a* to transactivate the reporter is not unexpected due to its inability to activate the wild-type promotor; however, it is interesting to note that all of the mutant constructs, excluding $gata2b^{G355T}$, were able to cause a 2-fold increase. Although promoter context is lost by using these isolated GATA binding elements, our results indicate that *gata2b* and *gata2a* behave differently in terms of transcriptional activation. It is highly probable that these differences are a result of the single amino acid mutations within their second zinc finger domain. Furthermore, it is remarkable that the luciferase construct with a mutated GATA2 consensus, unable to be activated by human GATA2, is strongly activated by the zebrafish *gata2b.* Moreover, the ability of the mutant GATA2 constructs to activate the promoter indicated that the difference within the second zinc finger domain does impact the function of the GATA proteins.

Figure 4: *gata2b* and *gata2a* have different gene regulation across cell lines

(A) qPCR results of HUVEC, HEK, and ZEST overexpressing GATA2, gata2a, and gata2b. (B) qPCR results of HUVEC, HEK, and ZEST overexpressing GFP, GATA2, gata2a, and gata2b. HUVEC and HEK were analyzed for RUNX1, RAB44, KDR, and TAL1. ZEST analyzed for zebrafish homologs runx1, rab44, kdrl, and kdr. All values normalized using β-actin and RSP18 primers in HUVEC and HEK and using β-actin and tBP in ZEST.


A
GATA2
gata2a
gata2b
HUVEC
HEK
ZEST
Fold Change (mRNA expression)
GFP GATA2 gata2a gata2b
B
RUNX1
RAB44
KDR
TAL1
runx1
rab44
kdrl
kdr
GFP GATA2 gata2a gata2b

Gene regulation varies across cell lines

Given the distinct functions of *gata2a* and *gata2b*, it is predicted that they regulate different gene targets resulting in unique transcriptional programs. The restricted expression patterns of *gata2a* within vasculature and *gata2b* within hemogenic endothelium might suggest inhibition of the opposing transcriptional program. Comparative analysis of downstream targets of *gata2a, gata2b,* and human GATA2 would provide insight into the molecular mechanisms governing their functional distinction. To do so, quantitative PCR was performed for common genes known to be either more specific for HSCs or endothelial cells. Hematopoietic genes analyzed include *RUNX1, RAB44*, and *TAL1* whereas analysis of endothelial markers was based on *KDR*, its zebrafish homolog *kdrl* and zebrafish ortholog *kdr*. To enforce the expression of GATA2, Gata2a and Gata2b target genes, we ectopically expressed these GATA genes within Human Umbilical cord Vein Endothelial Cells (HUVEC), Human Embryonic Kidney Cells (HEK) and Zebrafish Stromal Cells (ZEST) (Figure 4).

Analysis of hematopoietic markers in HUVEC and HEK cells indicates a down-regulation by gata2a and gata2b whereas GATA2 had no effect on its regulation. In general, the most stark effect is observed in overexpression of gata2b, resulting in at least a 2-fold decrease in expression of hematopoietic genes. However, within ZEST cells, these hematopoietic genes were highly up-regulated, particularly *runx1*, with a 3-fold increase in mRNA levels. These observations indicate an inhibitory role of *gata2b* within human cells but a highly activating one in zebrafish cells. This highlights the importance of cell context in understanding the innate molecular mechanisms of a

particular gene. The data also serves to indicate that gata2a and gata2b possess some functional redundancy and the likelihood of more unique targets of *GATA2*.

Examination of the endothelial marker KDR presents an interesting finding, which is the lack of regulation within HUVEC and ZEST cells but an up-regulation in HEK293. This suggests that within kidney cells, GATA2 expression drives it towards an endothelial character. In contrast, the absence of change in mRNA levels within HUVEC and ZEST cells indicate one of two possibilities; either KDR is not a target regulated by GATA2 transcription factors, or GATA2 is not potent enough to divert the endothelial program of these cells. The genes analyzed in this study are extremely limited and biased attributing to the minimal data of known targets. It is therefore our goal to examine, in a genome-wide and unbiased manner, downstream genes regulated by these transcription factors to delineate the dual function of GATA2 as well as the unique mechanisms of gata2a and gata2b.

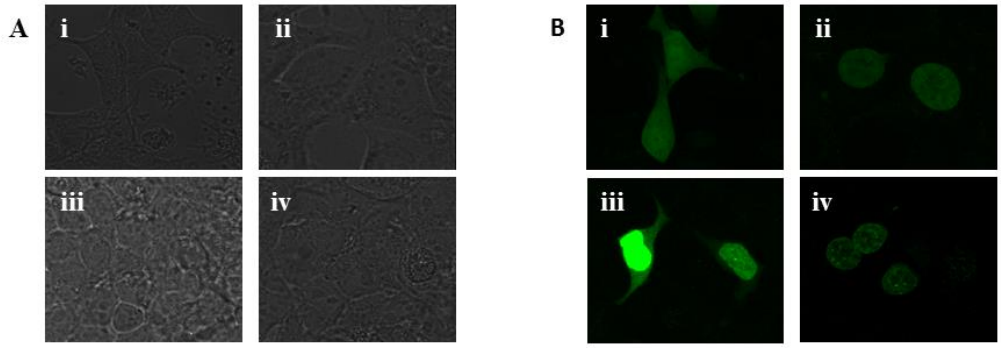


Figure 5: Transfection of GFP Fusion Plasmids

Imaging through confocal microscopy of HEK cells 24 hpt. Bright field image (A) and GFP image (B) of GFP (i), GATA2 (ii), gata2b (iii), gata2a (iv) transfected cells.

Transfection and localization of GATA factors in the Nucleus

Identifying a specific trend or function of *gata2b* and *gata2a* has proven difficult due to the limited genes analyzed by qPCR. Consequently, this provides little information into the molecular mechanisms of the zebrafish paralogs. Therefore, we attempted to gain more insight through a more global set of data using RNA sequencing. For this purpose, we choose HUVEC cells as this primary cell type is of an endothelial origin from which HSCs are predicted to arise. However, these primary cells proved difficult to transfect, resulting in a muting effect by un-transfected cells. Therefore, to enable visualization of transfection efficiency as well as enrichment through Fluorescence Activated Cell Sorting, we fused a flex-linked GFP to the N-terminus of our proteins. The addition of the flexible linker was made to prevent interaction of the large GFP protein with the zinc finger binding domains necessary for transcription factor function.

To test whether our constructed fusion proteins retained their transcriptional activity and character, we transfected HEK293 cells with either a GFP control or GATA2, Gata2a or gata2b. Upon transfection of GFP, high levels of fluorescence were seen throughout the cells. In addition, GFP expression was dispersed throughout the cell given the lack of signaling to home to specific regions (Figure 5B i). In contrast, transfection of any of the GATA proteins resulted in localization of GFP into the nucleus, indicative of transcription factor function (Figure 5B ii-iv). Additionally, this confirmed that the flex-linked GFP did not interrupt ability of the transcription factor to home to the nucleus. Interestingly, GFP expression was visualized within the cytoplasm of *gata2b* transfected cells (Figure 5B iii). A similar observation was noted within $gata2b^{G355T}$ and $GATA2^{T355G}$ transfected cells as well (data not shown), indicating the *gata2b* zinc finger binding domain affects the translocation of the transcription factor into the nucleus. Following the verification of functional constructs, we tested new transfection reagents and titrated the lipofectamine reagent to plasmid ratio to optimize transfection ratio within HUVECs. Optimization increased efficiency to ~40% visually, a significant increase from the ~5% observed previously. Next, to specifically look at only the cells that have been transfected successfully, we sorted GFP positive cells by flow cytometry.

The 10% and 8% transfection efficiency in $GATA2^{GFP}$ and $gata2a^{GFP}$ cells compared to the lower 5% in $gata2b^{GFP}$cells (Data not shown) indicate they may be more stable or have a higher turnover rate than gata2b. Due to these low levels of transfection, isolating sufficient RNA for sequencing has proven challenging. Difficulty in sorting, transfection, and RNA isolation have hindered progress in obtaining the

sample sets necessary for clean informative readouts. However, our data indicate a robust transfection method and enrichment process in which to expedite this process. Obtaining the RNA-sequencing data will provide valuable insight into both the mechanism regulating GATA2 function as well as the distinction between *gata2b* and *gata2a*.

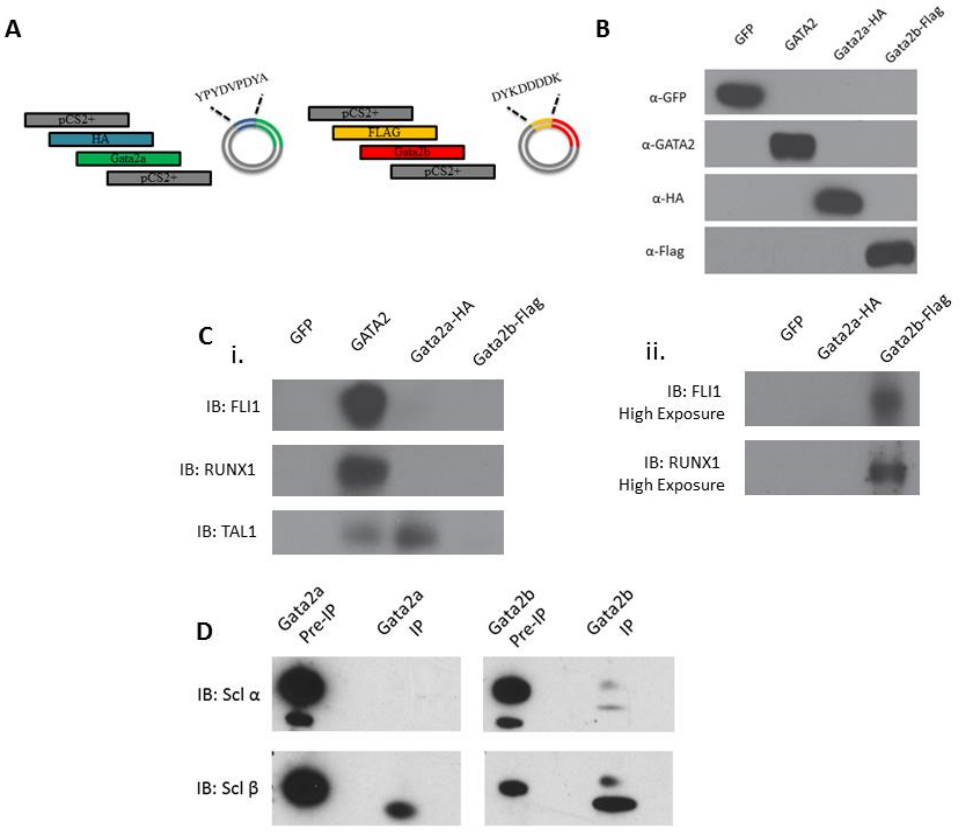


Figure 6: Transcriptional Complexes of *gata2a* and *gata2b*

(A) Fusion of HA-tag to gata2a and FLAG-tag to gata2b. (B) Immunoblot of protein harvested from transfected HEK cells for GFP, GATA2, HA, and FLAG. (C) Immunoblot for FLI1, RUNX1, and TAL1 after immunoprecipitation for GFP, GATA2, HA, and FLAG. (ii) High exposure of FLI1 and RUNX1 immunoblot with GATA2 immunoprecipitated samples removed. (D) Immunoblot for scl α and scl β after immunoprecipitation of HA or FLAG.

Transcriptional Complexes of *gata2b* and *gata2a*

The results shown previously suggest that gata2b and gata2a bind and activate promoters in distinct manners. However, it has been shown that GATA2 forms a comprehensive transcriptional complex with several transcriptions factors known to be key drivers of endothelial and hematopoietic transcriptional programs (Gering et al., 2003). With this knowledge, we investigated the possible protein-protein interaction of gata2b and gata2a with hematopoietic and vascular transcription factors known to form complexes with GATA2 (Göttgens et al., 2002, Nottingham et al., 2007, Wilson et al.,2010). As zebrafish specific antibodies for gata2b and gata2a are unavailable, we generated tagged versions of gata2b and gata2a (Figure 6A). We first tested whether the tagged versions could be recognized by their corresponding antibodies. To this extent, we transfected HEK293 cells with either GFP, GATA2, gata2a-HA, or gata2b-FLAG and detected the fusion proteins by western blot (Figure 6B).

Once verified, we transfected HEK293 cells with the indicated constructs and performed immunoprecipitation assays to detect protein-protein interactions. Immunoprecipitation of GFP, human GATA2, zebrafish gata2a and b followed by western blot for human FLI1, RUNX1, and TAL1 showed that only TAL1 was able to interact with gata2a (Figure 6C i), while FLI1 and RUNX1 with gata2b, after increased exposure time (Figure 6C ii). These results strongly suggest that gata2b and gata2a form differing transcriptional complexes that could determine their functionally separate roles during development. Moreover, the fact that gata2b interacts with the hematopoietic marker RUNX1 underlines its importance during HSC formation. Although the protein-protein interactions of gata2a and gata2b are not restricted only to

vascular or hematopoietic proteins respectively, it is important to note that such complexes can be both activating and inhibitory.

Because zebrafish antibodies for the TAL1 homologue scl were available, we explored the interaction between scl, gata2b and gata2a in more detail. However, the scl gene in zebrafish also underwent a chromosomal duplication event in the teleost lineage resulting in the paralogues sclα and sclβ (Zhen et al. 2013). Although both play functional roles in the development of HSCs, with sclβ expressed early in hemogenic endothelial cells and sclα maintaining newly formed HSCs within the AGM (Zhen et al. 2013), scl also plays a critical role in endothelial developments (Dooley et al., 2005). This suggests the scl paralogs may pair with the gata2 paralogs, prompting us to investigate the interaction of gata2b and gata2a with either sclα or sclβ by co-transfection of gata2 with scl in HEK293 cells.

The transfected cells were harvested 24 hpt and immunoprecipitated for HA and FLAG, then subsequently immunoblotted using antibodies targeting the N-terminus of SCL (sclα) and the C-terminus (sclα and sclβ). Presence of scl protein was verified by using a pre-immunoprecipitation sample. Despite initial results with TAL1 showing binding with gata2a, it seems that both gata2a and gata2b are able to bind at least one of the scl isoforms. While both gata2a and gata2b bind sclβ, a weak interaction between gata2b and sclα is also observed (Figure 6D). In sum, consistent with our prediction that gata2b has a more hematopoietic function than gata2a, we found that gata2b specifically interacts with RUNX1. Furthermore, these results also indicated that this gata2b-RUNX1 interaction has been conserved between species as HEK293 cells are of

human origin. The gata2b-sclα interaction also proves to be an interesting finding but further insight is necessary to identify its significance.

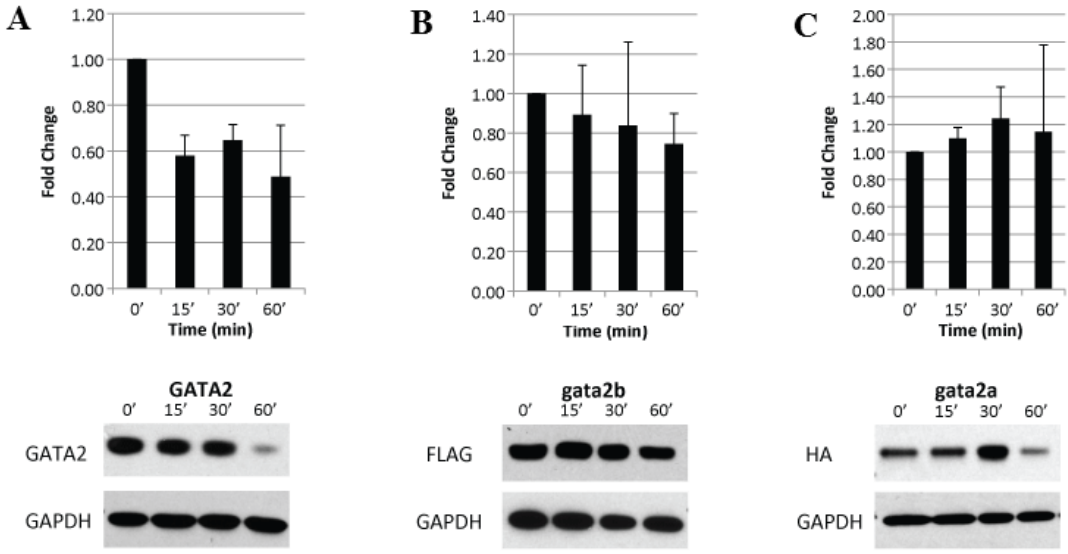


Figure 7: *gata2b* and *gata2a* have different protein stability

Graph and immunoblot of transfected HEK293 cells treated with 10uM of Cycloheximide. Band density calculated using ImageJ and represented in fold-change relative to untreated samples. Protein from *GATA2* (A), *gata2b* (B), and *gata2a* (C) transfected cells.

gata2b and gata2a, differences in protein stability

Extensive studies on the ubiquitin-proteasome protein degradation pathway show regulation of protein levels which ultimately affect transcriptional expression. Due to their potency in gene regulation, precise control over transcription factor expression is necessary. Therefore, these proteins possess high turnover rates to allow immediate activation or inhibition of their function. This is evident in the rapid degradation of GATA2 (Minegishi et al., 2005) and other hematopoietic transcription

factors (Kanei-Ishii et al., 2004). Furthermore, the effectiveness of a transcriptional activator to bind its activation domain has been shown to correlate with its rate of degradation (Molinari et al., 1999). Therefore, investigation into the degradation of gata2b and gata2a would prove informative of their potency and importance.

To this end, the different GATA2 proteins were overexpressed in HEK293 cells and subsequently treated with cycloheximide, a chemical that blocks protein synthesis. Next, cells were lyzed and protein was harvested from cells treated with cycloheximide for 0, 15, 30, and 60 minutes. Protein was analyzed for presence of GATA2 proteins by western blotting, and GAPDH was used to normalize for total protein loaded. Band density was quantified using ImageJ and averaged over two independent immunoblots.

In line with Minegishi et al. results, a rapid degradation of GATA2 was observed with a two-fold decrease in protein levels (Figure 7A). Degradation of gata2b was not quite as pronounced but degradation of the protein was still observed after 60' of treatment (Figure 7B). Remarkably, there was no degradation of gata2a over the time-course of the treatment (Figure 7C) suggesting a low turnover rate. The observed differences in degradation indicate differential regulation of these proteins, however the mechanism and purpose for this reason has yet to be determined.

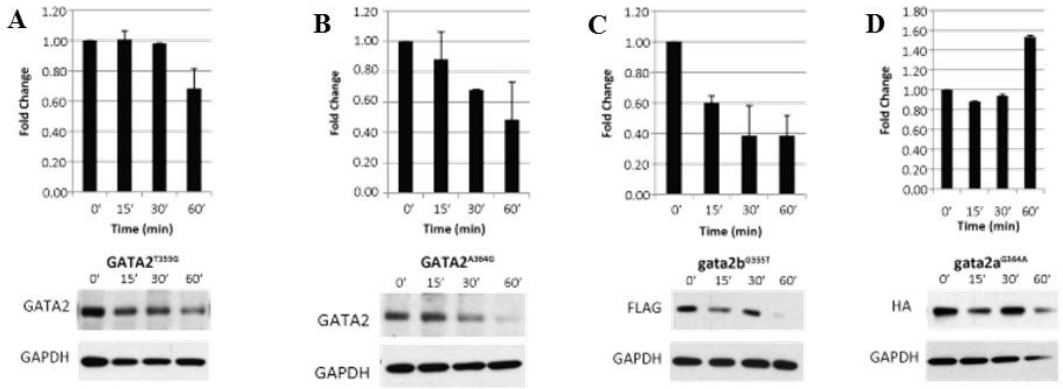


Figure 8: Mutations in the zinc finger affect protein stability

Graph and immunoblot of transfected HEK293 cells treated with 10uM of Cycloheximide. Band density calculated using ImageJ and represented in fold-change relative to untreated samples. Protein from $GATA2^{T355G}$ (A), $GATA2^{A364G}$ (B), $gata2b^{G355T}$ (C), and $gata2a^{T364G}$ (D) transfected cells.

Because ubiquitination is reliant on enzymes that recognize specific proteins to target for degradation, it is likely that this recognition is modulated by zinc finger specificity. Furthermore, as shown in Figure 3B, the amino acid substitution within the gata2b C-terminal zinc finger domain modulates its transcriptional activity. Therefore, it is plausible that mutations within the GATA2 binding domain are likely to also affect the stability of the transcription factor. To investigate this, we determined the protein stability of the mutated GATA2 factors in the method described above. Interestingly, stability of GATA2 was increased as a result of the GATA2^{T355G} (Figure 8A) but unaffected due to the GATA2^{A364G}mutation (Figure 8A). This seems to suggest the threonine is necessary to facilitate the high turnover rate of GATA2. To verify if this was indeed true, transfection and cycloheximide treatment of gata2b^{G355T} (Figure 8C) and gata2a^{G364A} (Figure 8D) was performed. Consistent with our hypothesis, substitution of the glycine for threonine resulted in high protein instability. In contrast, the protein stability of gata2a remained unchanged as a result of the alanine

substitution. Therefore, these results indicate that the degradation rate of transcription factors can be modulated by their zinc finger binding domains. Furthermore, the threonine to glycine substitution resulting in a gata2b-like zinc finger dramatically affects the stability of the protein. Together, these findings indicate the differences within the zinc finger binding domains of these GATA2 orthologs can affect the molecular mechanisms of the transcription factor in more ways than those investigated in these studies.

Discussion

Transcription factor function *in vivo* is complex making it difficult to isolate their specific mechanisms. Therefore, *in vitro* studies utilizing cell culture can provide a more simplistic model in which to gain preliminary insight into mechanisms of the *GATA2* orthologs. Due to the ease of transfection and maintenance of Human Embryonic Kidney Cells (HEK293), we performed our experiments primarily in this cell line. However, to verify our findings in a more *in vivo* cell context, we utilized Human Umbilical cord Vein Endothelial cells (HUVEC) as a cell type retaining a more endothelial characteristic and Zebrafish Stromal Tissue (ZEST) used as structural support for the growth of hematopoietic cells. Previous ChIP-sequencing and RNA-sequencing data have shown that the methylation of DNA varies between cell-types indicating the availability of binding sequences for interactions would differ amongst different cell-types (Zhang et al., 2013). Therefore, it is important to consider the cell context in which our experiments were performed.

Although previous studies showed that *gata2a* and *gata2b* differ in their expression and function, with *gata2a* functioning in a vascular role (Brown et al., 2000; Detrich et al., 1995; Zhu et al., 2011) and *gata2b* in a hematopoietic specific role (Butko et al., 2015), to date, there is no evidence that two paralogs drive distinct and unique genetic pathways. Therefore, our studies attempted to gain insight into the gene regulation of the *GATA2* orthologs to identify if *gata2a* and gata2b are indeed unique in their functionality. To do so, we observed regulation of select hematopoietic and endothelial favored genes based upon previous studies implicating their regulation by *GATA2*. Although a clear distinction in regulation was not observed in the manner we

predicted, with *gata2a* downregulating hematopoietic targets and *gata2b* upregulating them, there does seem to be differences depending on the cell line. Most notably is the regulation of the hematopoietic marker *runx1* which has been shown to be essential for HSC development (Chen et al., 2009). Within HUVEC and HEK cells, *RUNX1* is significantly downregulated by both *gata2a* and *gata2b* overexpression. In contrast, *runx1* expression within ZEST cells was highly upregulated, which might indicate that cellular context is important. Despite the endothelial origin from which HSCs arise (Bertrand et al., 2010) HUVEC and HEK cells are not be the specific endothelial cell type from which HSCs arise. ZEST cells however, cultured for the purpose of supporting growth of hematopoietic cells, likely has some of the factors and signals necessary for hematopoiesis (Stachura et al., 2009). It is equally likely that there is not enough conservation between the human and zebrafish orthologs for the zebrafish GATA2 factors to interact with endogenous human proteins.

In general, HUVEC and ZEST cells, both primary cells, have not been immortalized like HEK293 and prove difficult to transfection, typically between1% - 5%. This low transfection efficiency might be the case that we do not observe a strong regulation of known GATA2 target genes. In order to enrich for successfully transfected cells, we generated GFP fusion constructs, linking GFP to our GATA2 constructs with a flexible-linker to avoid steric hindrance. These constructs enables us to visualize the transfection efficiency as well as enrich for transfected cells by FACS. Therefore, we can control for the difficulty in transfecting primary cells as well as remove any background noise generated by non-transfected cells. As such, repetition of our qPCR data is necessary but more importantly, RNA-sequencing to identify gene

targets regulated by *gata2a* and *gata2b* will provide a broader list as the targets we chose to investigate are based upon targets known to be regulated by *GATA2*.

Sequence alignment between *GATA2* and the zebrafish paralogs *gata2b* and *gata2a* shows that conservation remains primarily within the dual zinc finger binding domains (Figure 1). It is known that zinc finger structures are critical for regulating DNA binding and protein-protein interactions which modulate transcription factor function (Laity et al., 2001). We identified one major amino acid substitution within the zinc finger domain of both *gata2a* and *gata2b* (Figure 3A i) and could most likely define their unique roles in either endothelial or hematopoietic specification. Due to the conservation of the alanine to serine substitution of amino acid 302 within the N-terminal zinc finger, it was predicted that differences in function would likely not result from this amino acid substitution. Furthermore the second mutation of the *gata2b* C-terminal zinc finger binding domain, an aspartate to glutamate substitution, results in a conservation of negatively charged amino acid. Therefore, although structure could be altered, it is expected to be minimal and unlikely to affect zinc finger function. However, compared with the human GATA2 zinc finger domain, both *gata2a* and *gata2b* obtained a glycine, a small amino acid especially used to create sharp turns within the protein structure.

Our studies indicate the T355G substitution resulting in a gata2b-like zinc finger significantly affects transcription factor function and could likely be a key component in regulating the hematopoietic specificity of *gata2b*. Most indicative is the increased ability of *gata2b* to transactivate a mutated GATA binding sequence (Figure 3C). The increase of luciferase activation was markedly higher than activation levels seen

utilizing a reporter containing the conserved WGATAR sequence (Figure 3B). Furthermore, the mutated GATA2 constructs were able to activate the mutated reporter construct, though to a lesser extent. In line with our expectations, the $gata2b^{G355T}$ construct was no longer able to transactivate the mutant promoter. These results have indicated that in terms of transactivation potential, the *gata2b* mutation affects the zinc finger enough to alter its preferred binding consensus. Furthermore, it is predicted that *GATA2* and *gata2b* bind different sequences within the genome, thereby activating different genes, possibly a more specific set of genes regulated by *GATA2*. This could indicate the *gata2b* zinc finger has evolved to become more specific for hematopoietic targets while *GATA2* relies more on the transcriptional complexes formed to regulate its role. Although, *gata2a* is unable to activate either reporter constructs, we are unable to determine if this is due to an inability to bind the consensus or a lack of activation potential. Performing a ChIP assay would prove beneficial to this end as we can then determine if *gata2a* has bound to the reporter construct. Overall ChIP-sequencing would confirm if these GATA2 orthologs bind to different DNA consensus as well as indicate what other binding motifs are commonly found in close proximity.

It is known that protein levels are regulated through various degradation pathways, which control the duration of protein expression. This modulation of transcription factor levels is important as overexpression of certain genes is detrimental to cell survival. Furthermore, given the importance of zinc finger structure to activation potential, we hypothesized that the differences we identified in the C-terminal zinc finger binding domains would likely alter protein stability. Cycloheximide treatment, a chemical blocking *de novo* protein synthesis, indicates that the T355 does indeed have

significance on the molecular mechanisms of *GATA2*. We demonstrated that mutation of the threonine to glycine results in increased stability of the protein; whereas a glycine to threonine substitution of the same corresponding region results in decreased stability. Furthermore, mutation of the same threonine has been observed within AML patients (Hahn et al., 2011) indicating it is critical in maintaining proper *GATA2* function. The reason for this phenotypic response remains to be determined. However, we can speculate that this particular mutation has resulted in an inability of *GATA2* to properly bind and activate necessary genes or rather an increased level of target genes as a result of increased protein stability. Phosphorylation of transcription factors have been shown to affect their stability and activation potential (Biggs et al., 2005) indicating the loss of a phosphorylation site due to the T355G substitution is the cause of the changes observed. It would be interesting to see if this substitution is capable of specifying GATA2 function toward a more hematopoietic function. This necessitates further comparative analysis on the similarities of *GATA2* and *gata2b* and subsequent confirmation to see if $GATA^{T355G}$ can replicate similar results.

Interestingly, we observed that *gata2b* fused to GFP was not fully able to translocate into the nucleus (Figure 5B iii). This observation is likely a result of the threonine to glycine substitution of its C-terminal zinc finger binding domain. However, it is possible that *gata2b* is unable to interact with the necessary proteins within HEK293 cells that translocate it to the nucleus. Our study has shown that *gata2b* and *gata2a* form different complexes with known interaction partners of GATA2. As such, it is not unlikely that *gata2b* requires a different protein to facilitate its localization than GATA2 or gata2a. Preliminary observation of cells transfected with

our GFP-fusion mutant constructs have shown that the translocation of GATA2 is affected when its zinc finger binding domain is mutated. Therefore, as mentioned previously, the gata2b zinc finger likely facilitates a more hematopoietic function thereby suggesting the $GATA^{T355G}$would be a prime candidate for driving HSC formation *in vitro.*

In conclusion, we have shown that *gata2a* and *gata2b* have different molecular mechanisms and are regulated differently. These differences are a result of differences within the zinc finger binding domains. However, as these studies have been performed *in vitro,* verification is necessary *in vivo.* Using the GAL4/UAS system within the zebrafish, we can drive overexpression of our *GATA2* constructs within specific cell lines. The inability of *gata2a* to rescue *gata2b* knockout within the mutant line we are characterizing would show that indeed *gata2a* and *gata2b* drive different cellular fates. Consequently, if *gata2a* can rescue the *gata2b* knockout phenotype, it would suggest that *gata2a* is capable of compensating for *gata2b*. More interestingly would be to see if $GATA^{T355G}$would be able to rescue the *gata2b* mutants while *GATA2* cannot, indicating that there is enough conservation between *GATA2* and *gata2b* that this single amino acid substitution is sufficient to drive *GATA2* toward a hematopoietic role. Isolation of vascular endothelial cells and hematopoietic stem cells through FACS and subsequent qPCR analysis for genes modulated would verify the data obtained *in vitro*.

Our current data has been limited to interaction partners and target genes of *GATA2*shown in previous studies thereby limiting the conclusions that can be drawn. Therefore, a genome-wide approach to identifying these factors would provide better insight into their individual molecular mechanisms. A more comprehensive list can

then be obtained through comparative analysis between the three *GATA2* orthologs isolating three subsets of mechanisms: endothelial specific mechanisms (shared by *GATA2* and *gata2a*), hematopoietic specific (shared by *GATA2* and *gata2b*) and overlapping mechanisms (shared by *gata2b* and *gata2a*). Verification of these mechanisms would provide better knowledge on the individual roles of *GATA2* ultimately coalescing in its use as a specific driver of long-term transplantable HSC formation from induced pluripotent stem cells.

REFERENCES

Aggarwal, R., Lu, J., Pompili, V. & Das, H. Hematopoietic Stem Cells: Transcriptional Regulation, Ex Vivo Expansion and Clinical Application. *CMM* **12,** 3449 (2012).

Auer, T. O., Duroure, K., De Cian, A., Concordet, J.-P. P. & Del Bene, F. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res.* **24,** 142–53 (2014)

Auer, T. O. & Del Bene, F. CRISPR/Cas9 and TALEN-mediated knock-in approaches in zebrafish. *Methods*(2014). doi:10.1016/j.ymeth.2014.03.027

Biggs, J. R., Zhang, Y., Peterson, L. F., Garcia, M., Zhang D.E., Kraft A. S. Phosphorylation of AML1/RUNX1 regulates its degradation and nuclear matrix association. *Mol Cancer Res*,3(7):391 – 401(2005)

Bertrand, J., Kim, A., Teng, S. & Traver, D. CD41+ cmyb+ precursors colonize the zebrafish pronephros by a novel migration route to initiate adult hematopoiesis.*Development (Cambridge, England)* **135,** 1853–62 (2008)

Bertrand, J. Y., Chi, N. C., Santoso, B., Teng, S., Stainier, D. Y. R. and Traver, D. . Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature* **464**, 108-111. (2010)

Brown, L. *et al.* Insights into early vasculogenesis revealed by expression of the ETS-domain transcription factor Fli-1 in wild-type and mutant zebrafish embryos. *Mechanisms of Development* **90,** (1999)

Butko, E., Distel, M., Pouget, C., Weijts, B., Kobayashi, I., Ng, K., Mosimann, C., Poulain, F., McPherson, A., Ni, C.-W., Stachura, D., Cid, N., Espín-Palazón, R., Lawson, N., Dorsky, R., Clements, W., Traver, D. *Development* **142,** 1050–1061 (2015).

Bussmann, J. & Schulte-Merker, S. Rapid BAC selection for tol2-mediated transgenesis in zebrafish. *Development* **138,**4327–32 (2011).

Chen, M. J., Yokomizo, T., Zeigler, B. M., Dzierzak, E. and Speck, N. A. Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. Nature 457, 887-891. (2009).

Clements, W. K. & Traver, D. Signalling pathways that control vertebrate haematopoietic stem cell specification. *Nat. Rev. Immunol.* **13,** 336–48 (2013).

Dooley, K., Davidson, A. & Zon, L. Zebrafish scl functions independently in hematopoietic and endothelial development.*Dev. Biol.* **277,** (2004).

Fujiwara, T., O'Geen, H., Keles, S, Blahnik, K., Linnemann, A., Kang, Y.-A., Choi, K., Farnham, P., Bresnick, E. Discovering Hematopoietic Mechanisms through Genome-wide Analysis of GATA Factor Chromatin Occupancy. *Mol. Cell* **36,**(2009).

Fujiwara, Y., Chang, A. N., Williams, A. M., Orkin S. H. Functional overlap of GATA-1 and GATA-2 in primitive hematopoietic development. *Blood,*103:583-585(2004)

Gao, X., Johnson, K. D., Chang, Y.-I., Boyer, M. E., Dewey, C. N., Zhang, J. and Bresnick, E. H.. Gata2 cis-element is required for hematopoietic stem cell generation in the mammalian embryo. J. Exp. Med. 210, 2833-2842. (2013)

Gao, J., Gentzler, R., Timms, A., Horwitz, M., Frankfurt, O., Altman, J., Peterson, L. . Heritable GATA2 mutations associated with familial AML-MDS: a case report and review of literature.*Journal of hematology & oncology* **7,**36 (2014).

Gillis, W., John, J., Bowerman, B. & Schneider, S. Whole genome duplications and expansion of the vertebrate GATA transcription factor gene family. *BMC Evol. Biol.* **9,** 207 (2009).

Göttgens, B., Nastos, A., Kinston, S., Piltz, S., Delabesse, E., Stanley, M., Sanchez, M., Ciau-Uitz, A., Patient, R., Green, A. . Establishing the transcriptional programme for blood: the SCL stem cell enhancer is regulated by a multiprotein complex containing Ets and GATA factors. *EMBO J.* **21,** 3039–3050 (2002).

Hahn, C. N., Chong, C.-E. E., Carmichael, C. L., Wilkins, E. J., Brautigan, P. J., Li, X.-C. C., Babic, M., Lin, M., Carmagnac, A., Lee, Y. K., Kok, C. H., Gagliardi, L., Friend, K. L., Ekert, P. G, Butcher, C. M., Brown, A. L., Lewis, I. D., To, L. B., Timms, A. E., Storek, J., Moore, S, Altree, M., Escher, R., Bardy, P. G., Suthers, G. K., D'Andrea, R. J., Horwitz, M. S., Scott, H. S. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat. Genet.* **43,** 1012–7 (2011).

Heasman, J. Morpholino Oligos: Making Sense of Antisense?*Developmental Biology* **243,**(2001).

Hruscha, A., Krawitz, P., Rechenberg, A., Heinrich, V., Hecht, J., Haass, C., Schmid, B. Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development* **140,**4982–7 (2013).

Jao, L.-E., Wente, S. & Chen, W. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc. Natl. Acad. Sci. U.S.A.* **110,**13904–13909 (2013).

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., Zhang, J., Holland, S.M., Bresnick, E.H. Cis-element mutated in GATA2- dependent immunodeficiency governs hematopoiesis and vascular integrity. J. Clin. Invest. 122, 3692-3704. (2012)

Laity, J., Lee, B. & Wright, P. Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struct. Biol.* **11,** (2001).

Lim, K.-C., Hosoya, T., Brandt, W., Ku, C.-J., Hosoya-Ohmura, S., Camper, S., Yamamoto, M., Engel, J. . Conditional Gata2 inactivation results in HSC loss and lymphatic mispatterning. *J. Clin. Invest.* **122,**37053717 (2012).

Lin HF, Traver D, Zhu H, Dooley K, Paw BH, Zon LI, Handin RI. Analysis of thrombocyte development in CD41-GFP transgenic zebrafish. Blood. (2005)

Ling, K.-W., Ottersbach, K., Hamburg, J., Oziemlak, A., Tsai, F.-Y., Orkin, S., Ploemacher, R., Hendriks, R., Dzierzak, E. GATA-2 Plays Two Functionally Distinct Roles during the Ontogeny of Hematopoietic Stem Cells.*Journal of Experimental Medicine***200,** 871–882 (2004).

Linnemann, A., O'Geen, H., Keles, S., Farnham, P. & Bresnick, E. Genetic framework for GATA factor function in vascular biology. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 13641–13646 (2011).

Minegishi, N., Suzuki, N., Kawatani, Y., Shimizu, R. & Yamamoto, M. Rapid turnover of GATA-2 via ubiquitin-proteasome protein degradation pathway. *Genes Cells* **10,** 693–704 (2005).

Molinari, E., Gilman, M. & Natesan, S. Proteasome-mediated degradation of transcriptional activators correlates with activation domain potency in vivo. *EMBO J* **18,** 6439–6447 (1999).

Network, C. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *The New England journal of medicine* **368,** 2059–74 (2013).

Nottingham, W., Jarratt, A., Burgess, M, Speck, C., Cheng, J.-F., Prabhakar, S., Rubin, E., Li, P.-S., Sloane-Stanley, J.; Kong-a-San, J.; de Bruijn, M. Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer. *Blood* **110,** 4188 (2007).

Orkin, S. Transcription Factors and Hematopoietic Development. *Journal of Biological Chemistry* **270,** 4955–4958 (1995).

Pater, E., Kaimakis, P., Vink, C., Yokomizo, T., Yamada-Inagawa, T., Linden, R., Kartalaei, P., Camper, S., Speck, N., Dzierzak, E. Gata2 is required for HSC generation and survival. *J. Exp. Med.* **210,**2843–2850 (2013).

Shi, X., Richard, J., Zirbes, K., Gong, W, Lin, G., Kyba, M., Thomson, J, Koyano-Nakagawa, N., Garry, D. Cooperative interaction of Etv2 and Gata2 regulates the development of endothelial and hematopoietic lineages. *Developmental Biology***389,** 208218 (2014).

Stachura, DL., Reyes, JR., Bartunek, P., Paw, B.H., Zon, L.I., Traver, D., Zebrafish kidney stromal cell lines support multilineage hematopoiesis. *Blood* **114,** 279-289 (2009)

Tsai, F.-Y., Orkin, S.H. Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. *Blood.* **89**, 3636-3643 (1997)

Tsai, F.-Y., Keller, G., Kuo, F., Weiss, M., Chen, J., Rosenblatt, M., Alt, F., Orkin, S. An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature* **371,** 221–226 (1994).

Wilson, N., Calero-Nieto, F., Ferreira, R. & Göttgens, B. Transcriptional regulation of haematopoietic transcription factors. *Stem Cell Res Ther* **2,** 6 (2011).

Wilson, N., Foster, S., Wang, X., Knezevic, K, Schütte, J., Kaimakis, P., Chilarska, P., Kinston, S., Ouwehand, W., Dzierzak, E., Pimanda, J., Bruijn, M, Göttgens, B. Combinatorial Transcriptional Control In Blood Stem/Progenitor Cells: Genome-wide Analysis of Ten Major Transcriptional Regulators. *Cell Stem Cell* **7,**(2010).

Zhang, B., Zhou, Y., Lin, N., Lowdon, R., Hong, C, Nagarajan, R., Cheng, J., Li, D., Stevens, M., Lee, H., Xing, X., Zhou, J., Sundaram, V., Elliott, G., Gu, J., Shi, T., Gascard, P., Sigaroudinia, M,; Tlsty, T., Kadlecek, T., Weiss, A., O'Geen, H., Farnham, P., Maire, C., Ligon, K., Madden, P., Tam, A., Moore, R., Hirst, M,; Marra, M., Zhang, B., Costello, J., Wang, T. Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. *Genome Research* (2013). doi:10.1101/gr.156539.113

Zhen, F., Lan, Y., Yan, B., Zhang, W. & Wen, Z. Hemogenic endothelium specification and hematopoietic stem cell maintenance employ distinct Scl isoforms.*Development* **140,** 39773985 (2013).

Zhu, C., Smith, T., McNulty, J., Rayla, A., Lakshmanan, A., Siekmann, A., Buffardi, M., Meng, X., Shin, J., Padmanabhan, A., Cifuentes, D., Giraldez, A., Look, A., Epstein, J., Lawson, N., Wolfe, S. Evaluation and application of modularly assembled zinc-finger nucleases in zebrafish. *Development***138,** 4555–4564 (2011).