Universidad de Navarra

Facultad de Ciencias

CONTRIBUTION OF THE GATA2 TRANSCRIPTION FACTOR TO THE DEVELOPMENT AND PROGRESSION OF MYELOID DISORDERS

Xabier Cortés Lavaud

Tesis doctoral

2014

Universidad de Navarra

Facultad de Ciencias

CONTRIBUTION OF THE GATA2 TRANSCRIPTION FACTOR TO THE DEVELOPMENT AND PROGRESSION OF MYELOID DISORDERS

Memoria presentada por D. Xabier Cortés Lavaud para aspirar al grado de Doctor por la Universidad de Navarra.

(firma del doctorando)

El presente trabajo ha sido realizado bajo mi dirección en el Departamento de Bioquímica y Genética y autorizo su presentación ante el Tribunal que lo ha de juzgar.

Pamplona, 8 de abril de 2014

(firma del Director de la Tesis Doctoral)

Dra. María Dolores Odero

A mis padres

A mis hermanos

A Sofía

ACKNOWLEDGEMENTS

Son muchísimas las personas con las que he interaccionado estos años en Pamplona, comenzando por las Licenciaturas de Química y Bioquímica, el Master de Biología Celular y Molecular y finalmente el doctorado en el Centro de Investigación Médica Aplicada (CIMA). Debo agradecer en primer lugar a los profesores de la Universidad de Navarra todo su esfuerzo docente y humano para formar personas competentes. En segundo lugar, quiero agradecer al decano de la Facultad de Ciencias, D. Ignacio López Goñi y al director del Departamento de Bioquímica y Genética, D. Carlos de Miguel, la oportunidad que me brindaron para completar mi formación con la realización de la tesis doctoral. Por último, debo agradecer a D. Francisco Errasti y D. Jesús Hernández la posibilidad de poder comenzar mi formación como investigador en el CIMA, realizando el master y posteriormente la tesis doctoral.

Cómo no agradecer a Lola Odero, mi directora de tesis, toda la labor realizada para formarme, día a día, como investigador y como docente. Tu empeño y apoyo para que aprendiera lo mejor posible a pensar, estudiar y presentar los datos científicos, tus críticas constructivas a mis ideas y opiniones y tu paciencia a la hora de observar mis lentos progresos me han ido enseñando, poco a poco, a trabajar en esto de la investigación. Además, me has permitido participar muy activamente en la parte práctica de la asignatura de Genética Clínica del Grado en Medicina, con lo que he disfrutado mucho y estoy seguro de que me ha aportado un grado de formación adicional tremendamente importante.

Hablando de docencia, debo agradecer al Departamento de Química y Edafología y al Departamento de Bioquímica y Genética la oportunidad de poder participar en tareas docentes de múltiples asignaturas. He podido aprender mucho de Maribel Zudaire, Paula Aranaz y Karmele Valencia. Su dedicación y entusiasmo a la hora de impartir las clases me ha servido de ejemplo y motivación, hasta tal punto que tomo docencia universitaria como algo a considerar muy seriamente en el futuro.

Los largos períodos de trabajo en el laboratorio han sido compartidos con muchas personas que no se reducen al propio grupo de investigación, y ni siquiera al propio CIMA. Nombrarlas a todas haría la lista larguísima, así que me limitaré principalmente a dar el agradecimiento a todos los que hemos pertenecido al mismo grupo: Ana Conchillo, Ana Domínguez, Carmen, Choni, Cris, Daniella, Gemma, Ion, Iria, Isa, Laura, Leire, Miren, Nerea, Raffy, Reme y Santi. Quiero también hacer mención en estos agradecimientos a Eloy, quien fue mi supervisor directo durante el master, hace ya casi 5 años. Vuestras muchas

enseñanzas, los consejos técnicos y la amable compañía han hecho de mi estancia en el CIMA algo memorable.

Durante el tiempo vivido en Pamplona, también he tratado de cultivar, a veces con más constancia que otras, las amistades forjadas durante la Licenciatura de Química. Los químicos siempre habéis estado ahí y he disfrutado mucho de nuestros encuentros. Las amistades no se reducen solo a este círculo, y puedo decir que he conocido a unas personas extraordinarias en Pamplona, con las que me gustaría seguir en contacto.

Además de la labor investigadora y docente, he participado en la vida universitaria coordinando la Escuela de Kendo, organizando las clases y diferentes actividades. Desde este año ha pasado a ser el Club de Kendo de la Universidad de Navarra, lo que supone un honor para mí porque reconoce la labor realizada. Si un día vuelvo a Pamplona de visita, me producirá una gran satisfacción saber que hay un grupo de Kendo con el que entrenar.

Estos más de 10 años en Pamplona me han alejado, sólo un poco, de mi ciudad natal, donde está mi cuadrilla, y con la que he compartido innumerables momentos que han dado lugar a las más divertidas anécdotas. El haber escrito la tesis en San Sebastián me ha permitido retomar el contacto más que nunca, y no puedo estar más agradecido de vuestros ánimos, sabiendo que además varios de nosotros hemos estado terminando nuestras respectivas tesis casi simultáneamente y tocaba repartir esos ánimos.

El estar en San Sebastián estos últimos meses también me ha permitido hacer algo parecido a un viaje atrás en el tiempo. Aunque estos meses haya sido más bien un habitante de la biblioteca más que de casa, tengo que agradecer enormemente a mis padres el haberme apoyado incondicionalmente, con una generosidad y paciencia inmensas, desde mi más tierna infancia hasta la actualidad. Si he de destacar algo que admire de ellos y que quisiera hacer mío, sería la capacidad de ser tan constante y eficiente de mi padre, y el aprecio que tiene mi madre por el detalle. A mis hermanos les agradezco el gran interés que les ha suscitado mi tesis, preguntándome con enorme curiosidad y atendiendo a mis torpes explicaciones sobre genes, proteínas y cáncer.

Y por último, debo dar un agradecimiento especial a esa persona que uno no sabe que le falta hasta que se la conoce. Sofía, si tú me has enseñado algo a la hora de realizar la tesis, ha sido tu entusiasmo y tu pasión por el trabajo científico, del que me he contagiado. No hay nada que me haga más ilusión que

tener un proyecto de vida contigo, sea donde sea. El estar contigo me ayuda a ser mejor cada día, me motiva para dar lo mejor de mí mismo. Creo que tu potencial en el campo científico es enorme, mucho mayor que el mío, y quiero apoyarte en todo aquello que te propongas.

Finalmente, estoy seguro de que con estos reconocimientos no hago justicia a todas las personas por los que me siento profundamente agradecido, y quisiera disculparme por si no los he mencionado al menos de forma indirecta.

Ezina ekinez egina

Antiguo proverbio vasco

ABBREVIATIONS

AML	Acute myeloid leukemia
ATRA	All-trans-retinoic acid
bp	base pairs
BM	Bone marrow
CD	Cluster of differentiation
cDNA	Complementary deoxyribonucleic acid
CFU-G	Granulocyte colony forming units
CFU-GEMM	Granulocyte/Erythroid/Macrophage/Megakaryocyte colony forming units
CFU-GM	Granulocyte/Macrophage colony forming units
CFU-M	Macrophage colony forming units
ChIP	Chromatin immunoprecipitation
CLP	Common lymphoid progenitor
CML-BC	Chronic myeloid leukemia in blast crisis
CMML	Chronic myelomonocytic leukemia
CMP	Common myeloid progenitor
CMV	Cytomegalovirus
DCML	Dendritic cell, monocyte, and lymphoid deficiency
DFS	Disease-free survival
DNA	Deoxyribonucleic acid
EBV	Epstein-Barr virus
EFS	Event-free survival
ES	Embryonic stem
FAB	French-American-British
GMP	Granulocyte-macrophage progenitor
HPC	Hematopoietic progenitor cell
HPV	Human papillomavirus
HSC	Hematopoietic stem cell
HSCT	Hematopoietic stem cell transplantation
ITD	Internal tandem duplication
kb	kilobases
LGL	Large granular lymphocytes
LT	Long-term
MEP	Megakaryocyte-erythroid progenitor
MDS	Myelodysplastic syndrome

MonoMAC	Monocytopenia and mycobacterium avium complex infection
MPP	Multipotent progenitor
MCSF	Macrophage colony-stimulation factor
NLS	Nuclear localization signal
NRD	Negative regulatory domain
NTM	Non-tuberculous mycobacterium
NK	Natural killer
OS	Overall survival
PAP	Pulmonary alveolar proteinosis
PCR	Polymerase chain reaction
qRT-PCR	Quantitative real time polimerase chain reaction
RNA	Ribonucleic acid
siRNA	Short interfering ribonucleic acid
ST	Short-term
TAD	Transactivation domain
T-ALL	T-cell acute lymphoblastic leukemia
TPA	12-O-tetradecanoylphorbol-13-acetate
TSS	Transcription start site
WT	Wild type
ZnF	Zinc-finger

TABLE OF CONTENTS

INTRODUCTION

Hematopoiesis is a hierarchically regulated process committed for immunologic surveillance and tissue oxygenation under the control of cytokines, signaling pathways, bone marrow microenvironment, and transcription factors (Wickrema et al., 2007). Together, they coordinate the progression from the hematopoietic stem cell (HSC), characterized by pluripotency and self-renewal, through hematopoietic progenitor cell (HPC) stages, retaining proliferative ability but increasing restricted self-renewal and differentiation potential, to lineage-specific cells that undergo terminal differentiation (Figure 1) (Orkin et al., 2008).

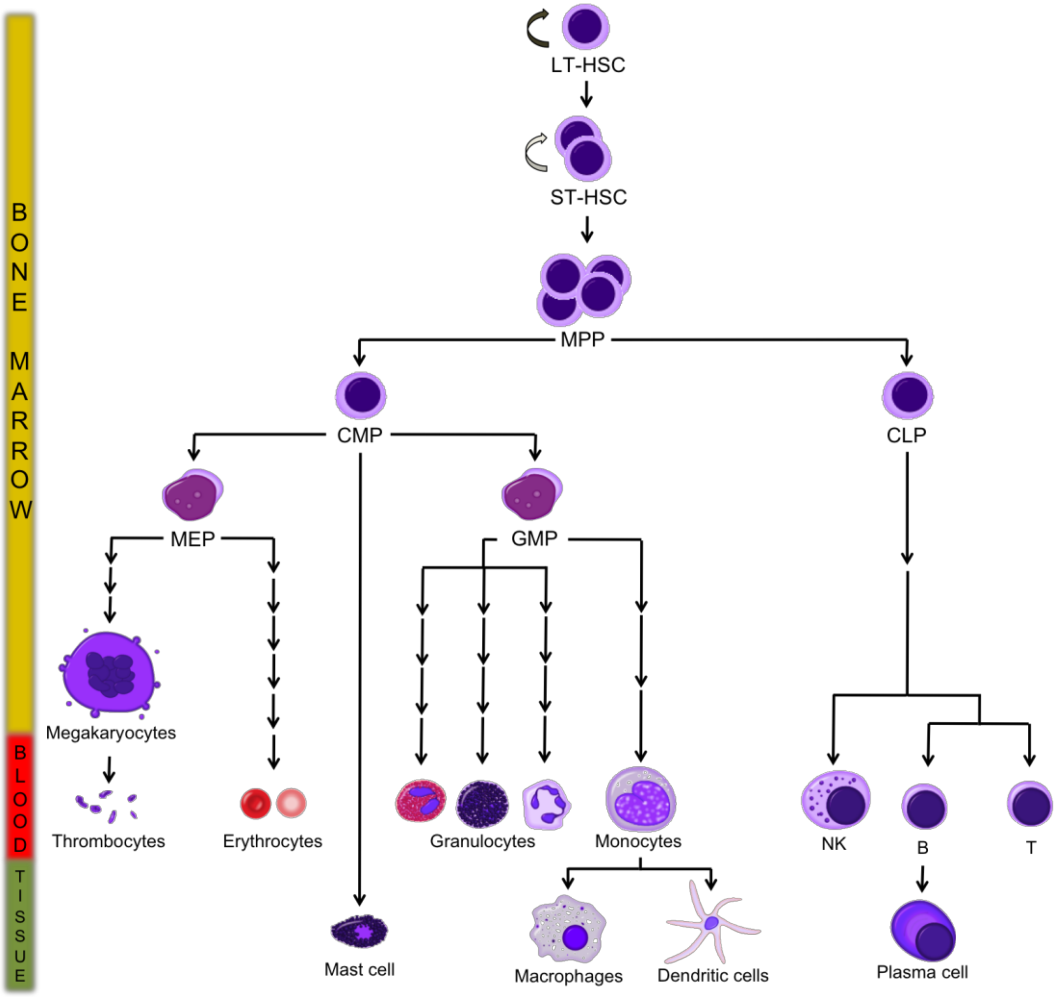

Figure 1: Schematic representation of hematopoiesis. Hematopoietic stem cells (HSC) can be classified as long-term (LT) or short-term (ST) HSC depending on their division rate and self-renewing potential. They give rise to multipotent progenitor cells (MPP) and then common myeloid or lymphoid progenitors (CMP and CLP, respectively), which show restricted self-renewal and differentiation potential. The CMP can give rise to the megakaryocyte-erythroid or granulocyte-macrophage progenitor (MEP and GMP, respectively). Differentiated cells leave the bone marrow and exert their function in blood or other tissues.

An interconnected network of transcription factors modulates downstream gene expression and determines cellular fate; therefore, their function must be tightly regulated during hematopoietic development to maintain tissue homeostasis. The GATA family of nuclear regulatory proteins arises as a prototype for the action of lineage-restricted transcription factors, with important roles during primitive and definitive hematopoiesis, and at different stages of hematopoietic cell differentiation (reviewed by Vicente et al., 2012).

GATA factor family

The GATA family of transcription factors comprises evolutionarily conserved proteins with crucial roles in the development and differentiation of eukaryotic cells. They are named after their ability to bind the consensus DNA sequence WGATAR through their zinc-finger domains (Ko et al., 1993). There are six recognized members of the GATA family in mammals, which are divided into two subfamilies based on their expression profile and domain structure (Morrisey et al., 1996, Lowry et al., 2000, Molkentin et al., 2000, Ohneda et al., 2002, Cantor et al., 2002, Grogan et al., 2002, Patient et al., 2002). *GATA1*, *GATA2* and *GATA3* are particularly relevant in hematopoietic development, while *GATA4*, *GATA5* and *GATA6* are mainly expressed in tissues of mesodermal and endodermal origin such as heart, liver, lung, gut and gonads, where they play critical roles in regulating tissue-specific gene expression. With the exception of *GATA5*, null mutations for each of the *GATA* genes results in embryonic lethality in mice, indicating that they are crucial transcriptional regulators of cell development (Pevny et al., 1991, Tsai et al., 1994, Pandolfi et al., 1995, Molkentin et al., 1997, Laforest et al., 2011, Koutsourakis et al., 1999).

The hematopoietic GATA factors

GATA1 is mainly expressed in cells committed towards the definitive erythroid and megakaryocyte lineages (reviewed by Bresnick et al., 2012). Disruption of *GATA1* in embryonic stem (ES) cells induces a primitive and definitive erythropoiesis block, with only occasional appearance of proerythroblasts (Simon et al., 1992). GATA1 supports megakaryocyte differentiation at multiple stages, controlling growth of immature megakaryocytic progenitor cells and mediating terminal maturation of megakaryocytes and platelet formation by interacting with FOG1 (Muntean et al., 2005, Kuhl et al., 2005). In addition, *GATA1* is expressed in the more differentiated eosinophilic and mast cell lineages (reviewed by Bresnick et al., 2012).

GATA2 expression has a broad distribution among hematopoietic cells, with a particularly prominent expression in early progenitors, as well as in megakaryocytes and in mast cell lineages. Furthermore, it is crucial for the proliferation and maintenance of hematopoietic stem and progenitor cells (Bresnick et al., 2012).

GATA3 is the only GATA factor expressed in T lymphocytes. It contributes to the development of T lymphocytes at different stages of differentiation, controlling the appearance of the most early T-cell progenitors in the thymus from the thymus-seeding progenitors, promoting its advance through double negative stages and favoring the arise of $CD4^+$ single positive cells at the expense of $CD8^+$ single positive cells. It is also the only GATA factor expressed in natural killer (NK) cells, where it contributes to their maturation in bone marrow and thymus (Hosoya et al., 2010).

The expression of these termed hematopoietic GATA factors is not limited to the hematopoietic compartment, as they are present in other tissues where they play important roles in development. *GATA1* is expressed in the Sertoli cells of the testis; *GATA2* is expressed in endothelial cells and in the central nervous system, placenta, fetal liver, and fetal heart; and *GATA3* is present in kidney, central nervous system, placenta, skin and mammary gland (reviewed by Vicente et al., 2012).

Both *GATA1* and *GATA2* have alternative first exons depending on the host cell, suggesting a tissue-specific regulation (Shimizu et al., 2005, Vicente et al., 2012). In the case of GATA3, only one alternative first exon has been confirmed, although other two can be found in the Gene database.

Partially overlapping functions of GATA factors

Despite that GATA members have a restricted expression pattern, the study of their function in a biologically relevant context has suggested that many properties of these factors are roughly interchangeable. Although only GATA1 and GATA2 are normally expressed in megakaryocytes, forced expression of GATA3 was able to induce megakaryocytic differentiation of a murine myeloid cell line (Visvader et al., 1993). In a murine model where *GATA3* cDNA substituted *GATA1* locus and thus, *GATA3* was expressed under *GATA1* regulatory elements, GATA3 was able to restore embryonic yolk sac cell numbers to 70% of normal levels, compared to only a 10% in *GATA1* null embryos (Tsai et al., 1998). Another murine model with a 5% of normal *GATA1*

expression showed that embryonic lethality was rescued when *GATA2* or *GATA3* transgenes were expressed, but adult mice presented thrombocytopenia and anemia with accumulation of immature red cells (Takahashi et al., 2000). Finally, GATA3 activates its own transcription in T cells, repressing genes associated with Th1 development (e.g.: *IFNG*), and favoring Th2 development with the expression of IL4 and IL5. Although heterologous GATA factors are not significantly expressed in T cells, GATA1, GATA2 and GATA4 were able to induce endogenous *GATA3* activation (without induction of other endogenous GATA factors), and promote Th2 development (Ranganath et al., 2001). This apparent partial interchangeability of GATA factors under certain conditions suggests that they share some functional abilities, but also that each GATA factor exerts its function in a tissue-specific manner, probably due to their differential regulation.

GATA2 structure and characteristics

The human *GATA2* gene was first cloned from an umbilical vein endothelial cell library, and was found highly homologous to the chicken *Gata2* (Lee et al., 1991). Subsequent isolation of *GATA2* cDNA allowed gene mapping and protein prediction, as a first approach of structural determination (reviewed by Bresnick et al., 2012).

The GATA2 gene

The human *GATA2* gene is located within chromosome 3 (3q21.3), and spans 13,760 kilobases (kb) containing 8 different exons (Figure 2). The first three are alternatively transcribed depending on the host tissue, and exons 7a and 7b differ in their length. The most distal first alternative exon is termed IS, and it is specifically transcribed in hematopoietic tissue. The second alternative exon is named IG, and is more broadly transcribed (Minegishi et al., 1998). These two isoforms contain the 7a exon and are translated into the same protein (although each protein has its specific designation, NP_116027.2 and NP_001139133.1, respectively). However, the most proximal third alternative exon is transcribed along with the shorter 7b exon, and gives rise to a slightly smaller protein (NP_001139134.1).

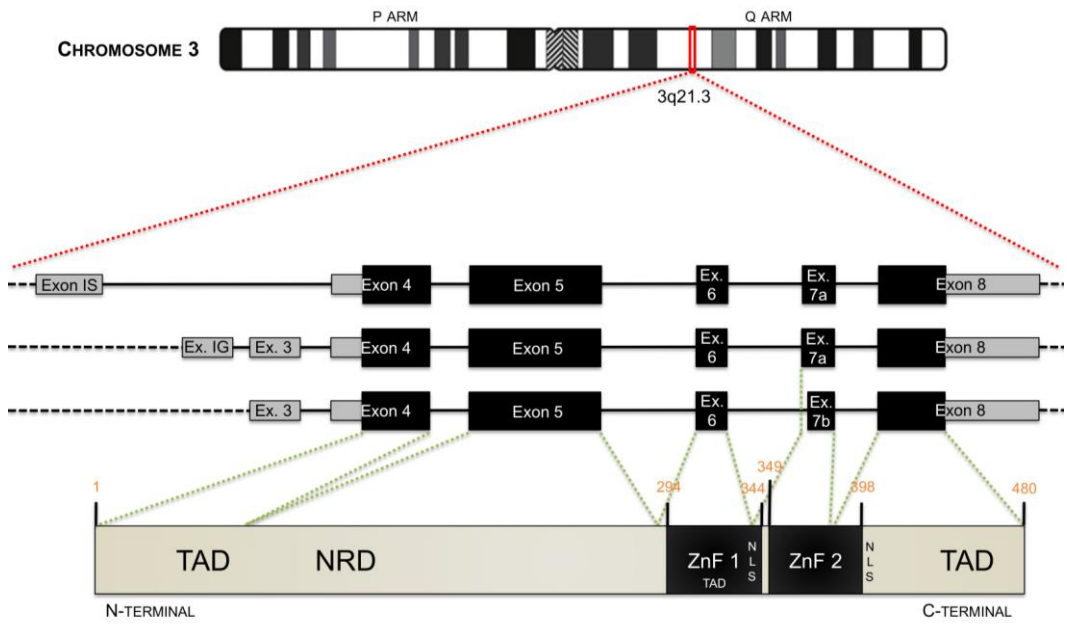

Figure 2: Schematic representation of *GATA2* location, transcripts and protein structure. Human *GATA2* is located in chromosome 3, gives rise to three different transcripts depending on the cellular context (shown in the middle). Non-translated region is shown in gray, while coding sequence is depicted in black. These transcripts can be translated into two different proteins (with the long isoform depicted below). GATA2 protein has two zinc-finger domains (ZnF), three transactivation domains (TAD), one negative regulatory domain (NRD) and two putative nuclear localization signals (NLS). Amino acid positions of relevant GATA2 characteristics are depicted in orange color.

The GATA2 protein

The longer GATA2 protein isoform contains 480 aminoacids, and the most relevant features are two zinc-finger (ZnF) domains of the Cys-X_2-Cys-X_{17}-Cys-X_2-Cys type, which belong to the treble clef fold group of ZnF domains (Figure 2) (Krishna et al., 2003). It has been postulated that both ZnF domains were originally a single motif, which was duplicated during evolution and nowadays have apparently diverged to incorporate discrete functions (Yang et al., 1992). The C-terminal ZnF (ZnF2) domain is critical for DNA binding to the WGATAR consensus sequence. On the other hand, the N-terminal ZnF (ZnF1) domain also shows independent DNA binding ability (Chlon et al., 2012), but its main role is to stabilize overall DNA binding, discriminating among different binding sites for subtle regulation of genes. ZnF1 also harbors a transactivation domain (TAD) to stimulate transcription (Yang et al., 1992). In addition, GATA2 has other two TADs outside the ZnF1, in the N and C terminal regions. All three TADs may function synergistically by properly positioning the protein to make a precise interaction with the DNA sequence (Yang et al., 1992). Between the N-terminal TAD and ZnF1, a negative regulatory domain (NRD) is able to moderate the production of early hematopoietic cells in mouse embryos, and shows little

resemblance with the NRDs of other GATA factors; the highest homology is shared with GATA3 at a 46%, and it has been reported as a TAD, not NRD (Minegishi et al., 2003). Finally, being a nuclear transcription factor, GATA2 may possess two nuclear localization signals. Although no functional analyses have been reported, bioinformatic tools identify putative conserved nuclear localization signals at the end of each ZnF domain. In fact, GATA1 harbors two conserved basic sequences in the C-terminal region of both ZnF domains, and subcellular localization analyses showed that each nuclear localization signal could direct nuclear translocation independently (Visvader et al., 1995). GATA3 also harbors two basic amino acid clusters containing sequence similarity to the conserved nuclear localization signal, and their simultaneous deletion (along with both ZnF domains) affected nuclear localization (Yang et al., 1994).

The shorter GATA2 protein isoform possesses 466 aminoacids, lacking 14 residues that bridge both ZnFs (amino acids 340 to 353). Despite its depiction in the Gene database, little is known about its function and expression pattern throughout tissues with presence of GATA2.

Regulation of GATA2 and function as a transcription factor

Regulation of *GATA2* is a complex process that includes transcriptional regulation to initiate, sustain or shut down *GATA2* expression, and post-transcriptional and post-translational regulation through the interaction with microRNAs (miRNAs) and other proteins. This regulatory network is oriented to achieve a context-specific transcriptional function in hematopoietic cells.

Transcriptional regulation of GATA2 expression

Various studies have established the role of NOTCH1, BMP4 and EVI1 in the activation of *GATA2* in primitive hematopoiesis. Notch1 and its ligands Jag1 and Jag2 are expressed at the ventral wall of the dorsal aorta in mouse embryos at E9.5-10.5 with a characteristic patched pattern that is shared with other hematopoietic transcription factors. In these cells with Notch1 expression, Notch1 binds *Gata2* promoter and activates its transcription for the generation of murine intra-embryonic hematopoiesis (Robert-Moreno et al., 2005). BMP4 signaling, which is crucial for hematopoietic and vascular development, has been also implicated in activating the expression of *Gata2* (Maeno et al., 1996). Finally, Evi1, which is highly expressed in the yolk-sac and para-aortic splanchnopleural region, also contributes to *Gata2* activation during embryonic development

binding to an enhancer located 7 kb far from *Gata2* transcription start site (TSS) (Figure 3) (Yuasa et al., 2005).

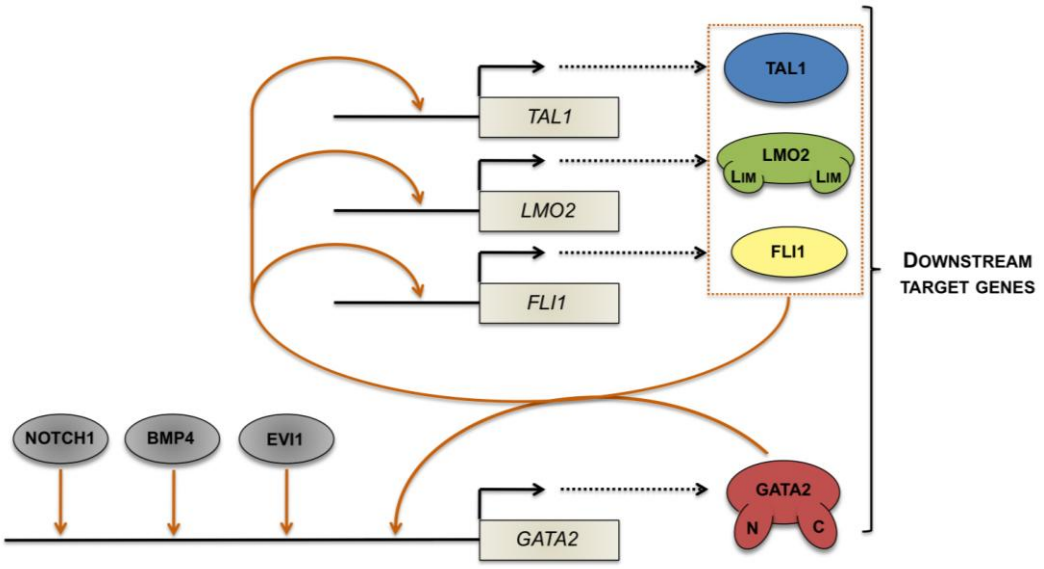

Figure 3: Activation of *GATA2* and cross-regulation with cofactors. The expression of *GATA2* initiates by the action of NOTCH1, BMP4 and EVI1. Once this signaling ceases, GATA2 can maintain its own transcription cross-activating TAL1, LMO2 and FLI1 cofactors. GATA2 can regulate downstream target genes in cooperation with these cofactors.

Post-transcriptional regulation of GATA2

In addition to transcriptional regulation, *GATA2* can be regulated at the mRNA and protein level. The translation of *GATA2* can be blocked by miRNAs, and GATA2 interacts with different proteins to enhance or inhibit its function. The translated protein can suffer covalent modifications to regulate DNA-binding activity, protein stability, subcellular localization and association with cofactors. GATA2 can also interact with other proteins to form multimeric complexes that do not imply covalent modifications of the protein.

Interactions with microRNAs

There are three studies about the role of miRNAs in the regulation of *GATA2* expression. miRNAs are small non-coding RNAs that regulate gene expression at post-transcriptional levels, either by repressing translation or by inducing sequence-specific degradation of target mRNAs (Bartel et al., 2004). A study in zebrafish identified two closely located miRNAs, miR-144 and miR-451, which were transcribed as a single transcript. Of those, only miR-451 was shown to decrease *Gata2* expression and delay erythropoietic maturation *in vivo* (Pase et al., 2009). Of note, it was previously described that Gata1 upregulated miR-144

and miR-451 expression (Doré et al., 2008). It was later found that the SON protein, a negative regulatory element-binding protein, and GATA2 itself, repressed the transcription of the miR-27a-24 cluster in hematopoietic stem and progenitor or immature erythroid progenitor cells. As erythropoiesis proceeded, GATA1 directly activated miR-27a~24 transcription, and this involved a GATA1-mediated displacement of GATA2 (this replacement process is termed GATA switch, see page 14). Furthermore, the mature miR-27a and miR-24 cooperatively inhibited GATA2 translation (Ahn et al., 2013, Wang et al., 2013). In summary, these data indicate that Gata1 downregulates *Gata2* expression indirectly via miRNAs in differentiated cells, while SON maintains higher levels of GATA2 by repressing miRNAs.

Interactions introducing covalent modifications

The first described covalent modification of GATA2 was the phosphorylation in hematopoietic progenitor cells (Figure 4).

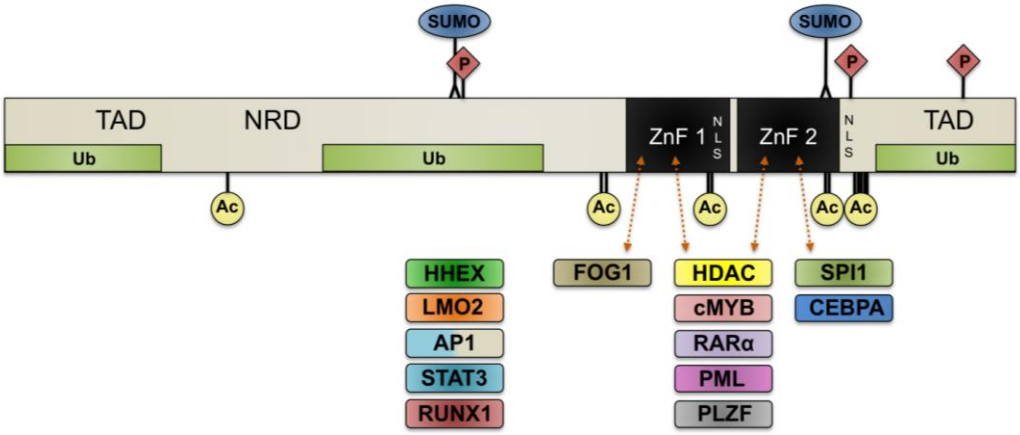

Figure 4: Covalent modifications of GATA2 and interaction with other proteins. GATA2 can be covalently modified by phosphorylation (P), sumoylation (SUMO), acetylation (Ac), and ubiquitinilation (Ub) at marked sites. Moreover, GATA2 can interact with other proteins through one or both zinc finger domains (dotted orange arrows) or through a non-specified domain.

GATA2 is phosphorylated in response to IL3-dependent activation of ERK signaling pathway by MAPK, enhancing the transactivational activities of a GATA-dependent CAT promoter (Towatari et al., 1994). Akt phosphorylates GATA2 at Ser401 in response to insulin, which reduces the nuclear localization of GATA2, blocking its transactivation ability, allowing adipogenesis, and reducing expression of proinflammatory cytokines (Menghini et al., 2005). Finally, CDK4/Cyclin D1 complexes phosphorylate T457 of GATA2, allowing proteasome-dependent degradation of GATA2 in the G1 phase. Newly synthesized IG exon derived GATA2 is phosphorylated at S227 by CDK2/Cyclin

A2 complexes in the S phase, but the low activity of anaphase-promoting complex/cyclosome (APC/C) in this phase prevents its degradation, increasing its concentration until the G2/M phase, where the activity of APC/C is resumed, interacts with Cyclin A2, and GATA2 is degraded. Conversely, due to the low CDK activity during the G0 phase, IS exon derived GATA2 is more stable, allowing a higher expression. Thus, this showed that GATA2 expression is cell cycle-dependent in synchronized leukemic cells, cultured cord blood $CD34^+$ cells, and in lin^- mouse bone marrow cells (Koga et al., 2007).

Another covalent modification, SUMOylation, was described to occur in endothelial cells by interaction of the N- and C- terminal regions of PIASγ with the C-terminal region of GATA2. PIASγ exerted a SUMO E3 ligase activity, covalently modifying GATA2 probably through two predicted SUMOylation sites, MKME (aminoacids 221 to 224) and MKKE (aminoacids 388 to 391). This modification was preferentially with SUMO2 rather than SUMO1, and it suppressed GATA binding site-dependent activation of endothelin 1 (ET1) promoter activity, without altering GATA2 DNA binding ability (Chun et al., 2003).

Moreover, GATA2 can be acetylated by p300 and GCN5 (Hayakawa et al., 2004). p300 acetylated lysines throughout the entire protein, namely K102, K281, K285, K334, K336, K389, K390, K399, K403, K405, K406, K408 and K409, while GCN5 acetylated preferentially the C-terminal region of GATA2. Acetylation of GATA2 by p300 enhanced its DNA binding and transactivation abilities of a reporter bearing eight GATA responsive elements. Once the pattern of acetylation by p300 was determined, it was suggested that a homologous protein of p300, CREBBP, would be able to acetylate GATA2 to activate ectodermic genes required for primitive erythropoiesis in Xenopus (Dalgin et al., 2007). Before these studies, it was also suggested that the interaction of GATA2 with HDAC3 could deacetylate not only histones but GATA2 itself, decreasing its activity (Ozawa et al., 2001).

Finally, ubiquitinilation was the latest covalent modification described for GATA2. It was reported that three regions of GATA2, encompassing aminoacids 1 to 70, 153 to 256 and 412 to 480, promoted ubiquitin-mediated proteasome degradation, and thus were termed degrons. The presence of these degrons was associated with a short half-life of GATA2 in multiple cell lines. The first and the third degrons overlapped with TADs, concluding that the transcription of GATA2 target genes may be coupled with its degradation, which occurs with other transcription factors (Minegishi et al., 2005).

Protein-protein interactions including GATA2

In addition to interactions with proteins that covalently modify the GATA2 protein, GATA2 associates with other cofactors to exert its function. Both ZnF domains can bind DNA, but they can also mediate association with other proteins to form multimeric regulatory complexes. GATA2 interacts with proteins involved in hematopoiesis through one or both ZnF domains (Figure 4). ZnF1 interacts with FOG1 (Chang et al., 2002). ZnF2 interacts with SPI1 (Zhang et al., 1999) and CEBPA (Tong et al., 2005). Moreover, there are some proteins that are able to bind both ZnFs to associate with GATA2, such as HDAC3 and HDAC5 (Ozawa et al., 2001), MYB (Kitajima et al., 2002), PML (Tsuzuki et al., 2000), PLZF (Tsuzuki et al., 2002) and RARα (Tsuzuki et al., 2004). Finally, although association with STAT3 (Ezoe et al., 2004), HHEX (Minami et al., 2004), LMO2 (Osada et al., 1995), AP1 (Masuda et al., 2004), and RUNX1 (Wilson et al., 2010) has been described, the GATA2 region required for the interaction has not been elucidated. In the case of STAT3, the homologous ZnF1 of GATA1 also interacts with STAT3, suggesting that it could be the same region for GATA2. These interactions can result in either activating (AP1, FOG1, PML, RARα) or inhibitory (STAT3, HHEX) complexes. The opposite is also valid in certain cases, since the interaction can be activating or inhibitory depending on the target. For instance, CEBPA is able to activate PPARγ to promote adipogenesis, but the interaction with GATA2 inhibits this activation (Tong et al., 2005). Conversely, GATA2 boosts the activation of an artificial CEBPA-responsive reporter (Greif et al., 2011). On the other hand, it has been described that SPI1 inhibits GATA2 transactivation function of an artificial GATA binding site reporter (Zhang et al., 1999), but they also cooperate to activate CSF1R transcription (Hahn et al., 2011). Thus, the role of GATA2 in multimeric complexes appears highly dependent on the cofactors and targets to which it is bound.

GATA2 as transcription factor

GATA2 functions as a transcription factor, and various approaches have explored the basics of its mechanism of action along with GATA1 and GATA3. The first notable fact is that despite the large number of WGATAR motifs throughout the whole genome (rising up to 6,976,111 motifs), ChIP based studies revealed that few are occupied by GATA1 or GATA2 (Figure 5) (Wozniak et al., 2008, Fujiwara et al., 2009).

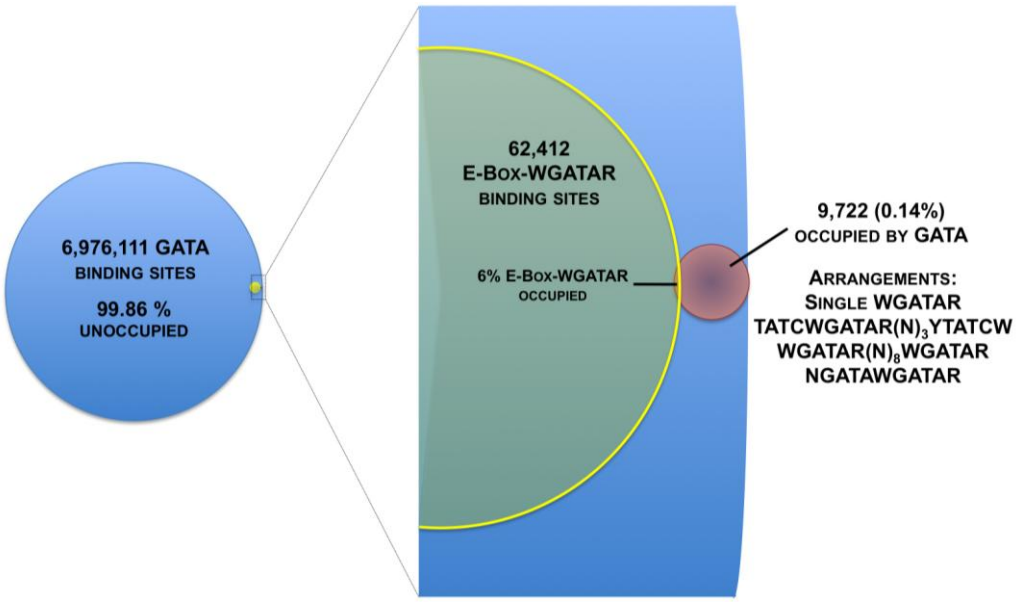

Figure 5: Occupancy and heterogeneity of GATA binding sites in the human genome. The total number of GATA binding sites in the human genome is depicted as a blue circle. GATA binding sites present different arrangements, including single WGATAR sequences and double motifs. E-box-WGATAR composite elements are shaded in yellow color. The red circle represents occupied GATA sites with any arrangement, 6% of which correspond to E-Box-WGATAR composite elements.

In the case of GATA1, only 0.14% of all WGATAR consensus sequences are occupied; occupancy was predominant in introns (37%) and more than 1 kb away from RefSeq genes (47%) (Fujiwara et al., 2009). The genome displays different arrangements of GATA binding motifs. Single WGATAR motifs can be found, where the ZnF2 is responsible for DNA binding. Moreover, three other arrangements have been described in vertebrates, in which both ZnF2 and ZnF1 binding would be required to maximize binding stability (Trainor et al., 1996). The first disposition corresponds to a complete and an incomplete motif forming a partially palindromic box and an additional motif separated by 3 base pairs (bp) from the palindromic motif. They are relatively infrequent but GATA1 showed highest binding affinity. The second disposition is a non-overlapping direct repeat separated by 8 bp with a strong interaction with GATA1; and the third includes a

complete GATA site and a second overlapping partial direct repeat with a relatively weak interaction with GATA1. The determination of the crystallographic structure of GATA1 and GATA3 has also helped to elucidate the configuration of the DNA-protein complex. Notable similarities were found in their binding disposition, where the zinc module bound to the major groove and the basic tail to the minor groove. The same study revealed a dimerization interface, which was important for GATA self-association when binding adjacent WGATAR consensus sequences (Bates et al., 2008).

GATA factors bind their consensus sequences as part of heteromultimeric protein complexes. GATA1 forms complexes with E2A, TAL1, LMO2 and LDB1, named LDB1 complexes (Love et al., 2014). There would be two DNA binding modules, provided by E47-TAL1 heterodimer (binding to E-box, with a CANNTG sequence) and GATA1 (binding to a WGATAR sequence). LMO2 and LDB1 would link the two DNA binding moieties, with LMO2 directly interacting with GATA1 (Love et al., 2014) (Figure 6). In a study analyzing GATA1 binding to DNA, 6% of the 9,722 GATA1 occupied WGATAR sites corresponded to E-box-WGATAR composite elements, which accounted for a 0.93% of all the 62,412 composite elements present in the genome (Wozniak et al., 2008).

The proposed mechanism of E-protein-GATA nucleoprotein complex assembly for GATA1 was supposed to be similar for GATA2 (Wozniak et al., 2008). In fact, GATA2 was shown to interact directly with LMO2 (Osada et al., 1995), and TAL1 occupancy in composite elements was predictive of GATA2 occupancy (Wozniak et al., 2008). Nevertheless, the presence of GATA1 or GATA2 in similar multimeric protein complexes exert overlapping as well as exclusive effects, suggesting subtle regulatory mechanism divergences (see GATA switch).

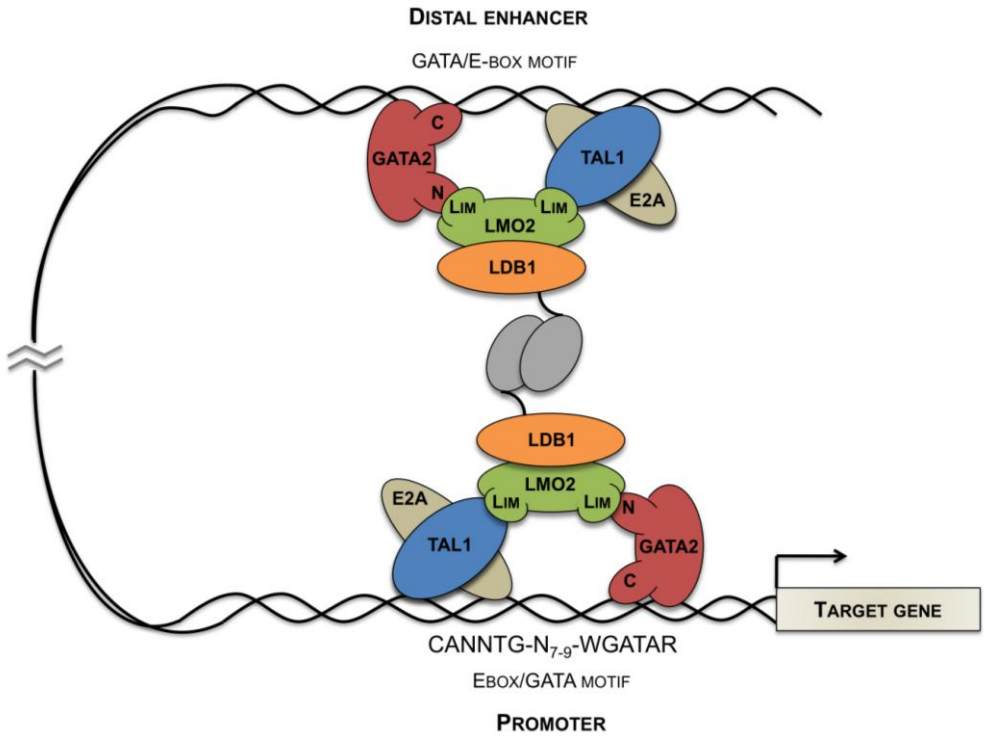

Figure 6: Participation of GATA2 in multimeric complexes for gene regulation. GATA2 takes part in protein complexes to bind the E-box-WGATAR composite elements of the genome. This binding can occur at proximal promoters and distal enhancers, where chromatin looping brings complexes together and exert transcriptional activity (adapted from Love et al., 2014).

The functionality of E-box-WGATAR composite elements has been directly related to their structure: it was very sensitive to the distance between the E-box and the WGATAR sequences and their mutual orientation (Wozniak et al., 2008). In addition to the E-box, other motifs are detected near composite elements: 37.5% of GATA2-occupied composite elements showed SP1 binding sites when considering ±250 bp of flanking sequence, compared to only 9.4% of SP1 binding sites in GATA2-unoccupied sites. A motif not related to protein binding or transcription occurred within ±100 bp of the composite elements in 62.5% of GATA2-occupied versus 4.7% of GATA2-unoccupied sites, and had a [TC][CT][CT]TG[GT][GC][CG][AT]G[TG] sequence (Wozniak et al., 2008). Finally, GATA2 occupancy was associated with diacetylated H3, tetraacetylated H4, dimethylated H3K4, which are commonly described as active chromatin marks (Wozniak et al., 2008, Bresnick et al., 2006). In contrast, trimethylated H3K9, which is often present at repressed chromatin sites (Bresnick et al., 2006), was enriched at 50% of GATA2-unoccupied sites, while only was enriched at 6.25% of GATA2-occupied sites (Wozniak et al., 2008).

The study of partners of GATA2 was widened in a more recent report, which integrated genome-wide computational analysis derived from ChIP-seq of ten

hematopoietic transcription factors, and acetylated H3K9 and H3K14 (for the search of accessible genomic regions). They found that a heptad of transcription factors formed by TAL1, LYL1 (a TAL1 ortholog), LMO2, GATA2, RUNX1, ERG, and FLI1 bound to specific regions, of which nearly all contained GATA and ETS binding sites, 76% E-box elements, and 39% RUNX1 binding motifs. Moreover, these regions were next to 927 genes enriched in HSC-specific expression (Wilson et al., 2010).

In summary, these data indicate that the mechanism of action of GATA2 as a transcription factor comprises binding to single or complex GATA motifs and forming protein complexes with TAL1 and ETS factors in E-box-WGATAR composite elements, exquisitely distinguishing among the large number of WGATAR sites of the genome. Also, GATA2 binding would favor the presence of active chromatin marks.

The *GATA switch*

The mechanism of action of the hematopoietic GATA factors presented above partially illustrates the complex processes associated with these proteins at molecular level. Probably, as a consequence of that complexity, few direct transcriptional targets of GATA2 have been validated. The most prominent target is *GATA2* itself: this autoregulatory loop has been extensively studied in murine models, and it serves as an example to further understand the mechanism of action of GATA2 and related GATA factors (Grass et al., 2003, Pal et al., 2004, Martowicz et al., 2005, Grass et al., 2006, Wozniak et al., 2007, Snow et al., 2010, Snow et al., 2011). These studies reported that Gata2 activates its own transcription in hematopoietic stem and early progenitor cell stages. Using G1E cells as a model, this self-sustained activation of *Gata2* was quickly abrogated in retrovirally transduced G1E cells, in which *Gata1* expression was induced in an estrogen dependent manner. Upon ectopic *Gata1* expression Gata2 was substituted by Gata1 in the *Gata2* locus along with chromatin modifications, and was termed the "GATA switch".

In normal erythroid development, GATA switch occurs at the entrance to the proerythroblast stage, where it activates erythroid-specific genes and contributes to the repression of genes associated with HSC maintenance (Figure 7A) (Bresnick et al., 2010, Love et al., 2014). The GATA factor interchange is primarily caused by their differential stability: GATA2 is much less stable than GATA1, and the non-finger regions appeared to contribute to this difference

(Minegishi et al., 2005). It was later suggested that GATA1 could replace GATA2 during the M or G_1 phases of the cell cycle: GATA2 would be quickly degraded at those stages, while GATA1 would remain stable to repress *GATA2* transcription. Successive cell cycles could lead to shut down *GATA2* transcription and promote mature erythroid development (Figure 7B) (Koga et al., 2007).

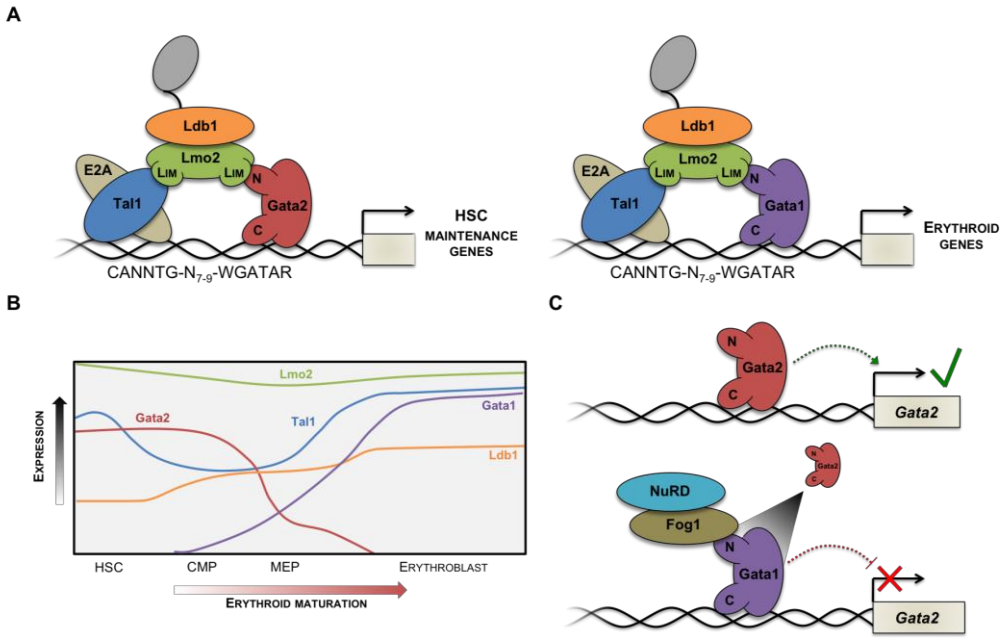

Figure 7: The GATA switch model. (A) Gata1 and Gata2 participate in similar multimeric complexes, but those containing Gata2 regulate the expression of HSC maintenance genes, while those containing Gata1 regulate erythroid genes. (B) Graph representing the expression of proteins present in multimeric complexes containing GATA factors. Gata2 expression is highest in immature cells, and as cells become committed to the erythroid lineage, *Gata1* expression is induced whereas *Gata2* is repressed. (C) Schematic representation of the GATA switch at the murine *Gata2* locus. Gata2 activates its own transcription in early hematopoietic cells. The activation of Gata1 triggers the downregulation of *Gata2* by displacement of the Gata2 protein from its own promoter (adapted from Love et al., 2014).

Among the more than 80 WGATAR sites 80 kb upstream and 30 kb downstream the murine *Gata2* locus, five relevant regulatory domains were located: 77, 3.9, 2.8 and 1.8 kb upstream and 9.5 kb downstream of IS transcription start site (TSS) in the murine genome (Table 1). Each had its specific GATA binding site configuration. In non-induced G1E cells (with no expression of Gata1), Gata2 bound to all these sites with the highest strength at -1.8, associating with Crebbp and Fog1. The association with Crebbp would aid to maintain acetylated histone marks, and complementarily, dimethylation marks of H3K4 were low in general. These histone marks are commonly associated with an accessible chromatin, and this was indicated by the higher DNAse I hypersensitivity in all analyzed sites.

Upon *Gata1* induction, Gata1 concomitantly replaces Gata2 with a similar kinetics in all sites, with the highest occupancy at -77 and -3.9. Fog1 is crucial in this substitution process (Figure 7C): although it is dispensable for Gata2 autoregulation, Gata1 requires the presence of Fog1 to replace Gata2, since Gata1 overexpression in Hox11 immortalized *Fog1*$^{-/-}$ hematopoietic cell precursors decreases *Gata2* transcription by a 30%, compared to a 90% repression in G1E cells, which express Fog1 (Pal et al., 2004). The substitution by Gata1 induces different chromatin modifications in each region. This could be caused by the interaction of FOG1 with subunits of the nucleosome remodeling and histone deacetylase (NuRD) complex (Vicente et al., 2012). As a whole, -77, -3.9, -2.8, -1.8 and +9.5 regions function as four distinct regulatory domains, with -2.8 and -1.8 regions having partially overlapping functions (Table 1). -3.9, -2.8 and -1.8 displayed cell type-specific enhancer activities, while -77 had a broader activity (both in MEL and G1E cells) and +9.5 had the broadest activity, (MEL, high activation in G1E and also in endothelial cells). The GATA2-specific activity, the strong GATA2 binding and the necessity to achieve maximal GATA2 expression show that the -2.8 and -1.8 regions are particularly relevant for GATA2 positive autoregulation in hematopoiesis.

Table 1: Summary of GATA switch.

Site summary			3 GATA motifs: **WGATAR*** NGATAN WGATAGATAN	2 GATAmotifs: **WGATAR** **WGATAR** Opposite strands, separated by 15 bp	2 GATA motifs: **WGATAR** **WGATAR** Opposite strands, separated by 14 bp with a CCAAT box	2 GATA motifs: **WGATAR** **WGATAR** Opposite strands, separated by 3 bp	IS exon Active specifically in hematopoietic cells	IG exon Active in all GATA2 expressing cells	3 GATA motifs: NGATAR **E-box-WGATAR** WGATAN
GATA2 binding		Active	+	+	++	+++	-	-	+
		Inactive	-	-	-	-	-	-	-
GATA1 binding		Active	-	+/-	+/-	+/-	-	-	-
		Inactive	+++	+++	++	+	-	-	+
CREBBP binding		Active	++	+	+	+	-	+/-	+/-
		Inactive	++	++	+/-	+/-	-	+/-	+
FOG1 binding		Active	+	+	+	+	-	-	+
		Inactive	++	+	++	+	-	-	+
DNA hypersensitivity		Active	Yes	Yes	Yes (weak)	Yes	ND	ND	ND
		Inactive	Yes	Yes	No	No	ND	ND	ND
Histone marks	AcH3	Active	+	++	+	+	+++	+++	+
		Inactive	+	+	-	-	++	+	-
	AcH4	Active	+	+	+	+	++	+	+
		Inactive	+	-	-	-	+	-	-
	$H3K4Me_2$	Active	+	+	++	+	++	++	+
		Inactive	++	+++	++	+	++++	+++	+
Activity in MEL†			+	+	-	-	ND	ND	+
Activity in G1E			+	-	+	+	ND	ND	+++

*Bold GATA motifs correspond to functionally relevant sites.
†MEL cells only express GATA1, while G1E only express GATA2.

Other GATA2 targets

In addition to self-activation, GATA2 forms a cross-regulatory network that includes transcriptional target activation and protein-protein interactions with gene products of those targets. The first example is the aforementioned Bmp4, which activates *Gata2* in primitive hematopoiesis. In turn, Gata2 binds and activates *Bmp4*, as it was shown in murine ES cell derived hemangioblasts. This constituted a positively regulated loop that activated mesodermal genes and shut down ectodermal and endodermal genes (Lugus et al., 2007).

GATA2 expression initiates a transcriptional program that includes activation of critical cofactors. It was first described that Gata2 induced the expression of a crucial hematopoietic gene, *Tal1*, through a 3' enhancer that directed expression to dorsal plate mesoderm entering the hemangioblast stage in *Xenopus* embryos. *In vitro* experiments showed that Gata2, in combination with Fli1 and Elf1, bound to an enhancer of *Tal1* located 19 kb downstream of TSS and activated *Tal1* transcription (Göttgens et al., 2002). Once activated, *TAL1* expression is maintained in hematopoietic cells with erythroid potential, including HSCs, multipotential progenitors (MPP), common myeloid progenitors (CMP), megakaryocyte/erythroid progenitors (MEP), and nucleated erythroid cells (Zhang et al., 2005).

Fli1, a member of the ETS factor family, is activated by Gata2 through an enhancer located 12 kb downstream of TSS (Pimanda et al., 2007). The same study suggested that Notch1 and Bmp4 would activate *Gata2* and *Fli1* expression, which would combine to initiate *Tal1*. These three factors could then maintain the expression of each other in HSCs after Notch1 and Bmp4 signaling would have ceased.

LMO2 is included in the GATA2 target and GATA2 partner list. *Lmo2* is activated by Tal1/Lmo2/Gata factor binding to multiple distal hematopoietic enhancers at the hemangioblast stage (Landry et al., 2009), and it is also upregulated by GATA2 at the progenitor stage (Huang et al., 2009).

Finally, another example of gene regulated by GATA2 and GATA2 partners in hematopoiesis is *RUNX1*, which is crucial for the emergence of definitive HSCs. An intragenic enhancer controlled by Gata2, Ets proteins, Tal1, Lmo2 and Ldb1 was responsible for Runx1 expression in the dorsal aorta of murine embryos (Nottingham et al., 2007).

In summary, GATA2, TAL1, ETS factors, LMO2 and RUNX1 cross-activate and cooperate forming multimeric complexes that also contain other factors such as E2A and LDB1, called LDB1 complexes (Love et al., 2014). They regulate the transcription of downstream targets in the developing hematopoietic and endothelial cells. GATA2 also regulates transcriptional targets through LDB1 complexes in other cellular contexts: the promoter region of *ANGPT2* was responsible for GATA2 containing complex-dependent activation in endothelial cells (Simon et al., 2008, Deleuze et al., 2012).

It has been shown that GATA2 directly regulates targets at intermediate differentiation stages of hematopoiesis, but the complexes involved have not been investigated. For instance, GATA2 inhibits erythroid differentiation repressing *EPO* expression (Imagawa et al., 2002). On the other hand, downregulation of *Gata2* in the murine G1E cell line decreased the expression of genes related to megakaryocytic potential such as *Slamf1*, and upregulation of genes related to myeloid potential *Spi1* and *Hhex*. Moreover, Gata2 binds directly to cis-regulatory regions of these genes (Huang et al., 2009). The regulation of *Spi1* by GATA factors was further characterized in another study: Both Gata1 and Gata2 were able to bind *Spi1* at two conserved regions and both were able to repress *Spi1* expression. However, Gata1 exerted a more profound downregulation of *Spi1* than Gata2. Gata2 binding was associated with an undifferentiated state, intermediate levels of *Spi1* expression, and low expression of myeloid genes. Gata1 expression replaced Gata2 at *Spi1* locus and induced erythromegakaryocytic differentiation, extinguishing *Spi1* expression (Chou et al., 2009).

Finally, in the context of cancer, it was described that GATA2 and GATA1 bound to a 3' enhancer of *WT1*, which activated the expression of *WT1* in both hematopoietic and solid tumor cell lines. However, there was a low correlation index between *WT1* and GATA1 or GATA2 expression in 23 acute myeloid leukemia (AML) and 3 acute lymphoid leukemia cases (Furuhata et al., 2009).

GATA2 requirement in hematopoiesis

The information shown above anticipates the indispensable role that GATA2 exerts in normal hematopoiesis. Various approaches have determined the effects of a deregulated *GATA2* expression above or below normal levels at different hematopoietic stages, and they have shown that *GATA2* expression must be within a narrow window to ensure an appropriate cell function.

Mouse models

Mice homozygous for a disrupted *Gata2* fail to develop beyond E11.5, dying from severe anemia. Moreover, Gata2 disrupted ES cells do not contribute to primitive (E13) and definitive (beyond E17) hematopoiesis when injected into normal embryos. Their impaired response to c-kit ligand or stem cell factor was suggested as responsible for cell defects (Tsai et al., 1994). Another mouse model used a version of Cre recombinase fused to a tamoxifen-sensitive ligand binding domain of the estrogen receptor, whose expression was under the control of the +9.5 regulatory element. Upon endogenous +9.5 element activation and tamoxifen exposure, *Gata2* was deleted and this led to anemia, hemorrhage, and eventual death (E16.5) in edematous embryos, involving HSC loss in the fetal liver and defective lymphatic development (Lim et al., 2012). The relevance of the +9.5 element was further highlighted in another mouse model. The homozygous deletion of the +9.5 element demonstrated that it enhanced *Gata2* expression and was required to generate long-term repopulating HSCs in the aorta-gonad-mesonephros (Gao et al., 2013). Both studies underscored the role of Gata2 in HSC. In fact, another study using Vec-Cre and Vav-Cre systems to conditionally delete *Gata2* reported that Gata2 function was essential for the generation of HSC during the stage of endothelial-to-hematopoietic cell transition and for HSC survival once they were generated (de Pater et al., 2013).

On the other hand, haploinsufficiency of *Gata2* perturbs both embryonic and adult HSC homeostasis. A haploid dose of *Gata2* markedly reduces HSC production and expansion in the aorta-gonad-mesonephros region, where the first HSCs are generated (Ling et al., 2004). Although haploinsufficient mice are born at normal Mendelian ratios, haploinsufficient cells show a decrease in the granulocyte-macrophage progenitor colony forming ability, suggesting progenitor cell impairment. Competitive transplants of haploinsufficient and wild type (WT) cells show additional defects: 7 weeks after transplantation, a lower contribution of haploinsufficient HSC and HPC is detectable. Between 4 and 8 months, there is a pronounced decrease in HPC contribution. At 11 months, there is a reduced contribution of specific cell lineages: the monocyte compartment is the most affected, followed by the B-cell compartment and T-cell compartment. On the other hand, the granulocyte compartment is not affected. Primitive haploinsufficient cells are quiescent and undergo a higher rate of apoptosis than their WT counterparts, with decreased expression of the antiapoptotic gene *Bcl-*

xL, although both display similar self-renewal ability (Rodrigues et al., 2005, Rodrigues et al., 2008).

Effects of GATA2 ectopic expression in hematopoiesis

The expression of GATA2 above normal levels also alters hematopoietic cell development. Murine stem/progenitor bone marrow cells transduced to stably express ectopic GATA2 display proliferation and differentiation blocking, and hematopoietic repopulation of lethally irradiated mice is delayed. In fact, cells responsible for hematopoietic repopulation expressed ectopic GATA2 at relatively low levels (Persons et al., 1999). Although high levels of GATA2 do not induce apoptosis, expression of *c-myc*, *Skp2* and *Cul1* is reduced (Ezoe et al., 2002). In both studies, transduced cells with a GATA2 expression in the lower range were able to proliferate and cycle more frequently. These effects on proliferation and cell cycle were later assessed in human cord blood samples. Although $p21^{Cip1}$ and $p27^{Kip1}$ components were not induced, cell cycle components CDK4 and CDK6 were repressed upon ectopic GATA2 expression, and the proliferation was effectively slowed down by modulation of *Mef* and *Hes1* (Tipping et al., 2009). In summary, these studies suggested a role of GATA2 in inducing quiescence at the primitive cell stage with a dose-dependent effect.

During normal hematopoietic development, *GATA2* expression is highest in the most immature population, where it induces quiescence. *GATA2* expression can decrease each time a HSC enters cell cycle, which would gradually allow HSCs to undergo progressive differentiation (Koga et al., 2007). This partial decrease in *GATA2* expression also allows cells to actively proliferate, entering the progenitor stage (Tsai et al., 1997). At the progenitor stage, however, *GATA2* overexpression impairs proper differentiation of hematopoietic cells. Ectopic expression of *GATA2* under estrogen control inhibits chicken erythroid progenitor cell differentiation (Doré et al., 2011). Macrophage differentiation of murine ES cells is also blocked by *Gata2* overexpression in the presence of macrophage colony-stimulation factor (MCSF) (Kitajima et al., 2006). On the other hand, although overexpression of *GATA2* leads to excessive megakaryocytic progenitor expansion, maturation is not blocked in this cell compartment (Huang et al., 2009). In summary, in order to achieve proper hematopoietic progenitor proliferation and differentiation, *GATA2* expression must decrease. The most notable exception is the mast cell lineage, where *Gata2* expression increases significantly during the differentiation of murine bone marrow cells towards mast

cells (Ohmori et al., 2012). The other exception where *GATA2* is also expressed is the megakaryocytic differentiation (Terui et al., 2000).

GATA2 and hematological diseases

Given the importance of an adequate *GATA2* expression during normal hematopoiesis, aberrant GATA2 regulation or function can lead to disease. Indeed, a study showed that the expression of GATA2 in patients suffering from aplastic anemia was lower, compared to healthy controls. Aplastic anemia is a bone marrow condition characterized by a fatty replacement of HSCs that leads to pancytopenia. The loss of direct repression of the adipogenic transcription factor *PPARγ* by the lower GATA2 expression was suggested as responsible for the fatty replacement of bone marrow (Fujimaki et al., 2001). It was later reported that this downregulation of *GATA2* could be caused by an increase in the IFNγ expression (Xu et al., 2009). On the other hand, abnormal expression of *GATA2* was observed in 40 out of 45 patients with severe myelodysplastic syndrome (MDS) (Fadilah et al., 2002).

Contribution of GATA2 to myeloid malignancies

GATA2 contributes to myeloid malignancies through various mechanisms. Overexpression of *GATA2* is a recurrent event in AML, with a poor prognostic impact in both adult and pediatric AML. Moreover, recent studies have found somatic *GATA2* mutations in chronic myeloid leukemia in blast crisis (CML-BC) and in AML. These mutations are strongly associated with a biallelic *CEBPA* mutation status (bi*CEBPA*), a rare subgroup of normal karyotype AML. Finally, several families with a history of early development of MDS, chronic myelomonocytic leukemia (CMML) and/or AML displayed germline *GATA2* mutations.

Overexpression of *GATA2* in acute myeloid leukemia

Two studies found *GATA2* expression in 54/62 AML patients by qRT-PCR (Shimamoto et al., 1995) and in 21/41 AML patients (Ayala et al., 2009). Our group was the first to correlate *GATA2* expression with cytogenetic data, molecular markers, and prognosis in a cohort of 259 patients with AML. Overexpression of *GATA2* (97/259 patients, 37.4%) was associated with clinical features related to poor prognosis in AML. It was more prevalent in the poor cytogenetic risk group (46.1%), in elderly patients and in those with secondary AML, and it was correlated with *WT1* and *EVI1* overexpression. Moreover,

GATA2 overexpression was associated with a significantly lower overall survival (OS) and event-free survival (EFS) compared to patients with normal *GATA2* expression in a sub-cohort of 112 AML patients. Finally, *GATA2* overexpression was also highly represented in AML with normal karyotype (23/57 patients, 40%, overexpressed *GATA2*), and there was also associated with worse OS, EFS, and disease-free survival (DFS). Consistent with data in patients, 5 out of 11 (45.5%) AML cell lines overexpressed *GATA2* (Vicente et al., 2012).

The expression of GATA2 has been also analyzed in 237 pediatric AML patients. *GATA2* expression was higher in 65% of patients, compared to normal bone marrow. Moreover, high expression of *GATA2* at diagnosis was an independent poor prognostic factor for OS, EFS and DFS. This was particularly evident in patients with French-American-British (FAB) M5 classification, inversion of the chromosome 16, or high *WT1* expression (Luesink et al., 2012).

Somatic mutations of *GATA2* in myeloid malignancies

In addition to *GATA2* overexpression, acquired (somatic) mutations of *GATA2* have been found in myeloid malignancies (Table 2 and Figure 8). Two *GATA2* mutations were first reported in CML patients in accelerated phase or blast crisis. The most common was the p.Leu359Val mutation, which was present in 8/85 (9%) cases. The p.Ala341_Gly346del in-frame deletion was detected in only one patient. p.Leu359Val is a missense mutation that affects the ZnF2 domain of GATA2, which is critical for DNA binding and protein association. Functional analyses showed that p.Leu359Val was a gain-of-function mutation, with enhanced DNA binding ability, higher transactivation activity of target genes, and a more effective inhibition of the ability of SPI1 to activate the MCSFR promoter. Moreover, p.Leu359Val ectopic expression conferred resistance to the monocytic differentiation of the HL60 cell line upon vitamin D3 or all-trans retinoic acid (ATRA) induction (Zhang et al., 2008, Hahn et al., 2011). Of note, p.Leu359Val was not found in other myeloid malignancies (Zhang et al., 2009).

Table 2: List of reported somatic *GATA2* mutations and associated malignancy.

Mutation (DNA)*	Mutation (protein)	Mutation type	Affected ZF domain	Reference	*CEBPA* status	Cases/total studied	Diagnosis
c.605G>C	p.Ala203Pro	Missense (NC)	None	Green et al., 2013	SM	1/153	AML
Unknown	p.His258fs	Frameshift	None	Green et al., 2013	SM	1/153	AML
c.878G>A	p.Arg293Gln†	Missense (NC)	None	Greif et al., 2012	DM	1/160	CN-AML
c.911C>A	p.Pro304His	Missense (NC)	1	Yan et al., 2011	NM	1/112	AML-M5
c.920G>T	p.Arg307Leu	Missense (NC)	1	Fasan et al., 2013	DM	1/212	CN-AML/IRK-AML
c.918C>T	p.Arg307Trp	Missense (NC)	1	Green et al., 2013	DM	1/153	AML
c.923G>C	p.Arg308Pro	Missense (NC)	1	Fasan et al., 2013	DM	1/212	CN-AML/IRK-AML
				Green et al., 2013	DM (2/2)	2/153	AML
				Niimi et al., 2013§	NM	1/96	CN-AML
c.949A>C	p.Asn317His‡	Missense (NC)	1	Greif et al., 2012	DM	1/160	CN-AML
c.950A>T	p.Asn317Ile	Missense (NC)	1	Fasan et al., 2013	DM	1/212	CN-AML/IRK-AML
c.950A>G	p.Asn317Ser	Missense (NC)	1	Fasan et al., 2013	DM	1/212	CN-AML/IRK-AML
c.952G>A	p.Ala318Thr‡	Missense (NC)	1	Greif et al., 2012	DM (3/3)	3/160	CN-AML
				Green et al., 2013	DM (2/2)	2/153	AML
					SM	1/153	AML
c.953C>G	p.Ala318Gly†	Missense (C)	1	Greif et al., 2012	DM	1/160	CN-AML
c.953C>T	p.Ala318Val	Missense (C)	1	Luesink et al., 2012§	DM	1/230	Pediatric *de novo* AML-M2
				Greif et al., 2012	DM	1/160	CN-AML
				Fasan et al., 2013	DM (3/3)	3/212	CN-AML/IRK-AML
				Green et al., 2013	DM (3/3)	3/153	AML
c.953_954insTCC	p.Cys319_Gly320insSer	In-frame insertion	1	Yan et al., 2011	SM	1/112	AML-M5
c.958G>A	p.Gly320Asp	Missense (NC)	1	Greif et al., 2012	DM (3/3)	3/160	CN-AML
				Fasan et al., 2013	DM	1/212	CN-AML/IRK-AML
c.959G>T	p.Gly320Val	Missense (C)	1	Green et al., 2013	DM	1/153	AML
c.961C>G	p.Leu321Val	Missense (C)	1	Fasan et al., 2013	DM	1/212	CN-AML/IRK-AML
				Greif et al., 2012	DM	1/160	CN-AML
c.961C>T	p.Leu321Phe	Missense (NC)	1	Greif et al., 2012	DM	1/160	CN-AML
				Fasan et al., 2013	DM (4/5)	5/212	CN-AML/IRK-AML
				Green et al., 2013	DM (3/3)	3/153	AML
c.962T>A	p.Leu321His	Missense (NC)	1	Fasan et al., 2013	DM	1/212	CN-AML/IRK-AML
c.962T>C	p.Leu321Pro	Missense (NC)	1	Greif et al., 2012	DM	1/160	CN-AML
				Green et al., 2013	SM	1/153	AML
c.962T>G	p.Leu321Arg	Missense (NC)	1	Fasan et al., 2013	DM	1/212	CN-AML/IRK-AML
c.983A>C	p.Gln328Pro	Missense (NC)	1	Greif et al., 2012	DM	1/160	CN-AML

Table 2 (cont.)

Mutation (DNA)*	Mutation (protein)	Mutation type	Affected ZF domain	Reference	*CEBPA* status	Cases/total studied	Diagnosis
c.989G>A	p.Arg330Gln	Missense (NC)	1	Greif et al., 2012	DM	1/160	CN-AML
				Fasan et al., 2013	DM	1/212	CN-AML/IRK-AML
				Green et al., 2013	DM	1/153	AML
c.989G>T	p.Arg330Leu	Missense (NC)	1	Green et al., 2013	DM	1/153	AML
					SM	1/153	AML
c.989G>C	p.Arg330Pro	Missense (NC)	1	Green et al., 2013	DM	1/153	AML
c.1021_1038del18	p.Ala341_Gly346del	In-frame deletion	1	Zhang et al., 2008	Unknown	1/85	CML-BC
c.1061C>A	p.Thr354Lys	Missense (NC)	2	Fasan et al., 2013	NM	1/212	CN-AML/IRK-AML
c.1069A>T	p.Thr357Ser	Missense (C)	2	Shiba et al., 2013	NM	1/157	Pediatric *de novo* AML-M4
c.1075T>G	p.Leu359Val	Missense (C)	2	Zhang et al., 2008	Unknown	8/85	CML-BC
c.1082G>A	p.Arg361His	Missense (C)	2	Fasan et al., 2013	NM	1/212	CN-AML/IRK-AML
c.1084C>G	p.Arg362Gly	Missense (NC)	2	Luesink et al., 2012§	NM	1/230	Pediatric *de novo* AML-M4
				Shiba et al., 2013§	NM	1/157	Pediatric *de novo* AML-M3
c.1085delG	p.Arg362fsX24	Frameshift	2	Luesink et al., 2012§	NM	1/230	Pediatric *de novo* AML-M4
c.1085G>C	p.Arg362Pro	Missense (NC)	2	Luesink et al., 2012§	NM	1/230	Pediatric *de novo* AML-M4
				Shiba et al., 2013§	NM	2/157	Pediatric *de novo* AML-M1/M4
c.1085G>A	p.Arg362Gln	Missense (NC)	2	Luesink et al., 2012§	DM	1/230	Pediatric *de novo* AML-M1
				Fasan et al., 2013	DM (2/3)	3/212	CN-AML/IRK-AML
				Yan et al., 2011	NM	2/112	AML-M5
				Green et al., 2013	DM	1/153	AML
				Shiba et al., 2013	SM	1/157	Pediatric *de novo* AML-M1
c.1086_1087insCGA	p.Arg362dup	In-frame insertion	2	Green et al., 2013	SM	1/153	AML
c.1123C>A	p.Leu375Ile	Missense (C)	2	Shiba et al., 2013	SM	1/157	Pediatric *de novo* AML-M3
c.1136T>A	p.Leu379Gln	Missense (NC)	2	Fasan et al., 2013	DM	1/212	CN-AML/IRK-AML
c.1154C>T	p.Pro385Leu	Missense (NC)	2	Green et al., 2013	SM	1/153	AML

C, conserved amino acid type in missense mutation; NC, non-conserved amino acid type; NM, non-mutated; SM, single-mutated; DM, double-mutated; CN-AML, cytogenetically normal acute myeloid leukemia; IRK-AML, intermediate risk karyotype acute myeloid leukemia; CML-BC, chronic myeloid leukemia in blast crisis.
*The RefSeq sequence used as reference was NM_032638.4.
†These *GATA2* mutations were presented simultaneously in the same patient (Greif et al., 2012).
‡These *GATA2* mutations were presented simultaneously in the same patient (Green et al., 2013).
§The origin of several mutations described by Luesink et al. (2012), Niimi et al. (2013) or Shiba et al., (2013) could not be determined due to lack of germline or remission material. However, due to their similarity to other somatic mutations, they are listed in this table.

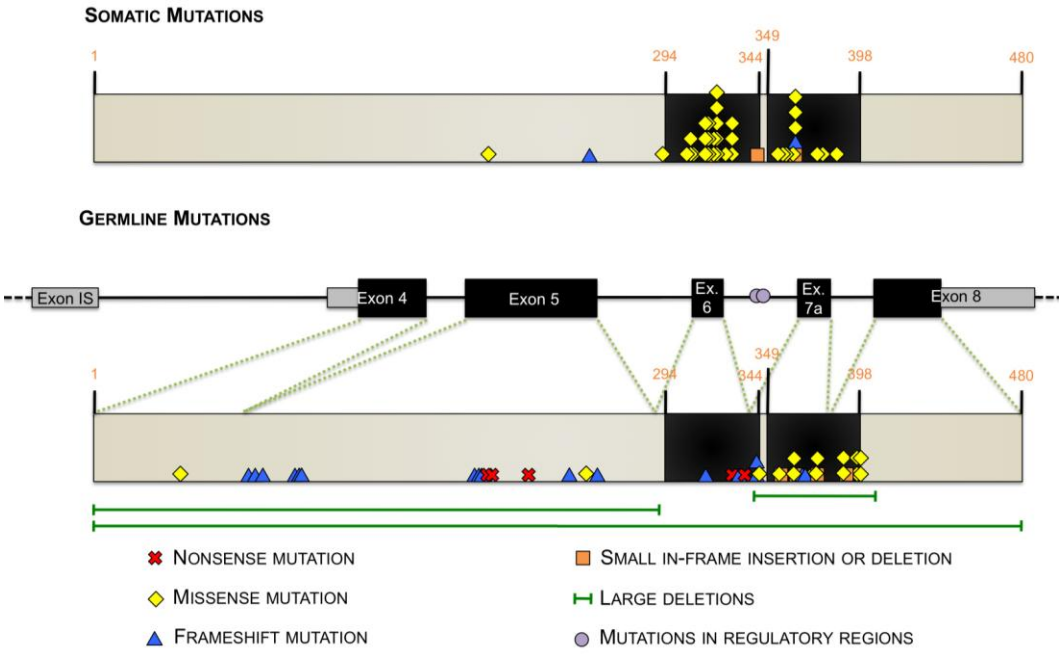

Figure 8: Schematic representation of somatic and germline *GATA2* mutations. Up to 38 different somatic mutations have been reported in sporadic myeloid malignancies, where there is a high frequency of missense mutations in the ZnF1 (21). On the other hand, 50 different germline mutations have been detected. Most germline mutations involve frameshift mutations before ZnF2 (15) and missense mutations in ZnF2 (12).

In AML, somatic *GATA2* mutations are more common in cases with double mutations of the *CEBPA* gene (bi*CEBPA*), a rare alteration in AML, which accounts approximately for 4% of all AMLs. This would explain that two groups did not detect mutations in large series of patient samples (Zhang et al., 2009; Vicente et al., 2012), while other two groups found *GATA2* mutations in less than 3% of cases (Yan et al., 2011; Luesink et al., 2012). However, recent studies found that somatic *GATA2* mutations were common in bi*CEBPA*-mutated AML, ranging between 18-40% of patients (Greif et al., 2012, Fasan et al., 2012, Green et al., 2013). Of note, our group found the R307W *GATA2* mutation in the MUTZ3 AML cell line, which was also present in the study published by Green et al., in 2013. The most common group of sporadic mutations in *GATA2* in AML corresponds to missense mutations. To date, 33 different missense somatic mutations have been found in AML, being the ZnF1 especially affected with 21. Other mutations include frameshift and in-frame insertions (Table 2).

Missense *GATA2* somatic mutations showed a reduced ability to enhance CEBPA-dependent activation of transcription. Indeed, *CEBPA* targets were altered in the bi*CEPBA* AML patients with *GATA2* mutated, compared to *GATA2* non-mutated bi*CEBPA* AML patients. Bioinformatic analyses suggested that

these somatic missense mutations could have hindered DNA binding ability through their ZnF1 (Greif et al., 2012).

GATA2 mutations were associated with an immature phenotype of leukemic blasts and AML FAB subtypes M1 and M2 (Fasan et al., 2012; Grossmann et al., 2013). Interestingly, the presence of bi*CEBPA* and *GATA2* mutations excluded the presence of internal tandem duplications of FLT3 (FLT3-ITD) (Greif et al., 2012, Grossmann et al., 2013). In fact, patients with bi*CEBPA* and *GATA2* mutations had a better OS compared to patients with bi*CEBPA* and WT-*GATA2* (Fasan et al., 2012).

In summary, a total of 38 different *GATA2* somatic mutations have been described in myeloid malignancies (Table 3). The p.Leu359Val is of particular relevance in progressing CML. However, most *GATA2* mutations in AML are found in AML with bi*CEBPA* mutations, are missense mutations affecting ZnF1, and are associated with a better OS in bi*CEBPA* AML cases.

Table 3: Summary of *GATA2* mutations.

Type of GATA2 mutation	Missense	Frameshift	Non-sense	In-frame insertion	In-frame deletion	Intronic mutation	Complete deletion of GATA2 locus	Total
Somatic	**33** 8 conservative 25 non-conservative 2 before ZnF1 21 located in ZnF1 10 located in ZnF2	**2** 1 before ZnF1 1 in ZnF2	**-**	**2** 1 in ZnF1 1 in ZnF2	**1** In ZnF1	**-**	**-**	**38**
Germline	**15** 2 conservative 13 non-conservative 2 before ZnF1 1 located in ZnF1 12 located in ZnF2 38 families reported	**17** 11 before ZnF1 4 located in ZnF1 2 located in ZnF2 17 families reported	**5** 3 before ZnF1 2 located in ZnF1 7 families reported	**-**	**6** 1 before ZnF1 1 overlaps ZnF1 and ZnF2 4 located in ZnF2 7 families reported	**2** 1 deletion of 28 bp 1 single nucleotide variant 5 families reported	**4** 4 families reported	**50**

Germline mutations of *GATA2*

GATA2 has been recently added to the list of germline mutations that predispose to MDS and AML, together with *TERT*, *TERC*, *RUNX1* and *CEBPA*, other genes that code for transcription factors with important roles in hematopoiesis (Holme et al., 2012). Germline GATA2 mutations have been associated with a quartet of diseases, termed Monocytopenia and Mycobacterium Avium Complex (MonoMAC) infection (Hsu et al., 2011), Dendritic Cell, Monocyte, B and NK Lymphoid (DCML) deficiency (Bigley et al., 2011, Dickinson et al., 2011), Emberger syndrome (primary lymphedema with MDS) (Ostengaard et al., 2011) and Familial MDS/AML (Hahn et al., 2011). Although the Emberger syndrome was identified in 1979, it was not until recent years that germline mutations of *GATA2* were identified as its cause. Similarly, historical reports of familial MDS/AML are now known to be caused by mutations in *GATA2* (Kaur et al., 1972, Robinson et al., 1983, Horwitz et al., 1996). Moreover, these diseases are recognized now as facets of a single genetic disorder (*GATA2* mutation) that may manifest heterogeneously even within a single pedigree (Ishida et al., 2012, Mutsaers et al., 2013).

To date, 50 different germline *GATA2* mutations have been reported, with more than 70 families affected (Table 3, Table 4, and Figure 8). The most common mutations include missense mutations, with 38 families reported, affecting particularly the ZnF2 domain. The second common type corresponds to frameshift mutations, which give rise to truncated proteins due to an early stop of protein translation, usually before ZnF2. Other mutation types include non-sense mutations, in-frame insertions or deletions, complete deletions of the *GATA2* locus (including neighboring genes), and mutations in regulatory regions. Notably, mutations in regulatory regions affect an intronic region between exons 6 and 7, homologous to the +9.5 enhancer element described above in the murine *Gata2* gene. Two mutations were described in that enhancer, including a 28 bp deletion that encompassed the E-box element, and a C>T transition, which disrupted an ETS factor binding element. Both altered elements flank a WGATAR sequence in the non-mutated genome, and are 7 bp and 20 bp far from the GATA binding site, respectively. The discovery of mutations in regulatory regions has suggested *GATA2* haploinsufficiency as the basis of syndromes caused by *GATA2* mutations (Johnson et al., 2012, Hsu et al., 2013).

Table 4: List of reported germline *GATA2* mutations and associated syndromes.

Mutation (DNA)*	Affected exon	Mutation (protein)	Mutation type	Affected ZnF	Reference	Cases/families	Associated syndrome
c.1-200_871+527del2033	4, 5	p.Met1_Ser290del	In-frame deletion	None	Vinh et al., 2010, Hsu et al., 2011, Kazenwadel et al., 2012, Hsu et al., 2013	2/1	MonoMAC
c.121C>G	4	p.Pro41Ala	Missense (NC)	None	Holme et al., 2012	3/1	Familial MDS/AML
c.230-1_230insC	5	p.Arg78ProfsX107	Frameshift	Not translated	Ostengaard et al., 2011	3/1	Emberger syndrome
c.243_244delAinsGC	5	p.Gly81GlyfsX103	Frameshift	Not translated	Vinh et al., 2010, Cuellar-Rodriguez et al., 2011, Hsu et al., 2011, Hsu et al., 2013	1/1	MonoMAC
c.257_258delGC	5	p.Arg86ProfsX98	Frameshift	Not translated	Dickinson et al., 2014	1/1	GATA2 deficiency
c.302delG	5	p.Gly101AlafsX16	Frameshift	Not translated	Hsu et al., 2013	1/1	MonoMAC
c.310_311insCC	5	p.Leu105ProfsX15	Frameshift	Not translated	Holme et al., 2012	2/1	Familial MDS/AML
					Ostengaard et al., 2011	5/1	Emberger syndrome
c.318_319insT	5	p.Ala106CysfsX78	Frameshift	Not translated	Dickinson et al., 2014	2/1	GATA2 deficiency
c.579_580insA	5	p.Ala194SerfsX8	Frameshift	Not translated	Ostengaard et al., 2011	1/1	Emberger syndrome
c.586_593dup	5	p.Gly199LeufsX20	Frameshift	Not translated	Hsu et al., 2013	1/1	MonoMAC
c.594delG	5	p.Gly200ValfsX19	Frameshift	Not translated	Dickinson et al., 2014	3/1	GATA2 deficiency
c.599_600insG	5	p.Ser201X	Nonsense	Not translated	Dickinson et al., 2011, Dickinson et al., 2014	2/1	DCML
c.610C>T	5	p.Arg204X	Nonsense	Not translated	Pasquet et al., 2013	1/1	MonoMAC
c.670G>T	5	p.Glu224X	Nonsense	Not translated	Pasquet et al., 2013	1/1	MonoMAC
c.735_736insC	5	p.Ile246HisfsX36	Frameshift	Not translated	Dickinson et al., 2014	1/1	GATA2 deficiency
c.751C>T	5	p.Pro254Leu	Missense (NC)	None	Hsu et al., 2011, Hsu et al., 2013	1/1	MonoMAC
c.769_778dup	5	p.Tyr260CysfsX24	Frameshift	Not translated	Cuellar-Rodriguez et al., 2011, Hsu et al., 2011, Hsu et al., 2013	1/1	MonoMAC
c.941_951dup	6	p.Asn317fsX11	Frameshift	1	Hsu et al., 2011, Hsu et al., 2013	1/1	MonoMAC
c.988C>T	6	p.Arg330X	Nonsense	1	Pasquet et al., 2013	4/1	MonoMAC
					Spinner et al., 2014	2/1	GATA2 deficiency
c.992_993insGACC	6	p.Leu332ThrfsX53	Frameshift	1	Kazenwadel et al., 2012	1/1	MonoMAC/DCML
c.1009C>T	6	p.Arg337X	Nonsense	1	Ostengaard et al., 2011	1/1	Emberger syndrome
					Hsu et al., 2013, Spinner et al., 2014	1/1	MonoMAC
c.1017+512del28	None	None	Intron 5 deletion	None	Vinh et al., 2010, Johnson et al., 2012, Hsu et al., 2013	3/1	MonoMAC
c.1017+572C>T	None	None	Intron 5 SNV	None	Hsu et al., 2013, Dickinson et al., 2014	10/4	MonoMAC
c.1018-1G>T	7a	p.Thr340_Thr381del	In-frame deletion	1, 2	Hsu et al., 2011	1/1	MonoMAC
					Dickinson et al., 2011	1/1	DCML

Table 4 (cont.)

Mutation (DNA)*	Affected exon	Mutation (protein)	Mutation type	Affected ZnF	Reference	Cases/families	Associated syndrome
c.1018-3_1031del17	7a	p.Ala341ArgfsX38	Frameshift	1	Ostengaard et al., 2011	1/1	Emberger syndrome
					Dickinson et al., 2014	1/1	GATA2 deficiency
c.1019_1022delCGGC	7a	p.Ala341ProfsX45	Frameshift	1	Ostengaard et al., 2011	1/1	Emberger syndrome
c.1024G>A	7a	p.Ala342Thr	Missense (NC)	1	Shiba et al., 2013	3/1	Pediatric *de novo* AML-M0
c.1061C>T	7a	p.Thr354Met	Missense (C)	2	Vinh et al., 2010, Cuellar-Rodriguez et al., 2011, Hsu et al., 2011, Hsu et al., 2013, Mace et al., 2013	5/3	MonoMAC
					Kazenwadel et al., 2012	3/1	MonoMAC/DCML
					Hahn et al., 2011	18/3	Familial MDS/AML
					Holme et al., 2012	4/1	Familial MDS/AML
					Dickinson et al., 2011	1/1	DCML
					Bödör et al., 2012	6/1	GATA2 deficiency
					Dickinson et al., 2014	5/1	GATA2 deficiency
c.1063_1065delACA	7a	p.Thr355del	In-frame deletion	2	Hahn et al., 2011	2/1	Familial MDS/AML
c.1081C>T	7a	p.Arg361Cys	Missense (NC)	2	Hsu et al., 2013, Spinner et al., 2014	1/1	MonoMAC
					Dickinson et al., 2014	1/1	GATA2 deficiency
c.1082G>C	7a	p.Arg361Leu	Missense (NC)	2	Ostengaard et al., 2011	1/1	Emberger syndrome
c.1083_1094del12	7a	p.Arg362_Asn365del	In-frame deletion	2	Hsu et al., 2011, Vinh et al., 2010	1/1	MonoMAC
c.1099_1100insG	7a	p.Asp367GlyfsX15	Frameshift	2	Spinner et al., 2014	2/1	GATA2 deficiency
c.1103_1104insG	7a	p.Val369CysfsX13	Frameshift	2	Hsu et al., 2013	1/1	MonoMAC
c.1113C>G	7a	p.Asn371Lys	Missense(NC)	2	Hsu et al., 2011, Vinh et al., 2010	1/1	MonoMAC
c.1114G>A	7a	p.Ala372Thr	Missense (NC)	2	Dickinson et al., 2014	1/1	GATA2 deficiency
c.1116_1130del15	7a	p.Cys373del5	In-frame deletion	2	Spinner et al., 2014	3/1	GATA2 deficiency
c.1117T>C	7a	p.Cys373Arg	Missense (NC)	2	Ostengaard et al., 2011	1/1	Emberger syndrome
c.1162A>G	8	p.Met388Val	Missense (NC)	2	Kazenwadel et al., 2012, Hsu et al., 2013	2/1	MonoMAC/DCML
c.1163T>C	8	p.Met388Thr	Missense (C)	2	Hsu et al., 2013, Spinner et al., 2014	3/1	MonoMAC
c.1168_1170delAAG	8	p.Lys390del	In-frame deletion	2	Dickinson et al., 2014	1/1	GATA2 deficiency

Table 4 (cont.)

Mutation (DNA)*	Affected exon	Mutation (protein)	Mutation type	Affected ZnF	Reference	Cases/families	Associated syndrome
c.1186C>T	8	p.Arg396Trp	Missense (NC)	2	Vinh et al., 2010, Hsu et al., 2011, Hsu et al., 2013, Spinner et al., 2014	2/1	MonoMAC
					Camargo et al., 2012	1/1	MonoMAC
c.1187G>A	8	p.Arg396Gln	Missense (NC)	2	Cuellar-Rodriguez et al., 2011, Hsu et al., 2011,	1/1	MonoMAC
					Hsu et al., 2013, Spinner et al., 2014	6/2	MonoMAC
					Ishida et al., 2012	1/1	MonoMAC/Emberger syndrome
					Holme et al., 2012	3/1	Familial MDS/AML
					Pasquet et al., 2013	4/1	MonoMAC
c.1192C>T	8	p.Arg398Trp	Missense (NC)	2	Vinh et al., 2010, Cuellar-Rodriguez et al., 2011, Hsu et al., 2011, Hsu et al., 2013	6/5	MonoMAC
					Dickinson et al., 2011, Dickinson et al., 2014	4/2	DCML
c.1193G>A	8	p.Arg398Gln	Missense (NC)	2	Dickinson et al., 2014	5/1	GATA2 deficiency
Complete deletion†	All	Complete deletion	Complete deletion	Not translated	Pasquet et al., 2013	1/1	MonoMAC
					Kazenwadel et al., 2012	1/1	MonoMAC/DCML
						1/1	MonoMAC/DCML
						1/1	MonoMAC/DCML

C indicates conserved amino acid type in missense mutation; NC non-conserved amino acid type; MonoMAC, monocytopenia and mycobacterial infection syndrome; DCML, dendritic cell, monocyte, B and NK lymphoid deficiency; SNV, single nucleotide variant.
*The RefSeq sequence used as reference was NM_032638.4.
†Although deletions are different in each case, all include a heterozygous deletion of the entire GATA2 locus.

Few functional analyses of germline *GATA2* mutations have been performed to date. The most notable study reported that two mutations affecting the ZnF2 domain of GATA2, p.Thr354Met and p.Thr355del, had their DNA binding ability moderately and severely decreased, respectively. Moreover, the transactivating abilities of these mutations were impaired, they were unable to activate GATA2-responsive enhancer elements, and the interaction with SPI1 was hampered. Finally, p.Thr354Met increased HL60 AML cell line proliferation, decreased apoptosis, and blocked the differentiation response upon treatment with ATRA, which promotes granulocytic differentiation. On the other hand, p.Thr355del had an effect comparable to an empty vector, and appeared to be a null mutant under the same conditions (Hahn et al., 2011). Despite the effects of p.Thr354Met in altering cell behavior in certain conditions, the general consensus is that these mutations act as loss-of-function mutations.

Until recently, there has been a lack of comprehensive studies of large series of patients diagnosed with germline *GATA2* mutations. Two studies have analyzed series of 57 (Spinner et al., 2014) and 30 patients (Dickinson et al., 2014), and both reported that these syndromes show an incomplete penetrance of symptoms, which leads to a different diagnosis. It is possible to subdivide the symptoms in four groups: those affecting hematopoiesis, infections derived of the altered hematopoietic compartment, symptoms affecting other parts of the body, and propensity to develop myeloid malignancy.

Patients with *GATA2* mutation can remain healthy for the first decades of life. However, as they age, the chance of symptoms increases (Table 2). After the onset of the syndrome at any age, the 5-year survival rate reaches 91%, 10-year survival 84%, and 20-year survival 67% (Spinner et al., 2014). On the other hand, frameshift mutations favor an earlier onset age (18 years old), compared to missense mutations, (26 years old) (Dickinson et al., 2014).

Table 5: *Evolution of patients with GATA2 mutation (adapted from Spinner et al., 2014).*

Age	Patients without symptoms	Overall survival	Survival without HSCT†
20	50%	96%	89%
30	25%	ND*	76%
40	16%	77%	53%
60	ND	45%	ND

*ND: no data
†HSCT: hematopoietic stem cell transplantation

GATA2 mutated patients can show an altered leukocyte profile, starting with a mild leukopenia in peripheral blood. Analysis of bone marrow shows hypocellularity compatible with diagnosis as aplastic anemia. However, when focusing on different leukocyte compartments, a pattern with more specific alterations appears (Figure 9). A study revealed circulating monocytopenia, dendritic cell cytopenia, and B and NK lymphocytopenia (DCML deficiency) (Dickinson et al., 2014). Analysis of 55 patients detected B lymphocytopenia in 86%, NK lymphocytopenia in 82%, and monocytopenia in 78% of cases, being the last two particularly severe. Moreover, $CD4^+$ lymphocytopenia and neutropenia were identified in 51% and 47% of patients, respectively, but this decrease in cell numbers was generally less marked than in the previous four cell compartments. The lower number of $CD4^+$ cells reduced the $CD4^+/CD8^+$ ratio in the T lymphocyte compartment. The decrease in NK, B and $CD4^+$ T lymphocyte number has been associated with a decreased expansion of multi-lymphoid and granulocyte-macrophage progenitors, with cell profile skewed towards a more mature phenotype (Mace et al., 2013, Spinner et al., 2014, Dickinson et al., 2014). On the other hand, macrophages in the lungs appear at normal levels, but their function could be compromised (Spinner et al., 2014, Dickinson et al., 2014). In the bone marrow, atypical megakaryocytes were detected in 92% of cases, and large granular lymphocyte (LGL) population expansion in 16% (Spinner et al., 2014).

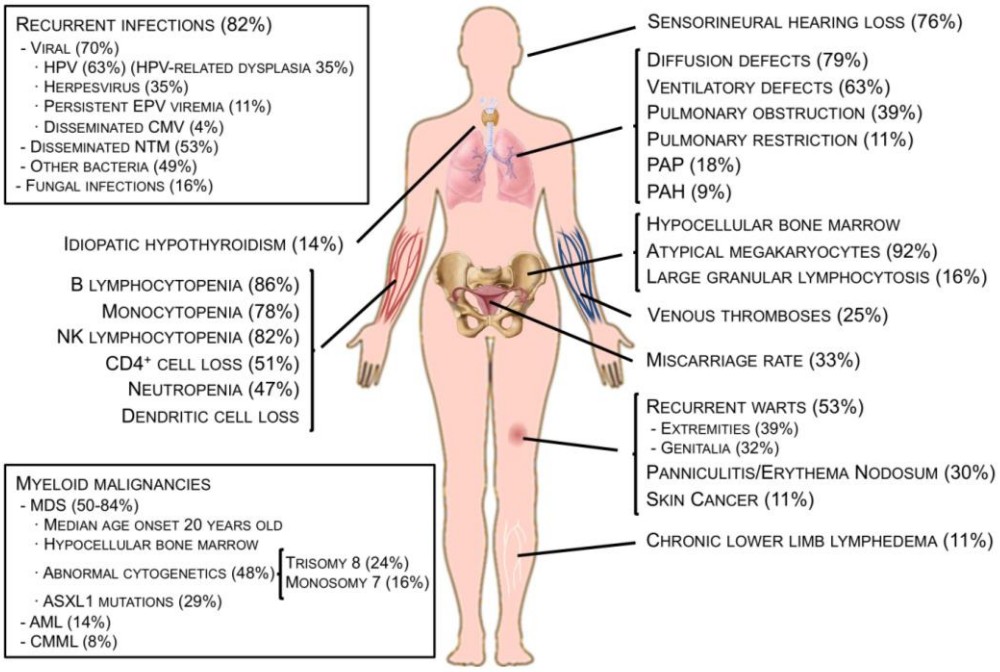

Figure 9: Summary of GATA2 deficiency symptoms: Patients with GATA2 deficiency show a wide variety of symptoms affecting bone marrow and blood, lymphatic and venous vessels, lungs and skin among others. Moreover, they are prone to suffer recurrent infections and develop myeloid malignancies. Percentages shown here correspond to the studies performed by Dickinson et al. and Spinner et al. in 2013, which analyzed a total of 30 and 57 patients, respectively.

Spinner et al. reported that 82% of patients were prone to suffer recurrent, severe and disseminated infections by a reduced number of pathogens: 70% of the patients have viral infections, including Human Papillomavirus (HPV) (63%) and herpes virus (63%), with a persistent Epstein-Barr viremia (EBV) in 11%, and disseminated Cytomegalovirus (CMV) infection in 4% of patients. Among bacterial infections, 53% are disseminated infections of non-tuberculous mycobacteria (NTM), while non-specific bacterial infections account for 49%. Finally, 16% of patients have suffered from invasive fungal infections. Most infection rates remain low until the third or fourth decade, but these infections can be life-threatening in some cases. Susceptibility to viral infections could be related to the loss of NK cells, which are critical for anti-viral immunity. Similarly, the dysfunction in monocytes/macrophages may predispose to infection by non-tuberculous mycobacteria, which are intramacrophagic pathogens (Spinner et al., 2014, Dickinson et al., 2014).

In addition to the symptoms presented above, patients can show other complications. Several accessory symptoms were related to the propensity of patients to develop severe infections. The study of pulmonary performance in 38 patients revealed diffusion (79%) and ventilatory (63%) defects, pulmonary obstruction (39%) and restriction (11%). More severe conditions such as

pulmonary alveolar proteinosis (PAP) or pulmonary arterial hypertension (PAH) were present in 18% and 9% of cases, respectively. Recurrent warts (53%) and skin cancer (11%) were also indicators of underlying infections. 35% presented HPV-related dysplasia and EBV^+ mesenchymal tumors were detected in two patients. Accessory symptoms such as chronic lymphedema, which was presented by 11% of patients, could involve unilateral or bilateral lower extremities and genitalia (Spinner et al., 2014). The appearance of lymphedema has been associated with the requirement of *GATA2* in lymphatic valves, and null (e.g. frameshift) mutations have been associated with its appearance, compared to missense mutations (Kazenwadel et al., 2012). On the other hand, mild to severe sensorineural hearing loss has been detected in 76% of patients (when tests were adjusted to age), venous thromboses in 25% and idiopathic hypothyroidism in 14% of patients. Finally, female patients suffered a higher rate of miscarriages at 33%, compared to 15-20% suspected in the general population (Spinner et al., 2014). Interestingly, it has been reported that Gata2 is expressed in the uterine luminal and glandular epithelium pre-implantation, spatio-temporally co-localizing with that of the progesterone receptor (Rubel et al., 2012).

The most life-threatening condition that these patients can face is MDS, AML or CMML, which they are prone to develop. According to studies, between the 50 and 84% of patients with any type of *GATA2* mutations develop MDS (Figure 9). The median age of MDS presentation is much lower at 20 years old, compared to 65 years old of sporadic MDS. Furthermore, typical MDS displays a hypercellular bone marrow, while MDS in *GATA2*-mutated patients shows a hypocellular bone marrow. 48% of MDS show abnormal cytogenetics, trisomy of chromosome 8 is present in 24% of cases, and monosomy of chromosome 7 is present in 16% of cases. Fourteen percent and 8% of patients develop AML or CMML, respectively, with or without a previous MDS event (Spinner et al., 2014, Dickinson et al., 2014). Mutations in *ASXL1* have been associated with the transformation of MDS to AML, and particularly to CMML. In a study of 48 patients, 29% presented heterozygous mutations in the exon 13 of *ASXL1* in ten different *GATA2* mutation backgrounds, suggesting it as an important second hit in myeloid transformation. Furthermore, the presence of *ASXL1* mutations predicts a rapid progression and poor OS in sporadic MDS/AML, and the same happened in patients with *GATA2* mutation (West et al., 2013).

It has been proposed FLT3L as a disease progression marker, since it was the only serum marker elevated among 118 screened in *GATA2* mutation. As the disease progresses, levels of FLT3L in serum gradually increase, suggesting hematopoietic stress. The loss of specific progenitors and the fact that female patients presented hematopoietic clonality supports this suggestion. Moreover, the severity of symptoms was correlated with FLT3L levels. As MDS arises, levels of FLT3L decrease, possibly due to expansion of cellularity and ligand consumption (Dickinson et al., 2014).

Finally, it has been shown the usefulness of hematopoietic stem cell transplantation (HSCT) in patients who have suffered life-threatening infections or have developed MDS or malignancy (Cuellar-Rodriguez et al., 2012). Although patients who underwent a HSCT improved their response against infections and consequently lung affections decreased, long term survival supposes a challenge: 72%, 65% and 54% survive one, two and four years, respectively (Dickinson et al., 2014).

In summary, germline *GATA2* mutations include loss-of-function mutations, being the missense in the ZnF2 and frameshift before the ZnF2 the most common (Table 5). Individuals affected with this condition experience hematopoietic stress through cell extrinsic and intrinsic mechanisms that gradually drain progenitor cell compartment. This provokes the loss of monocytes, dendritic cells, NK and B lymphocytes, making patients prone to suffer viral, bacterial and/or fungal infections and its complications. Moreover, they show a higher frequency of developing early MDS, AML or CMML.

HYPOTHESIS AND OBJECTIVES

HYPOTHESIS

The GATA2 transcription factor has an essential role in the proliferation and differentiation of hematopoietic cells. GATA2 contributes to myeloid malignancies through various mechanisms. Overexpression of *GATA2* is a recurrent event in acute myeloid leukemia (AML), with a poor prognostic impact in both adult and pediatric cases. Moreover, recent studies have found somatic *GATA2* mutations in chronic myeloid leukemia in blast crisis and in AML with a biallelic *CEBPA* mutation status. Finally, germline *GATA2* mutations have been recently associated with four related familial syndromes that predispose to myeloid malignancies, including AML: familial myelodysplastic syndrome/acute myeloid leukemia, monocytopenia and mycobacterial infection (MonoMAC) syndrome, dendritic cell, monocyte, B and NK lymphoid deficiency (DCML) syndrome, and Emberger syndrome.

We hypothesize that deregulation of either *GATA2* expression or GATA2 aberrant function caused by mutations could alter key downstream target genes, contributing to the pathogenesis of the aforementioned disorders. Thus, our general aim was to identify and functionally characterize novel target genes of the GATA2 transcription factor in AML, and to assess the effect of the *GATA2* mutations in the transcription of known target genes of GATA2.

OBJECTIVES

1. To identify novel target genes of the GATA2 transcription factor in human acute myeloid leukemia.

1.1. To knock-down *GATA2* expression by siRNA in acute myeloid leukemia cell lines, and to quantify the resulting expression of selected candidate genes.

1.2. To determine whether GATA2 regulates the expression of the selected candidate genes at the transcriptional level by chromatin immunoprecipitation and luciferase reporter assays.

1.3. To functionally analyze the role of these target genes of GATA2 in acute myeloid leukemia cells, by studying their effects in proliferation, apoptosis, cell cycle, and differentiation.

2. To analyze the role of the human GATA2 transcription factor in its own transcription.

2.1. To identify homologous GATA2 binding sites between human and murine *GATA2* promoters using bioinformatic tools.

2.2. To determine whether the human GATA2 protein regulates its own expression at transcriptional level by chromatin immunoprecipitation and luciferase reporter assays.

3. To assess the effect of the GATA2 mutations in the transcription of its target genes.

3.1. To analyze the effect of the p.Thr354Met, p.Thr355del, and p.Arg396Gln mutations in the promoter of the *GATA2* gene by luciferase reporter assays.

4. To functionally analyze the p.Arg396Gln GATA2 mutation of a patient with MonoMAC syndrome.

4.1. To quantify the mRNA expression of genes related to *GATA2*, in a patient with MonoMAC syndrome with the p.Arg396Gln mutation.

4.2. To perform an *in silico* analysis of p.Arg396Gln functionality using bioinformatic tools.

4.3. To transduce the HL60 cell line with mutated GATA2 and to assess cell differentiation upon ATRA or TPA treatment.

4.4. To transduce murine hematopoietic stem and progenitor cells with p.Arg396Gln GATA2 mutant and to determine the clonogenic ability, differentiation profile and gene expression of transduced murine cells.

RESULTS

CHAPTER 1: GATA2 ACTIVATES THE TRANSCRIPTION OF *MYB* IN ACUTE MYELOID LEUKEMIA

This manuscript has been submitted to *Haematologica* for its publication.

CHAPTER 1: GATA2 ACTIVATES THE TRANSCRIPTION OF *MYB* IN ACUTE MYELOID LEUKEMIA

This manuscript has been submitted to *Haematologica* for its publication.

GATA2 activates the transcription of *MYB* in acute myeloid leukemia

Xabier Cortés-Lavaud,[1,2] Carmen Vicente,[1] Leire Urquiza,[1] María A. García-Sánchez,[1] Marcos González,[3] Eva Barragán,[4] and María D. Odero[2]

[1]Center for Applied Medical Research (CIMA), University of Navarra, Pamplona, Spain.
[2]Department of Biochemistry and Genetics, University of Navarra, Pamplona, Spain.
[3]Hospital Universitario, Salamanca, Spain and Centro de Investigación del Cáncer-Instituto de Biología Molecular y Celular del Cáncer, Salamanca, Spain.
[4]Molecular Biology Unit, Service of Clinical Pathology, La Fe University Hospital, Valencia, Spain.

Running head: GATA2 activates MYB transcription in AML.

Contact information for correspondence: María D. Odero, CIMA, C/ Pío XII, 55, University of Navarra, 31008-Pamplona, Spain. E-mail: modero@unav.es. Phone: +34-948194700. Fax: +34-948194714.

Word count: 1,325

Number of tables: zero.

Number of figures: three.

Number of supplemental files: one.

Acknowledgements: the work was partially supported by Instituto de Salud Carlos III-Red Temática de Investigación Contra el Cáncer (grant number RD12/0036/0063); and Fundación para la Investigación Médica Aplicada (Spain).

CHAPTER 2: GATA2 GERMLINE MUTATIONS IMPAIR *GATA2* TRANSCRIPTION CAUSING HAPLOINSUFFICIENCY. FUNCTIONAL ANALYSIS OF THE p.Arg396Gln MUTATION

This manuscript has been submitted to *PLoS Genetics* for its publication.

CHAPTER 2: GATA2 GERMLINE MUTATIONS IMPAIR *GATA2* TRANSCRIPTION CAUSING HAPLOINSUFFICIENCY. FUNCTIONAL ANALYSIS OF THE p.Arg396Gln MUTATION

This manuscript has been submitted to *PLoS Genetics* for its publication.

GATA2 germline mutations impair *GATA2* transcription causing haploinsufficiency. Functional analysis of the p.Arg396Gln mutation

Xabier Cortés-Lavaud,[1,2*] Manuel F. Landecho,[3*] Miren Maicas[1], Leire Urquiza,[1] Juana Merino[4], Isabel Moreno-Miralles,[1] and María D. Odero[2]

[1]Division of Oncology, Center for Applied Medical Research (CIMA), University of Navarra, Pamplona, 31008, Spain

[2]Department of Biochemistry and Genetics, University of Navarra, Pamplona, 31008, Spain

[3]Department of Internal Medicine, Clínica Universidad de Navarra, Pamplona, 31008, Spain.

[4]Department of Immunology, Clínica Universidad de Navarra, Pamplona, 31008, Spain.

*These authors equally contributed to the manuscript.

Running head: GATA2 mutated proteins impair their own transcription.

Contact information for correspondence: Maria D. Odero, Department of Biochemistry and Genetics, University of Navarra, Edificio CIMA, C/ Pío XII, 55, 31008-Pamplona, Spain. E-mail: modero@unav.es. Telephone: +34-948194700. Fax: +34-948194714.

Author contributions: Conceived and designed the experiments: XCL, MM, IMM, MDO. Performed the experiments: XCL, LU, IMM. Collected phenotype data and biological samples: MFL. Analyzed the data: XCL, MFL, JM. Wrote the paper: XCL, MFL, MM, MDO.

DISCUSSION

Myeloid malignancies are clonal diseases that arise from genetic and epigenetic alterations that perturb critical processes in hematopoiesis, such as self-renewal, proliferation and differentiation. They include chronic stages, namely myelodysplastic syndromes (MDS), myeloproliferative neoplasms including chronic myeloid leukemia (CML), and chronic myelomonocytic leukemia (CMML), or an acute stage, acute myeloid leukemia (AML). AML can occur *de novo* (approximately 80% of the cases) or follow a chronic stage (secondary AML) (Murati et al., 2012).

GATA2 can contribute through different mechanisms to several of the myeloid malignancies mentioned above. First, overexpression of *GATA2* is a recurrent event in AML and is associated with poor prognosis in this disease (Vicente et al., 2012, Luesink et al., 2012). Second, somatic *GATA2* mutations have been detected in chronic myeloid leukemia in blast crisis (CML-BC) (Zhang et al., 2008), and in biCEBPA mutated AML (Greif et al., 2012, Fasan et al., 2012, etc.). Third, *GATA2* can suffer germline mutations, and this is associated with a predisposition to develop early MDS, CMML and *de novo* or secondary AML (Hahn et al., 2011, Hsu et al., 2011, Dickinson et al., 2011, Ostengaard et al., 2011, Hsu et al., 2013, Dickinson et al., 2014, Spinner et al., 2014).

In an effort to further characterize the role of GATA2 as transcription factor in human myeloid disorders, we sought for the identification of novel target genes of GATA2 in acute myeloid leukemia, and to assess the effect of the GATA2 mutated proteins on the transcription of other known target genes of GATA2, and on hematopoietic stem and progenitor cells.

GATA2 activates the transcription of *MYB* in acute myeloid leukemia

The *GATA2* and *MYB* genes have been reported to be deregulated in human leukemias, including CML, AML and T-cell acute lymphoblastic leukemia (T-ALL) (Manzotti et al., 2012, Vicente et al., 2012, Luesink et al., 2012, Hess et al., 2006, Zuber et al., 2011, Clappier et al., 2007, Lahortiga et al., 2007). Moreover, two studies have pointed to a relationship between these two transcription factors in the CML-BC K562 cell line (Soliera et al., 2008, Lorenzo et al., 2012). Here we show for the first time that the expression of both genes significantly correlate in a series of AML patients, validate these results in two independent series of previously published data of AML patients (Valk et al., 2004, Cancer Genome Atlas Research

Network, 2013), and demonstrate that the GATA2 transcription factor activate the transcription of the *MYB* oncogene in AML cells.

ChIP and luciferase reporter assays indicate that GATA2 drives the expression of *MYB* through binding to two different regions in the *MYB* promoter. The -7534 binding site harbors two consecutive canonical WGATAR motifs, which is one of the possible arrangements of GATA2 binding sites throughout the genome (Trayner et al., 1998). The proximal region II contains two functional GATA2 binding sites that belong to a DNAse I hypersensitive region of the *MYB* locus (Dassé et al., 2012). Interestingly, both sites have a GATG sequence. It has been described that the N-terminal zinc-finger of GATA2 binds to GATC, GATT or GATG sequences. Conversely, the C-terminal zinc-finger of GATA2 binds to WGATAR sequences (Pedone et al., 1997). Thus, our results show that GATA2 likely activates *MYB* transcription through distal and proximal promoter regions, probably using the corresponding zinc-finger domain depending on the specific binding site.

GATA2 knockdown downregulated all *MYB* transcripts in HL60, and the 9B transcript in F36P cells, suggesting that the regulation of this transcript could play a specific role in this AML cell line. The *MYB*-9B transcript contains 121 additional amino acids (between exon 9 and 10) within a MYB protein domain involved in protein-protein interactions and negative regulation. Interestingly, it has been reported that the resulting protein is more stable and more effective in transactivating MYB-regulated promoters than the p75 isoform, suggesting that the overall biological effects of MYB in BCR/ABL-transformed cells might be largely exerted by this isoform (Manzotti et al., 2012). Our results indicate that MYB-9B could also have a specific role in AML, pointing to the need of further functional studies to elucidate the role of this isoform in AML.

Additionally, our functional studies indicate that AML cells are highly dependent on *MYB* expression to sustain leukemogenic growth, and that GATA2 would contribute to AML transformation activating the transcription of *MYB*. *GATA2* knockdown downregulated the expression of MYB at mRNA and protein levels in both F36P and HL60 cells. However, the different baseline expression of *MYB* in these two cell lines could explain that the functional effects differ to some extent: in fact, the resulting *MYB* expression after *GATA2* knockdown in HL60 was still comparable to the baseline levels of *MYB* in F36P. On the other hand, it must be taken into account that GATA2 is not the only MYB activating factor in these cell lines, as other previously described mechanisms of MYB activation could cooperate (Zuber et al., 2011, Dassé et al., 2012). Interestingly, GATA3 forms a protein complex along with

TAL1 and RUNX1 to activate *MYB* in T-ALL (Sanda et al., 2012). Moreover, it has been reported that GATA1, TAL1 and LBD1 containing complexes activate *MYB* in erythroid progenitors (Stadhouders et al., 2012). These data suggest that GATA factors could be actively regulating the expression of *MYB* in normal and malignant hematopoiesis.

In summary, we report for the first time that the GATA2 transcription factor directly binds to the *MYB* promoter region, activating its transcription. Interestingly, mRNA expression levels of both genes displayed a positive correlation in three independent cohorts of AML patients. Our functional studies suggest that AML cells are highly dependent on *MYB* expression to sustain leukemogenic growth, and that GATA2 would contribute to AML transformation in part by maintaining high levels of *MYB*, opening new directions to further understand the mechanisms of GATA2 overexpressing leukemias.

GATA2 missense and small in-frame deletion germline mutations impair *GATA2* transcription. Functional analysis of the p.Arg396Gln mutation

Heterozygous mutations in GATA2 have been identified as the cause of four previously described clinical syndromes: monocytopenia and mycobacterium avium complex infection (MonoMAC) syndrome; dendritic cell, monocyte, B and NK lymphoid (DCML) deficiency; familial MDS and AML; and Emberger syndrome. These syndromes are now recognized as different manifestations of a single genetic disorder with protean disease manifestations (Spinner et al., 2014). However, there is considerable clinical heterogeneity among patients with *GATA2* deficiency, and the molecular basis of these diseases remains undetermined. A recent study suggested *GATA2* haploinsufficiency as the basis of these syndromes, on account of the finding that mutations in intronic regulatory regions of *GATA2* decreased *GATA2* transcript levels (Hsu et al., 2013). Here we show that the human GATA2 protein activates its own transcription through a specific region located at -2.4 kb from IS transcription start site, and that the p.Thr354Met, p.Thr355del, and p.Arg396Gln germline mutations impair *GATA2* promoter activation, confirming that *GATA2* haploinsufficiency would be the basis for these familial syndromes.

Initial efforts to characterize the GATA2 autoregulatory loop were performed in murine models, and found that the regions located 2.8 and 1.8 kb upstream of the IS TSS individually contributed to the transcription of *Gata2* (Grass et al., 2003, Snow et al., 2011). Our results indicate that human GATA2 activates its own promoter

through the -2.4 kb region, homologous to the murine -1.8 region, with a main activation role triggered by the -2409, and secondarily by the -2418 *GATA2* binding site. Moreover, we show that some of the most common GATA2 mutated proteins impair this activation, which would cause *GATA2* deficiency through a haploinsufficiency mechanism. As indicated above, GATA2 haploinsufficiency perturbs normal hematopoietic stem cell homeostasis in murine models (Rodrigues et al., 2005); of note, a comprehensive examination of the clinical features of 57 patients with *GATA2* deficiency found high prevalence of cytopenias and bone marrow failure, supporting the notion that a similar defect exists in human *GATA2* haploinsufficiency (Spinner et al., 2014), although it remains unclear why monocytes, dendritic cells, B cells, and NK cells are preferentially depleted. The finding that our patient with a p.Arg396Gln mutation showed a lower *GATA2* expression compared to controls supports this hypothesis. These results would match the previously reported phenotype in Gata2 haploinsufficient mice, which presented 20% of the normal Gata2 levels in total marrow, and 50% in the hematopoietic stem cell population, where Gata2 is predominantly expressed (Rodrigues et al., 2005). Interestingly, our results also point out that GATA2 germline mutations could affect other GATA2 target genes, explaining in part the clinical heterogeneity among patients with GATA2 deficiency (Spinner et al., 2014). In fact, our results show that the expression of several genes related to GATA2 is impaired. The finding that the expression of *SPI1* in our patient was about 17% of normal levels could be of special interest. Thirteen out of the 15 cases reported with the p.Arg396Gln mutation developed a myeloid malignancy, and mice carrying hypomorphic *Spi1* alleles that reduce *Spi1* expression to 20% of normal levels developed AML (Rosenbauer et al., 2004). Taken together, in addition to the insufficient *GATA2* expression, the putative dysregulation of specific GATA2 targets caused by mutated proteins could also contribute to the disease, and symptoms could differ from a pure haploinsufficiency phenotype. Along these lines, it has been reported that frameshift mutations favor an earlier onset age, compared to missense mutations (Dickinson et al., 2014). Therefore, further studies are needed to elucidate the pathogenesis of *GATA2* related syndromes.

In this study we have focused on the p.Arg396Gln mutation, for which no previous functional analyses had been reported. This missense mutation was reported previously in a total of 15 individuals, and there was a considerable variability in the overall phenotype. The mechanisms that could alter hematopoietic homeostasis and trigger the disease are unknown. Dickinson et al. suggested that disease evolution

may be consistent with both cell intrinsic (progressive deterioration of the HSPC compartment irrespective of environmental factors), and cell extrinsic (immunocompromising events such as infections) mechanisms (Dickinson et al., 2014). Our patient showed the first symptoms of the disease the same year she became pregnant and had febrile preeclampsia. Whether pregnancy is an immunocompromising condition is a matter of discussion (Griffiths et al., 2010). The case presented resembles a typical manifestation of the disease, with recurrent infections and altered leukocyte numbers beginning in early or mid-adulthood. To date, she has not displayed chronic neutropenia or lymphedema.

Our bioinformatic approach indicates that the substitution of the arginine for glutamine in the p.Arg396Gln mutation could impair binding to DNA. Several studies propose that cationic regions of proteins that contain arginine or lysine mediate the non-specific interaction between the protein and the negatively charged phosphate group of DNA backbone (Grant et al., 2013). Accordingly, Arg396 would stabilize the non-sequence-specific binding to DNA, mediating electrostatic interactions with the phosphate group. Although p.Arg396Gln would not necessarily alter the specificity for WGATAR sequences, it would probably affect the stability of GATA2 binding to DNA, which could lead to a loss of function. Moreover, it is possible that the interaction of Gln396 with neighboring residues (presumably Gln394 or Asn397 by our models) could also alter protein-protein interactions, as has been shown for p.Thr354Met and del355T (Hahn et al., 2011).

In vitro results in HL60 cells show that p.Arg396Gln is not able to function as WT-GATA2 does, which, in the context of these syndromes, could affect the maintenance of the progenitor compartment. It has been reported that the population of granulocyte/macrophage progenitor cells in *Gata2*$^{+/-}$ mice was diminished (Rodrigues et al., 2008), suggesting that appropriate WT-GATA2 expression levels maintain the characteristics of the progenitor-cell stage. A similar approach found that p.Thr354Met was able to inhibit differentiation, in the same way that WT-GATA2 does, while p.Thr355del was unable to do so (Hahn et al., 2011), indicating that each mutation exerts different activities. In this context, the activity of p.Arg396Gln would be similar to p.Thr355del.

On the other hand, previous reports have shown that ectopically overexpressed *GATA2* is able to induce hematopoietic cell quiescence, reducing colony output compared to cells with normal *GATA2* expression levels (Tipping et al., 2009). Thus, physiological *GATA2* expression levels allow a healthy balance between quiescent and cycling hematopoietic cell populations at stem and progenitor stages (Tipping et

al., 2009). Our analysis in murine HSPCs suggests that, while the results with WT-GATA2 are in accordance with the aforementioned report (which suggested that WT-GATA2 induces quiescence in stem and progenitor cells), p.Arg396Gln would be unable to stop immature cell division. Moreover, the increase in immature colony proportion with WT-GATA2 and not with p.Arg396Gln further suggests that, while WT-GATA2 retains the immature phenotype, p.Arg396Gln fails to do so. It has been suggested that hyperstimulation of the stem cell pool that produces hematopoietic progenitor cells with reduced expansion potential may lead to stem cell exhaustion in MonoMAC and DCML syndromes (Migliaccio et al., 2011). Our results would indicate that, in the context of these syndromes, p.Arg396Gln would force HSPCs to divide and differentiate prematurely, which would gradually drain the stem cell pool.

In summary, we show that human wild-type GATA2 contributes to its own transcription through a specific region located at -2.4 kb from IS TSS, and that p.Thr354Met, p.Thr355del and p.Arg396Gln fail to do so. In accordance with this finding, a patient with a p.Arg396Gln mutation and MonoMAC syndrome displayed lower *GATA2* expression than normal controls. Moreover, we show that p.Arg396Gln is a loss-of-function mutation, possibly triggered by a defective DNA binding, which as a consequence is unable to retain the progenitor phenotype in cells where it is expressed. This effect would cause HSPC pool damage, triggering cytopenias that would increase the risk of severe infections in patients with these diseases.

CONCLUSIONS

1. The GATA2 transcription factor directly binds and activates *MYB* expression in acute myeloid leukemia through two specific regions of the promoter: one distal, located at -7534 bp, and one proximal, located -669 and -612 bp. In fact, *MYB* expression is positively correlated with the expression of *GATA2* in three independent cohorts of patients with acute myeloid leukemia. Functional studies indicate that acute myeloid leukemia cells are highly dependent on *MYB* expression to sustain leukemogenic growth, suggesting that GATA2 contributes in part to acute myeloid leukemia transformation by maintaining high expression levels of *MYB*.

2. The human wild type GATA2 protein contributes to its own transcription through a specific region located at -2.4 kb from IS transcription start site, with a main activation role triggered by the -2409, and the -2418 *GATA2* binding sites.

3. *GATA2* missense or small in-frame deletion germline mutations in *GATA2* related syndromes code for mutated proteins that exert altered abilities as transcription factors, with a different mechanism depending on the *GATA2* mutation and the target gene. p.Thr354Met, p.Thr355del and p.Arg396Gln are unable to activate the *GATA2* promoter, causing a haploinsufficiency that may be responsible for *GATA2* deficit in individuals with those mutations.

4. *In silico* analyses indicate that p.Arg396Gln mutation is a disease-causing mutation, and may have an impaired ability to stabilize DNA binding by loss of electrostatic interaction with phosphate groups of DNA backbone.

5. *In vitro* functional approaches in the acute myeloid leukemia cell line HL60, and in murine hematopoietic stem and progenitor cells indicate that p.Arg396Gln is a loss-of-function mutation that is unable to retain the immature phenotype of hematopoietic progenitor cells. Moreover, the decrease in CFU-GEMM and increase in CFU-G obtained by the ectopic expression of p.Arg396Gln in murine hematopoietic stem and progenitor cells indicates that p.Arg396Gln might contribute to alter the leukocyte profile in affected individuals.

REFERENCES

Ahn EE, Higashi T, Yan M, Matsuura S, Hickey CJ, Lo MC, et al. SON protein regulates GATA-2 through transcriptional control of the microRNA 23a~27a~24-1 cluster. *J Biol Chem* 2013; **288**: 5381-5388.

Ayala RM, Martínez-López J, Albízua E, Diez A, Gilsanz F. Clinical significance of Gata-1, Gata-2, EKLF, and c-MPL expression in acute myeloid leukemia. *Am J Hematol* 2009; **84**: 79-86.

Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004; **116**: 281-297.

Bates DL, Chen Y, Kim G, Guo L, Chen L. Crystal structures of multiple GATA zinc fingers bound to DNA reveal new insights into DNA recognition and self-association by GATA. *J Mol Biol* 2008; **381**: 1292-1306.

Bigley V, Haniffa M, Doulatov S, Wang XN, Dickinson R, McGovern N, et al. The human syndrome of dendritic cell, monocyte, B and NK lymphoid deficiency. *J Exp Med* 2011; **208**: 227-234.

Bödör C, Renneville A, Smith M, Charazac A, Iqbal S, Etancelin P, et al. Germline GATA2 p.THR354MET mutation in familial myelodysplastic syndrome with acquired monosomy 7 and ASXL1 mutation demonstrating rapid onset and poor survival. *Haematologica* 2012; **97**: 890-894.

Bresnick EH, Johnson KD, Kim SI, Im H. Establishment and regulation of chromatin domains: mechanistic insights from studies of hemoglobin synthesis. *Prog Nucleic Acid Res Mol Biol* 2006; **81**: 435-471.

Bresnick EH, Katsumura KR, Lee HY, Johnson KD, Perkins AS. Master regulatory GATA transcription factors: mechanistic principles and emerging links to hematologic malignancies. *Nucleic Acids Res* 2012; **40**: 5819-5831.

Briegel K, Lim KC, Plank C, Beug H, Engel JD, Zenke M. Ectopic expression of a conditional GATA-2/estrogen receptor chimera arrests erythroid differentiation in a hormone-dependent manner. *Genes Dev* 1993; **7**: 1097-1109.

Camargo JF, Lobo SA, Hsu AP, Zerbe CS, Wormser GP, Holland SM: (2013) MonoMAC Syndrome in a Patient With a GATA2 Mutation: Case Report and Review of the Literature. *Clin Infect Dis* 2012; **57**: 697-699.

Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013; **368**: 2059-2074.

Cantor AB, Orkin SH. Transcriptional regulation of erythropoiesis: an affair involving multiple partners. *Oncogene* 2002; **21**: 3368-3376.

Chang AN, Cantor AB, Fujiwara Y, Lodish MB, Droho S, Crispino JD, et al. GATA-factor dependence of the multitype zinc-finger protein FOG-1 for its essential role in megakaryopoiesis. *Proc Natl Acad Sci USA* 2002; **99**: 9237-9242.

Chlon TM, Crispino JD. Combinatorial regulation of tissue specification by GATA and FOG factors. *Development* 2012; **139**: 3905-3916.

Chou ST, Khandros E, Bailey LC, Nichols KE, Vakoc CR, Yao Y, et al. Graded repression of PU.1/Sfpi1 gene transcription by GATA factors regulates hematopoietic cell fate. *Blood* 2009; **114**: 983-994.

Chun TH, Itoh H, Subramanian L, Iñiguez-Lluhí JA, Nakao K. Modification of GATA-2 transcriptional activity in endothelial cells by the SUMO E3 ligase PIASy. *Circ Res* 2003; **92**: 1201-1208.

Clappier E, Cuccuini W, Kalota A, Crinquette A, Cayuela JM, Dik WA, et al. The C-MYB locus is involved in chromosomal translocation and genomic duplications in human T-cell acute leukemia (T-ALL), the translocation defining a new T-ALL subtype in very young children. *Blood* 2007; **110**: 1251-1261.

Cuellar-Rodriguez J, Gea-Banaloche J, Freeman AF, Hsu AP, Zerbe CS, Calvo KR, et al. Successful allogeneic hematopoietic stem cell transplantation for GATA2 deficiency. *Blood* 2011; **118**: 3715-3720.

Dalgin G, Goldman DC, Donley N, Ahmed R, Eide CA, Christian JL. GATA-2 functions downstream of BMPs and CaM KIV in ectodermal cells during primitive hematopoiesis. *Dev Biol* 2007; **310**: 454-469.

Dassé E, Volpe G, Walton DS, Wilson N, Del Pozzo W, O'Neill LP, et al. Distinct regulation of c-myb gene expression by HoxA9, Meis1 and Pbx proteins in normal hematopoietic progenitors and transformed myeloid cells. *Blood Cancer J* 2012; **2**: e76.

De Pater E, Kaimakis P, Vink CS, Yokomizo T, Yamada-Inagawa T, van der Linden R, et al. Gata2 is required for HSC generation and survival. *J Exp Med* 2013; **210**: 2843-2850.

Deleuze V, El-Hajj R, Chalhoub E, Dohet C, Pinet V, Couttet P, et al. Angiopoietin-2 is a direct transcriptional target of TAL1, LYL1 and LMO2 in endothelial cells. *PLoS One* 2012; **7**: e40484.

Dickinson RE, Griffin H, Bigley V, Reynard LN, Hussain R, Haniffa M, et al. Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. *Blood* 2011; **118**: 2656-2658.

Dickinson RE, Milne P, Jardine L, Zandi S, Swierczek SI, McGovern N, et al. The evolution of cellular deficiency in GATA2 mutation. *Blood* 2014; **123**: 863-874.

Doré LC, Amigo JD, Dos Santos CO, Zhang Z, Gai X, Tobias JW, et al. A GATA-1-regulated microRNA locus essential for erythropoiesis. *Proc Natl Acad Sci USA* 2008; **105**: 3333-3338.

Doré LC, Crispino JD. Transcription factor networks in erythroid cell and megakaryocyte development. *Blood* 2011; **118**: 231-239.

Ezoe S, Matsumura I, Nakata S, Gale K, Ishikara K, Minegishi N, et al. GATA-2/estrogen receptor chimera regulates cytokine-dependent growth of hematopoietic cells through accumulation of p21(WAF1) and p27(Kip1) proteins. *Blood* 2002; **100**: 3512-3520.

Ezoe S, Matsumura I, Gale K, Satoh Y, Ishikawa J, Mizuki M, et al. GATA transcription factors inhibit cytokine-dependent growth and survival of a hematopoietic cell line through the inhibition of STAT3 activity. *J Biol Chem* 2005; **280**: 13163-13170.

Fadilah SA, Cheong SK, Roslan H, Rozie-Hanisa M, Yen GK. GATA-1 and GATA-2 gene expression is related to the severity of dysplasia in myelodysplastic syndrome. *Leukemia* 2002; **16**: 1563-1565.

Fasan A, Eder C, Haferlach C, Grossmann V, Kohlmann A, Dicker F, et al. GATA2 mutations are frequent in intermediate-risk karyotype AML with biallelic CEBPA mutations and are associated with favorable prognosis. *Leukemia* 2013; **27**: 482-485.

Fujimaki S, Harigae H, Sugawara T, Takasawa N, Sasaki T, Kaku M. Decreased expression of transcription factor GATA-2 in hematopoietic stem cells in patients with aplastic anemia. *Br J Haematol* 2001; **113**: 52-57.

Fujiwara T, O'Geen H, Keles S, Blahnik K, Linnemann AK, Kang YA, et al. Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. *Mol Cell* 2009; **36**: 667-681.

Furuhata A, Murakami M, Ito H, Gao S, Yoshida K, Sobue S, et al. GATA-1 and GATA-2 binding to 3' enhancer of WT1 gene is essential for its transcription in acute myeloid leukemia and solid tumor cell lines. *Leukemia* 2009; **23**: 1270-1277.

Gao X, Johnson KD, Chang YI, Boyer ME, Dewey CN, Zhang J, et al. Gata2 cis-element is required for hematopoietic stem cell generation in the mammalian embryo. *J Exp Med* 2013; **210**: 2833-2842.

Göttgens B, Nastos A, Kinston S, Plitz S, Delabesse EC, Stanley M, et al. Establishing the transcriptional programme for blood: the SCL stem cell enhancer is regulated by a multiprotein complex containing Ets and GATA factors. *EMBO J* 2002; **21**: 3039-3050.

Grant BN, Dourlain EM, Araneda JN, Throneberry ML, McFail-Isom LA. DNA phosphate crowding correlates with protein cationic side chain density and helical curvature in protein/DNA crystal structures. *Nucleic Acids Res* 2013; **41**: 7547–7555.

Grass JA, Boyer ME, Pal S, Wu J, Weiss MJ, Bresnick EH. GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodelling. *Proc Natl Acad Sci USA* 2003; **100**: 8811-8816.

Grass JA, Jung H, Kim SI, Martowicz ML, Pal S, Blobel GA, Bresnick EH. Distinct functions of dispersed GATA factor complexes at and endogenous gene locus. *Mol Cell Biol* 2006; **26**: 7056-7067.

Green CL, Tawana K, Hills RK, Bödör C, Fitzgibbon J, Inglott S, et al. GATA mutations in sporadic and familial acute myeloid leukaemia patients with CEBPA mutations. *Br J Haematol* 2013; **161**: 701-705.

Greif PA, Dufour A, Konstandin NP, Ksienzyk B, Zellmeier E, Tizazu B, et al. GATA2 zinc finger 1 mutations associated with biallelic CEBPA mutations define a unique genetic entity of acute myeloid leukemia. *Blood* 2012; **120**: 395-403.

Griffiths PD. Is pregnancy an immunocompromising condition? *Rev Med Virol* 2010; **20**: 341-343.

Grogan JL and Locksley RM. T helper cell differentiation: on again, off again. *Curr Opin Immunol* 2002; **14**: 366-372.

Grossmann V, Haferlach C, Nadarajah N, Fasan A, Weissmann S, Roller A, et al. CEBPA double-mutated acute myeloid leukaemia harbours concomitant molecular mutations in 76.8% of cases with TET2 and GATA2 alterations impacting prognosis. *Br J Haematol* 2013; **161**: 649-658.

Hahn CN, Chong CE, Carmichael CL, Wilkins EJ, Brautigan PJ, Li XC, et al. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat Genet* 2011; **43**: 1012-1017.

Hayakawa F, Towatari M, Ozawa Y, Tomita A, Privalsky ML, Saito H. Functional regulation of GATA-2 by acetylation. *J Leukoc Biol* 2004; **75**: 529-540.

Hess JL, Bittner CB, Zeising DT, Bach C, Fuchs U, Borkhardt A, et al. c-MYB is an essential downstream target for homeobox-mediated transformation of hematopoietic cells. *Blood* 2006; **108**: 297-304.

Holme H, Hossain U, Kirwan M, Walne A, Vulliamy T, Dokal I. Marked genetic heterogeneity in familial myelodysplasia/acute myeloid leukaemia. *Br J Haematol* 2012; **158**: 242-248.

Horwitz M, Sabath DE, Smithson WA, Radich J. A family inheriting different subtypes of acute myelogenous leukemia. *Am J Hematol* 1996; **52**: 295-304.

Hosoya T, Maillard I, Engel JD. From the cradle to the grave: activities of GATA-3 througout T-cell development and differentiation. *Immunol Rev* 2010; **238**: 110-125.

Hsu AP, Sampaio EP, Khan J, Calvo KR, Lemieux JE, Patel SY, et al. Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood* 2011; **118**: 2653-2655.

Hsu AP, Johnson KD, Falcone EL, Sanalkumar R, Sanchez L, Hickstein DD, et al. GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. *Blood* 2013; **121**: 3830-3837.

Huang Z, Dore LC, Li Z, Orkin SH, Feng G, Lin S, et al. GATA-2 reinforces megakaryocyte development in the absence of GATA-1. *Mol Cell Biol* 2009; **29**: 5168-5180.

Imagawa S, Tarumoto T, Suzuki N, Mukai HY, Hasegawa Y, Higuchi M, et al. L-arginine rescues decreased erythropoietin gene expression by stimulating GATA-2 with L-NMMA. *Kidney Int* 2002; **61**: 396-404.

Ishida H, Imai K, Honma K, Tamura S, Imamura T, Ito M, et al. GATA-2 anomaly and clinical phenotype of a sporadic case of lymphedema, dendritic cell, monocyte, B- and NK-cell (DCML) deficiency, and myelodysplasia. *Eur J Pediatr* 2012; **171**: 1273-1276.

Johnson KD, Hsu AP, Ryu MJ, Wang J, Gao X, Boyer ME, et al. Cis-element mutated in GATA2-dependent immunodeficiency governs hematopoiesis and vascular integrity. *J Clin Invest* 2012; **122**: 3692-3704.

Kaur J, Catovsky D, Valdimarsson H, Jensson O, Spiers AS. Familial acute myeloid leukaemia with acquired Pelger-Huet anomaly and aneuploidy of group C. *Br Med J* 1972; **4**: 327-331.

Kazenwadel J, Secker GA, Liu YJ, Rosenfeld JA, Wildin RS, Cuellar-Rodriguez J, et al. Loss-of-function germline GATA2 mutations in patients with MDS/AML or MonoMAC syndrome and primary lymphedema reveal a key role for GATA2 in the lymphatic vasculature. *Blood* 2012; **119**: 1283-1291.

Kitajima K, Masuhara M, Era T, Enver T Nakano T. GATA-2 and GATA-2/ER display opposing activities in the development and differentiation of blood progenitors. *EMBO J* 2002; **21**: 3060-3069.

Kitajima K, Tanaka M, Zheng J, Yen H, Sato A, Sugiyama D, et al. Redirecting differentiation of hematopoietic progenitors by a transcription factor, GATA-2. *Blood* 2006; **107**: 1857-1863.

Ko LJ, Engel JD. DNA-binding specificities of the GATA transcription factor family. *Mol Cell Biol* 1993; **13**: 4011-4022.

Koga S, Yamaguchi N, Abe T, Minegishi M, Tsuchiya S, Yamamoto M, et al. Cell-cycle-dependent oscillation of GATA2 expression in hematopoietic cells. *Blood* 2007; **109**: 4200-4208.

Krishna SS, Majumdar I, Grishin NV. Structural classification of zinc fingers: a survey and summary. *Nucleic Acids Res* 2003; **31**: 532-550.

Kuhl C, Atzberger A, Iborra F, Nieswandt B, Porcher C, Vyas P. GATA1-mediated megakaryocyte differentiation and growth control can be uncoupled and mapped to different domains in GATA1. *Mol Cell Biol* 2005; **25**: 8592-8606.

Lahortiga I, De Keersmaecker K, Van Vlierberghe P, Graux C, Cauweiler B, Lambert F, et al. Duplication of the MYB oncogene in T cell acute lymphoblastic leukemia. *Nat Genet* 2007; **39**: 593-595.

Landry JR, Bonadies N, Kinston S, Knezevic K, Wilson NK, Oram SH, et al. Expression of the leukemia oncogene Lmo2 is controlled by an array of tissue-specific elements dispersed over 100 kb and bound by Tal1/Lmo2, Ets, and Gata factors. *Blood* 2009; **113**: 5783-5792.

Lee ME, Temizer DH, Clifford JA, Quertermous T. Cloning of the GATA-binding protein that regulates endothelin-1 gene expression in endothelial cells. *J Biol Chem* 1991; **266**: 16188-16192.

Lim KC, Hosoya T, Brandt W, Ku CJ, Hosoya-Ohmura S, Camper SA, et al. Conditional Gata2 inactivation results in HSC loss and lymphatic mispatterning. *J Clin Invest* 2012; **122**: 3705-3717.

Ling KW, Ottersbach K, van Hamburg JP, Oziemlak A, Tsai FY, Orkin SH, et al. GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells. *J Exp Med* 2004; **200**: 871-882.

Love PE, Warzecha C, Li L. Ldb1 complexes. The new master regulators of erythroid gene transcription. *Trends Genet* 2014; **30**: 1-9.

Lorenzo PI, Brendeford EM, Gilfillan S, Gavrilov AA, Leedsak M, Razin SV, et al. Identification of c-Myb target genes in K562 cells reveals a role for c-Myb as a master regulator. *Genes Cancer* 2011; **2**: 805-817.

Lowry JA, Atchley WR. Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding domain. *J Mol Evol* 2000; **50**: 103-115.

Luesink M, Hollink IH, van der Velden VH, Knops RH, Boezeman JB, de Haas V, et al. High GATA2 expression is a poor prognostic marker in pediatric acute myeloid leukemia. *Blood* 2012; **120**: 2064-2075.

Lugus JJ, Chung YS, Mills JC, Kim SI, Grass J, Kyba M, et al. GATA2 functions at multiple steps in hemangioblast development and differentiation. *Development* 2007; **134**: 393-405.

Mace EM, Hsu AP, Monaco-Shawver L, Makedonas G, Rosen JB, Dropulic L, et al. Mutations in GATA2 cause human NK cell deficiency with specific loss of the $CD56^{bright}$ subset. *Blood* 2013; **121**: 2669-2677.

Maeno M, Mead PE, Kelley C, Xu RH, Kung HF, Suzuki A, et al. The role of BMP-4 and GATA-2 in the induction and differentiation of hematopoietic mesoderm in Xenopus laevis. *Blood* 1996; **88**: 1965-1972.

Maicas M, Vázquez I, Vicente C, García-Sánchez MA, Marcotegui N, Urquiza L, et al. Functional characterization of the promoter region of the human EVI1 gene in acute myeloid leukemia: RUNX1 and ELK1 directly regulate its transcription. *Oncogene* 2013; **32**: 2069-2078.

Manzotti G, Mariani SA, Corradini F, Bussolari R, Cesi V, Vergalli J, et al. Expression of p89(c-Mybex9b), an alternatively spliced form of c-Myb, is required for proliferation and survival of p210BCR/ABL-expressing cells. *Blood Cancer J* 2012; **2**: e71.

Martowicz ML, Grass JA, Boyer ME, Guend H, Bresnick EH. Dynamic GATA factor interplay at a multicomponent regulatory region of the GATA-2 locus. *J Biol Chem* 2005; **280**: 1724-1732.

Masuda A, Yoshikai Y, Kume H, Matsuguchi T. The interaction between GATA proteins and activator protein-1 promotes the transcription of IL-13 in mast cells. *J Immunol* 2004; **173**: 5564-5573.

Menghini R, Marchetti V, Cardellini M, Hribal ML, Mauriello A, Lauro D, et al. Phosphorylation of GATA2 by Akt increases adipose tissue differentiation and reduces adipose tissue-related inflammation: a novel pathway linking obesity to atherosclerosis. *Circulation* 2005; **111**: 1946-1953.

Migliaccio AR, Bieker JJ. GATA2 finds its macrophage niche. *Blood* 2011; **118**: 2647–2649.

Minami T, Murakami T, Horiuchi K, Miura M, Noguchi T, Miyazaki K, et al. Interaction between hex and GATA transcription factors in vascular endothelial cells inhibits flk-1/KDR-mediated vascular endothelial growth factor signaling. *J Biol Chem* 2004; **279**: 20626-20635.

Minegishi N, Ohta J, Suwabe N, Nakauchi H, Ishihara H, Hayashi N, et al. Alternative promoters regulate transcription of the mouse GATA-2 gene. *J Biol Chem* 1998; **273**: 3625-3634.

Minegishi N, Suzuki N, Yokomizo T, Pan X, Fujimoto T, Takahashi S, et al. Expression and domain-specific function of GATA-2 during differentiation of the hematopoietic precursor cells in midgestation mouse embryos. *Blood* 2003; **102**: 896-905.

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

Molkentin JD. The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression. *J Biol Chem* 2000; **275**: 38949-38952.

Morgenstern JP, Land H. Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. *Nucleic Acids Res* 1990; **18:** 3587-3596.

Morrisey EE, Ip HS, Lu MM, Parmacek MS. GATA-6: a zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. *Dev Biol* 1996; **177**: 309-322.

Muntean AG, Crispino JD. Differential requirements for the activation domain and FOG-interaction surface of GATA-1 in megakaryocyte gene expression and development. *Blood* 2005; **106**: 1223-1231.

Mutsaers PG, van de Loosdrecht AA, Tawana K, Bödör C, Fitzgibbon J, Menko FH. Highly variable clinical manifestations in a large family with a novel GATA2 mutation. *Leukemia* 2013; **27**: 2247-2248.

Niimi K, Kiyoi H, Ishikawa Y, Hayakawa F, Kurahashi S, Kihara R, et al. GATA2 zinc finger 2 mutation found in acute myeloid leukemia impairs myeloid differentiation. *Leuk Res Rep* 2013; **2**: 21-25.

Nottingham WT, Jarratt A, Burgess M, Speck CL, Cheng JF, Prabhakar S, et al. Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer. *Blood* 2007; **110**: 4188-4197.

Ohmori S, Takai J, Ishijima Y, Suzuki M, Moriguchi T, Philipsen S, et al. Regulation of GATA factor expression is distinct between erythroid and mast cell lineages. *Mol Cell Biol* 2012; **32**: 4742-4755.

Ohneda K, Yamamoto M. Roles of hematopoietic transcription factors GATA-1 and GATA-2 in the development of red blood cell lineage. *Acta Haematol* 2002; **108**: 237-245.

Orkin SH, Zon LI. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* 2008; **132**: 631-644.

O'Rourke JP, Ness SA. Alternative RNA splicing produces multiple forms of c-Myb with unique transcriptional activities. *Mol Cell Biol* 2008; **28**: 2091-2101.

Osada H, Grutz G, Axelson H, Forster A, Rabbitts TH. Association of erythroid transcription factors: complexes involving the LIM protein RBTN2 and the zinc-finger protein GATA1. *Proc Natl Acad Sci USA* 1995; **92**: 9585-9589.

Ostengaard P, Simpson MA, Connell FC, Steward CG, Brice G, Woollard WJ, et al. Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). *Nat Genet* 2011; **43**: 929-931.

Ozawa Y, Towatari M, Tsuzuki S, Hayakawa F, Maeda T, Miyata Y, et al. Histone acetylase 3 associates with and represses the transcription factor GATA-2. *Blood* 2001; **98**: 2116-2123.

Pal S, Cantor AB, Johnson KD, Moran TB, Boyer ME, Orkin SH, et al. Coregulator-dependent facilitation of chromatin occupancy by GATA-1. *Proc Natl Acad Sci USA* 2004; **101**: 980-985.

Pase L, Layton JE, Kloosterman WP, Carradice D, Waterhouse PM, Lieschke GJ. miR-451 regulates zebrafish erythroid maturation in vivo via its target gata2. *Blood* 2009; **113**: 1794-1804.

Pasquet M, Bellanné-Chantelot C, Tavitian S, Prade N, Beaupain B, Larochelle O, et al. High frequency of GATA2 mutations in patients with chronic neutropenia evolving to MonoMac syndrome, myelodysplasia, and acute myeloid leukemia. *Blood* 2013; **121**: 822-829.

Pattabiraman DR, Gonda TJ. Role and potential for therapeutic targeting of MYB in leukemia. *Leukemia* 2013; **27**: 269-277.

Patient RK, McGhee JD. The GATA family (vertebrates and invertebrates). *Curr Opin Genet Dev* 2002; **12**: 416-422.

Pedone PV, Omichinski JG, Nony P, Trainor C, Gronenborn AM, Clore GM, et al. The N-terminal fingers of chicken GATA-2 and GATA-3 are independent sequence-specific DNA binding domains. *EMBO J* 1997; **16**: 2874-2882.

Persons DA, Allay JA, Allay ER, Ashmun RA, Orlic D, Jane SM, et al. Enforced expression of the GATA-2 transcription factor blocks normal hematopoiesis. *Blood* 1999; **93**: 488-499.

Pevny L, Simon MC, Robertson E, Klein WH, Tsai SF, D'Agati V, et al. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* 1991; **349**: 257-260.

Pimanda JE, Ottersbach K, Knezevic K, Kinston S, Chan WY, Wilson NK, et al. Gata2, Fli1, and Scl form a recursively wired gene-regulatory circuit during early hematopoietic development. *Proc Natl Acad Sci USA* 2007; **104**: 17692-17697.

Ranganath S, Murphy KM. Structure and specificity of GATA protein in Th2 development. *Mol Cell Biol* 2001; **21**: 2716-2725.

Robert-Moreno A, Espinosa L, de la Pompa JL, Bigas A. RBPjkappa-dependent Notch function regulates GATA2 and is essential for the formation of intra-embryonic hematopoietic cells. *Development* 2005; **132**: 1117-1126.

Robinson CW, Normal JE, Cleland LG, Ford JH. Dermal necrosis and chromosome Iq abnormality in a man with familial myeloproliferative disorder. *Aust N Z J Med* 1983; **13**: 141-145.

Rodrigues NP, Janzen V, Forkert R, Dombkowski DM, Boyd AS, Orkin SH, et al. Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. *Blood* 2005; **106**: 477-484.

Rodrigues NP, Boyd AS, Fugazza C, May GE, Guo Y, Tipping AJ, et al. GATA-2 regulates granulocyte-macrophage progenitor cell function. *Blood* 2008; **112**: 4862-4873.

Rosenbauer F, Wagner K, Kutok JL, Iwasaki H, Le Beau MM, Okuno Y, et al. Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nat Genet* 2004; **36**: 624-630.

Rubel CA, Franco HL, Jeong JW, Lydon JP, DeMayo FJ. GATA2 is expressed at critical times in the mouse uterus during pregnancy. *Gene Expr Patterns* 2012; **12**: 196-203.

Sanda T, Lawton LN, Barrasa MI, Fan ZP, Kohlhammer H, Gutierrez A, et al. Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. *Cancer Cell* 2012; **22**: 209-221.

Shiba N, Funato M, Ohki K, Park MJ, Mizushima Y, Adachi S, et al. Mutations of the GATA2 and CEBPA genes in paediatric acute myeloid leukaemia. *Br J Haematol* 2013; [Epub ahead of print].

Shimamoto T, Ohyashiki K, Ohyashiki JH, Kawakubo K, Fujimura T, Iwama H, et al. The expression pattern of erythrocyte/megakaryocyte-related transcription factors GATA-1 and the stem cell leukemia gene correlates with hematopoietic differentiation and is associated with outcome of acute myeloid leukemia. *Blood* 1995; **86**: 3173-3180.

Shimizu R, Yamamoto M. Gene expression regulation and domain function of hematopoietic GATA factors. *Semin Cell Dev Biol* 2005; **16**: 129-136.

Simon MC, Pevny L, Wiles MV, Keller G, Costantini F, Orkin SH. Rescue of erythroid development in gene targeted GATA-1-mouse embryonic stem cells. *Nat Genet* 1992; **1**: 92-98.

Simon MP, Tournaire R, Pouyssegur J. The angiopoietin-2 gene of endothelial cells is up-regulated in hypoxia by HIF binding site located in its first intron and by the central factors GATA-2 and Ets-1. *J Cell Physiol* 2008; **217**: 809-818.

Smith ML, Cavenagh JD, Lister TA, Fitzgibbon J. Mutation of CEBPA in familial acute myeloid leukemia. *N Engl J Med* 2004; **351**: 2403-2407.

Snow JW, Trowbridge JJ, Fujiwara T, Emambokus NE, Grass JA, Orkin SH, et al. A single cis element maintains repression of the key developmental regulator Gata2. *PLoS Genet* 2010; **6**: e1001103.

Snow JW, Trowbridge JJ, Johnson KD, Fujiwara T, Emambokus NE, Grass JA, et al. Context-dependent function of "GATA switch" sites in vivo. *Blood* 2011; **117**: 4769-4772.

Soliera AR, Lidonnici MR, Ferrari-Amorotti G, Prisco M, Zhang Y, Martinez RV, et al. Transcriptional repression of c-Myb and GATA-2 is involved in the biologic effects of C/EBP in p210BCR/ABL-expressing cells. *Blood* 2008; **112**: 1942-1950.

Song WJ, Sullivan MG, Legare RD, Hutchings S, Tan X, Kufrin D, et al. Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat Genet* 1999; **23**: 166-175.

Spinner MA, Sanchez LA, Hsu AP, Shaw PA, Zerbe CS, Calvo KR, et al. GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics, and immunity. *Blood* 2014; **123**: 809-821.

Stadhouders R, Thongjuea S, Andrieu-Soler C, Palstra RJ, Bryne JC, van den Heuvel A, et al. Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development. *EMBO J* 2012; **31**: 986-999.

Takahashi S, Shimizu R, Suwabe N, Kuroha T, Yoh K, Ohta J, et al. GATA factor transgenes under GATA-1 locus control rescue germline GATA-1 mutant deficiencies. *Blood* 2000; **96**: 910-916.

Terui K, Takahashi Y, Kitazawa J, Toki T, Yokoyama M, Ito E. Expression of transcription factors during megakaryocytic differentiation of CD34+ cells from human cord blood induced by thrombopoietin. *Tohoku J Exp Med* 2000; **192**: 259-273.

Tijssen MR, Ghevaert C. Transcription factors in late megakaryopoiesis and related plateled disorders. *J Thromb Haemost* 2013; **11**: 593-604.

Tipping AJ, Pina C, Castor A, Hong D, Rodrigues NP, Lazzari L, et al. High GATA-2 expression inhibits human hematopoietic stem and progenitor cell function by effects on cell cycle. *Blood* 2009; **113**: 2661-2672.

Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A. Use of dipolar 1H-15N and 1H-13C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat Struc Biol* 1997; **4**: 732-738.

Tong Q, Tsai J, Tan G, Dalgin G, Hotamisligil GS. Interaction between GATA and the C/EBP family of transcription factors is critical in GATA-mediated suppression of adipocyte differentiation. *Mol Cell Biol* 2005; **25**: 706-715.

Towatari M, May GE, Marais R, Perkins GR, Marshall CJ, Cowley S, et al. Regulation of GATA-2 phosphorylation by mitogen-activated protein kinase and interleukin-3. *J Biol Chem* 1995; **270**: 4101-4107.

Trainor CD, Omichinski JG, Vandergon TL, Gronenborn AM, Clore GM, Felsenfeld G. A palindromic regulatory site within vertebrate GATA-1 promoters requires both zinc fingers of the GATA-1 DINA-binding domain for high-affinity interaction. *Mol Cell Biol* 1996; **16**: 2238-2247.

Trayner ID, Bustorff T, Etches AE, Mufti GJ, Foss Y, Farzaneh F. Changes in antigen expression on differentiating HL60 cells treated with dimethylsulphoxide, all-trans retinoic acid, alpha1,25-dihydroxyvitamin D3 or 12-O-tetradecanoyl phorbol-13-acetate. *Leuk Res* 1998; **22**: 537-547.

Tsai FY, Keller G, Kuo FC, Weiss M, Chen J, Rosenblatt M, et al. An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature* 1994; **371**: 221-226.

Tsai FY, Orkin SH. 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* 1997; **89**: 3636-3643.

Tsai FY, Browne CP, Orkin SH. Knock-in mutation of transcription factor GATA-3 into the GATA-1 locus: partial rescue of GATA-1 loss of function in erythroid cells. *Dev Biol* 1998; **196**: 218-227.

Tsuzuki S, Towatari M, Saito H, Enver T. Potentiation of GATA-2 activity through interactions with the promyelocytic leukemia protein (PML) and the t(15;17)-generated PML-retinoic acid receptor alpha oncoprotein. *Mol Cell Biol* 2000; **20**: 6276-6286.

Tsuzuki S, Enver T. Interactions of GATA-2 with the promyelocytic leukemia zinc finger (PLZF) protein, its homologue FAZF, and the t(11;17)-generated PLZF-retinoic acid receptor alpha oncoprotein. *Blood* 2002; **99**: 3404-3410.

Tsuzuki S, Kitajima K, Nakano T, Glasow A, Zelent A, Enver T. Cross talk between retinoic acid signaling and transcription factor GATA-2. *Mol Cell Biol* 2004; **24**: 6824-6836.

Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM, et al. Prognostically useful-gene expression profiles in acute myeloid leukemia. *N Engl J Med* 2004; **350**: 1617-1628.

Vicente C, Conchillo A, García-Sánchez MA, Odero MD. The role of the GATA2 transcription factor in normal and malignant hematopoiesis. *Crit Rev Oncol Hematol* 2012; **82**: 1-17.

Vicente C, Vazquez I, Conchillo A, García-Sánchez MA, Marcotegui N, Fuster O, et al. Overexpression of GATA2 predicts an adverse prognosis for patients with acute myeloid leukemia and it is associated with distinct molecular abnormalities. *Leukemia* 2012; **26**: 550-554.

Vinh DC, Patel SY, Uzel G, Anderson VL, Freeman AF, Olivier KN, et al. Autosomal dominant and sporadic monocytopenia with susceptibility to mycobacteria, fungi, papillomaviruses, and myelodysplasia. *Blood* 2010; **115**: 1519-1529.

Visvader J, Adams JM. Megakaryocytic differentiation induced in 416B myeloid cells by GATA-2 and GATA-3 transgenes or 5-azacytidine is tightly coupled to GATA-1 expression. *Blood* 1993; **82**: 1493-1501.

Visvader JE, Crossley M, Hill J, Orkin SH, Adams JM. The C-terminal zinc finger of GATA-1 or GATA-2 is sufficient to induce megakaryocytic differentiation of an early myeloid cell line. *Mol Cell Biol* 1995; **15**: 634-641.

Wang T, Nandakumar V, Jiang XX, Jones L, Yang AG, Huang XF, et al. The control of hematopoietic stem cell maintenance, self-renewal, and differentiation by Mysm1-mediated epigenetic regulation. *Blood* 2013; **122**: 2812-2822.

Watsuji T, Okamoto Y, Emi N, Katsuoka Y, Hagiwara M. Controlled gene expression with a reverse tetracycline-regulated retroviral vector (RTRV) system. *Biochem Biophys Res Commun* 1997; **234**: 769-773.

West RR, Hsu AP, Holland SM, Cuellar-Rodriguez J, Hickstein DD. Acquired ASXL1 mutations are common in patients with inherited GATA2 mutations and correlate with myeloid transformation. *Haematologica* 2014; **99**: 276-281.

Wickrema A, Crispino JD. Erythroid and megakaryocytic transformation. *Oncogene* 2007; **26**: 6803-6815.

Wilson NK, Foster SD, Wang X, Knezevic K, Schütte J, Kaimakis P, et al. Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of then major transcriptional regulators. *Cell Stem Cell* 2010; **7**: 532-544.

Wozniak RJ, Boyer ME, Grass JA, Lee Y, Bresnick EH. Context-dependent GATA factor function: combinatorial requirements for transcriptional control in hematopoietic and endothelial cells. *J Biol Chem* 2007; **282**: 14665-14674.

Wozniak RJ, Keles S, Lugus JJ, Young KH, Boyer ME, Tran TM, et al. Molecular hallmarks of endogenous chromatin complexes containing master regulators of hematopoiesis. *Mol Cell Biol* 2008; **28**: 6681-6694.

Xu Y, Takahashi Y, Wang Y, Hama A, Nishio N, Muramatsu H, et al. Downregulation of GATA-2 and overexpression of adipogenic gene-PPARgamma in mesenchymal stem cells from patients with aplastic anemia. *Exp Hematol* 2009; **37**: 1393-1399.

Yan XJ, Xu J, Gu ZH, Pan CM, Lu G, Shen Y, et al. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat Genet* 2011; **43**: 309-315.

Yang HY, Evans T. Distinct roles for the two cGATA-1 finger domains. *Mol Cell Biol* 1992; **12**: 4562-4570.

Yang Z, Gu L, Romeo PH, Bories D, Motohashi H, Yamamoto M, et al. Human GATA-3 trans-activation, DNA-binding, and nuclear localization activities are organized into distinct structural domains. *Mol Cell Biol* 1994; **14**: 2201-2212.

Yuasa H, Oike Y, Iwama A, Nishikata I, Sugiyama D, Perkins A, et al. Oncogenic transcription factor Evi1 regulates hematopoietic stem cell proliferation through GATA-2 expression. *EMBO J* 2005; **24**: 1976-1987.

Zhang Y, Payne KJ, Zhu Y, Price MA, Parrish YK, Zielinska E, et al. SCL expression at critical points in human hematopoietic lineage commitment. *Stem Cells* 2005; **23**: 852-860.

Zhang SJ, Ma LY, Huang QH, Li G, Gu BW, Gao XD, et al. Gain-of-function mutation of GATA-2 in acute myeloid transformation of chronic myeloid leukemia. *Proc Natl Acad Sci USA* 2008; **105**: 2076-2081.

Zhang SJ, Shi JY, Li JY. GATA-2 L359V mutation is exclusively associated with CML progression but not other hematological malignancies and GATA-2 P250A is a novel single nucleotide polymorphism. *Leuk Res* 2009; **33**: 1141-1143.

Zuber J, Rappaport AR, Luo W, Wang E, Chen C, Vaseva AV, et al. An integrated approach to dissecting oncogene addiction implicates a Myb-coordinated self-renewal program as essential for leukemia maintenance. *Genes Dev* 2011; **25**: 1628-1640.