

Supplemental text.

Additional clinical history

Immunodeficiency patient

The patient was born as the third child after an uneventful pregnancy to parents of northern European ancestry. During childhood she suffered from recurrent bronchitis. A routine screening for cervical cancer at 31 years-old revealed cervical intraepithelial neoplasia. Her subsequent course was complicated by recurrent HPV-associated cervical and vaginal intraepithelial neoplasia. At 40 years-old she was diagnosed with chronic hepatitis C virus (HCV) infection which cleared with therapy. She developed recurrent fevers of unknown source at 41 years-old. She has a history of pneumonia complicated by empyema at 42 years-old with subsequent recurrent episodes of pneumonia, including bilateral pneumonia. She also was diagnosed with cryptogenic organizing pneumonia. She developed an extensive deep vein thrombosis of the lower extremity at 45 years-old (vena cava, right iliac, femoral, popliteal, tibial veins). At age 48 she continued to suffer from recurrent fever, systemic inflammation (low-grade fever, anemia of chronic disease, c-reactive protein ~ 80mg/l, reference <5mg/l) and waxing and waning pulmonary symptoms with satisfactory clinical condition. The typical pattern of cytopenias for GATA2 deficiency (Figure 1B) was noted with monocytopenia, B-, NK-, dendritic cell and CD4+ T-cell cytopenia and GATA2 haploinsufficiency was diagnosed. Blood and bone marrow analysis showed anemia of chronic disease without evidence of MDS. Cytogenetic analysis of the bone marrow was performed and showed a normal karyotype. At 50 years-old she presented in markedly deteriorated general condition, fever resistant to antipyretic treatment and respiratory partial insufficiency due to pneumonia with *Serratia marcescens* and pulmonary alveolar proteinosis (Figure 1Aiii-iv). Her clinical condition worsened progressively. Follow-up cultures grew *Pseudomonas aeruginosa*, *Aspergillus fumigatus*, *Aspergillus nidulans* and *Mycobacterium intracellulare*. She died at age 50 years-old from overwhelming mycobacterial and fungal infections.

The course of the disease illustrates the difficulty to set the optimal timepoint for hematopoietic stem cell transplantation in adult GATA2 deficient patients without MDS.

Inherited AML/MDS family

A 22-year-old woman (III-1 on pedigree in Figure 1B) of northern European /Chickasaw Indian and Māori ancestry with no antecedent cytopenias presented with acute myeloid leukemia characterized by a 7q deletion detected by karyotype and FISH. Somatic mutation testing of her AML using an amplicon-based NGS platform identified a pathogenic *FLT3*

tyrosine kinase domain point mutation (p.Asp835Tyr; c.2503G>T, NM_004119.2). Of note, this initial somatic testing covered the region of synonymous mutation in *GATA2* in this family; however, this testing was initially reported as negative for a *GATA2* mutation. Re-review of the sequencing data revealed this pathogenic *GATA2* exon 3 synonymous mutation in this family (c.351C>G, NM_032638.4; p.Thr117=). Re-review of the data was prompted by the subsequent identification of this as a constitutional variant in her mother (II-1) by targeted capture and next generation sequencing of known inherited bone marrow failure and inherited AML/MDS predisposition genes using DNA isolated from cultured skin fibroblasts and functional characterization of its impact on RNA splicing. The 22-year-old was treated with one cycle of high dose-induction chemotherapy and achieved complete remission with incomplete count recovery and evidence of minimal residual disease by karyotype. She then underwent a matched unrelated donor PBSCT. Subsequent sequencing of tumor (AML) and constitutional (cultured skin fibroblasts) DNA from the proband confirmed the familial *GATA2* synonymous mutation. Her past medical history was otherwise unremarkable apart from an episode of respiratory syncytial viral infection requiring hospitalization as a child.

Her family history was significant for multiple family members with AML or MDS. Her mother (II-1) was diagnosed with MDS characterized by monosomy 7 and trisomy 8 at age 40 years and was treated with a mismatched unrelated donor double umbilical cord blood transplant. Notably, she additionally had a history of ptosis corrected surgically as a child. A maternal aunt (II-2) was diagnosed with AML at age 18 years who passed away of infectious complications seven years after sibling PBSCT (donor II-1). One of her half-brothers (III-2) at 24 years-old had normal peripheral blood counts and a history of ptosis corrected surgically and her other half-brother (III-3) at 17 years-old has mild thrombocytopenia and a hypocellular marrow with normal karyotype.

The genetic diagnosis in this family highlights several complexities of germline testing in hematologic malignancy. These include the need to 1) reconsider and re-classify variants as additional data emerge in the field, and 2) recognize and confirm germline (vs. somatic) variants identified by somatic mutation testing of tumor samples.

Additional genetic evaluation

We identified a SNP at chr3:128200534 that is present as a homozygous change in an individual sequenced from the inherited AML/MDS family (III-1) which is not present in the immunodeficiency patient. As the reported mutation in *GATA2* p.Thr117=, is at chr3:128205090, it is very unlikely that the reported mutation (*GATA2* p.Thr117=) occurred on the same haplotype in both families. We therefore conclude that these are likely

independent occurrences of the same mutation in the two separate families (immunodeficiency patient and the inherited AML/MDS family).

Table S1. In silico analyses of *GATA2* (exon 3) c.351C>G;NM_032638.4 donor and acceptor splice sites

Donor Splice site	position	type	NN Splice	HSF
tggtcagGTaaaggc	1070 1084	canonical donor	0.94	89.68
ctggacgGTgagccc	550 564	activated cryptic donor	0.61	88.29
Acceptor Splice site	position	type	NN Splice	HSF
cctctctctcttggccgcAGcccgcctgaccggaggccag	414 454	canonical acceptor	0.94	89.68
tggcctccctcaccctacAGcagcccactctggctcccac	672 712	alternative acceptor	0.61	88.29

Predictions were made using the Human Splicing Finder (HSF) and Neural Network algorithm (NN splice). The sequences of the canonical donor/acceptor splices predicted by the algorithms were verified by comparison to *GATA2* gene (ENSG00000179348) and cDNA (ENST00000341105.6) sequence information derived from the Ensembl database¹ (www.ensembl.org). The NN Splice algorithm² was run with default thresholds of 0.4 for both donor and acceptor sites with a maximum score of 1. The predicted splice sites are considered significant if the score is > 0.4. Scores closer to 1 indicate a stronger prediction of a splice site. For the HSF algorithm³, the consensus values go from 0 to 100 and a default threshold of significance is defined at 65 with a score >65 considered a likely donor or acceptor splice site.

¹Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Girón CG, Gil L, Gordon L, Haggerty L, Haskell E, Hourlier T, Izuogu OG, Janacek SH, Juettemann T, To JK, Laird MR, Lavidas I, Liu Z, Loveland JE, Maurel T, McLaren W, Moore B, Mudge J, Murphy DN, Newman V, Nuhn M, Ogeh D, Ong CK, Parker A, Patricio M, Riat HS, Schuilenburg H, Sheppard D, Sparrow H, Taylor K, Thormann A, Vullo A, Walts B, Zadissa A, Frankish A, Hunt SE, Kostadima M, Langridge N, Martin FJ, Muffato M, Perry E, Ruffier M, Staines DM, Trevanion SJ, Aken BL, Cunningham F, Yates A, Flicek P. Ensembl 2018. *Nucleic Acids Res.* 2018; 46(D1):D754-D761.

²Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol.* 1997 Fall;4(3):311-23.

³Desmet FO, Hamroun D, Lalande M, Collod-Bérout G, Claustres M, Bérout C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* 2009 May;37(9):e67.

Supplemental Methods

Sanger sequencing of GATA2 genomic DNA. Sanger sequencing of *GATA2* coding exons 2-6 including exon-intron boundaries from genomic DNA was performed after PCR amplification using previously published primers and conditions¹

RT-PCR and cloning of cDNA. Fibroblast RNA was extracted with High Pure RNA Isolation Kit (Roche Diagnostics) and transcribed with GoScript Reverse Transcription System (Promega) using oligo dT primers. RT-PCR was performed (Table S2). PCR products were cloned (TOPO™ TA Cloning™ Kit, Fisher) and transformed into One Shot™ TOP10 Chemically Competent *E. coli* (Invitrogen). Plasmid DNA isolation of single colonies was performed with Quick lyse Miniprep Kit (Qiagen). Subclones were Sanger sequenced after PCR amplification. In the inherited AML/MDS family, activation of the *GATA2* exon 3 cryptic donor splice was confirmed on RNA from the proband's AML peripheral blood mononuclear cells (PBMCs). PBMC RNA was extracted using TRIzol reagent (Invitrogen) and reverse-transcribed with SuperScript II Reverse Transcriptase (Invitrogen). RT-PCR was performed using primers spanning the exonic sequence flanking *GATA2* c.351C>G (Table S2). RT-PCR amplicons were subcloned into *E.coli* competent cells using TOPO TA Cloning kit (Invitrogen) and sequenced after colony PCR using Applied Biosystems Big Dye Terminator version 3.1 Cycle Sequencing Kit and ABI 3130xl genetic analyzer (Applied Biosystems). Sequence traces were analyzed using Sequencher version 5.4 software (Gene Codes Corporation).

¹Bodor C, Renneville A, Smith M, et al. Germ-line *GATA2* p.THR354MET mutation in familial myelodysplastic syndrome with acquired monosomy 7 and *ASXL1* mutation demonstrating rapid onset and poor survival. *Haematologica*. 2012;97(6):890-894.

Table S2: Primer sequences used for RT-PCR.

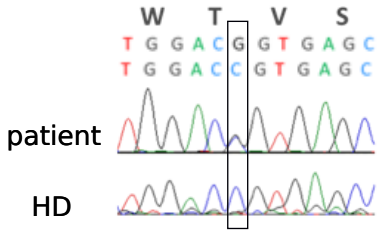
GATA2 RT-PCR	Forward primer	Reverse primer	Annealing temperature	Product size
RT-PCR ID Patient	5' GAATGCGCAGCAC CCCGACTCACA 3'	5' CTCCTTCTGGCGGC CGACAGTCTTC 3'	66°C	983bp
RT-PCR AML sample	5' CCCACACTTGTTGC ACAG 3'	5' TCAGTGACACCTGG TACTTGAC 3'	57°C	395bp
RT-PCR for cloning into TOPO vector	5' GAATGCGCAGCAC CCCGACTCACA 3'	5' TGGCCGAAGGAGA GGCTGGAGGAG 3'	66°C	1356bp
RT-PCR specific for aberrant splice product	5' CACCACCACAACC CCTGGACGCAG 3'	5' TGGCCGAAGGAGA GGCTGGAGGAG 3'	66°C	940bp

Figure S1:

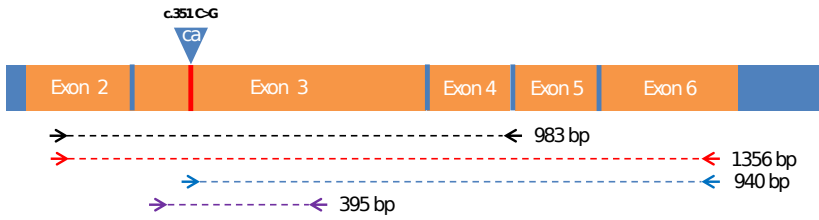
(A) Chromatogram of GATA2 genomic DNA exon 3 of immunodeficiency patient and healthy donor (HD). (B) Primer positions used for GATA2 mRNA RT-PCR studies. Black arrows and dashed line: RT-PCR immunodeficiency (ID) patient. Red arrows and dashed line: RT-PCR for cloning into TOPO vector of ID patient; Blue arrows and dashed line: RT-PCR specific for aberrant splice product. Violet arrows and dashed line: RT-PCR AML sample. (C) Relative percentages of three GATA2 mRNA transcripts expressed in cultured skin fibroblasts from the immunodeficiency patient and AML sample from patient III-1. Cloning of the RT-PCR products of the fibroblast sample from the immunodeficiency patient into *E.coli* and subsequent sequencing generated 70% wild-type GATA2 clones (symbolized in green), 13% full length mutant clones expressing a GATA2 transcript containing the c.351C>G mutation (red), and 17% expressing an aberrantly spliced GATA2 transcript with the 136bp deletion (orange). Cloning of the RT-PCR products of the AML sample generated 28% wild-type clones, 38% full length mutant clones, and 34% internally deleted mutant clones. Notably, the patient acquired a somatic GATA2 frameshift mutation in her AML affecting her second GATA2 allele (GATA2 NM_032638.4:c.72dupA, p.His25fs, VAF 0.53). This frameshift mutation likely induces nonsense mediated decay which would decrease the detection of wild-type GATA2 transcripts.

Supplementary Figure 1

A

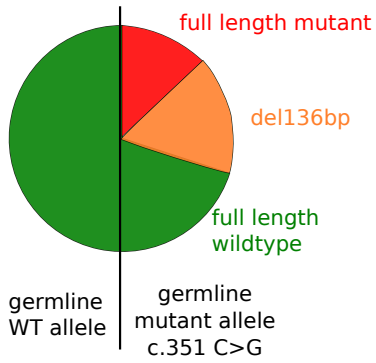


B



C

fibroblast, ID patient



AML, patient III-1

