haematologica
Journal of the Ferrata Storti Foundation

GATA2 at 14: genotype-phenotype correlations

by Amy P. Hsu, Subrata Paul, Jennifer L. Kwan, Philip L.F. Johnson, Justin Lack, Eva P. Szymanski, Jana Lovell, Ladan Foruraghi, Cindy Palmer, Christa S. Zerbe, Janine R. Daub, Joie Davis, Dennis D. Hickstein, Katherine R. Calvo and Steven M. Holland

Received: November 19, 2024.
Accepted: December 29, 2025.

Citation: Amy P. Hsu, Subrata Paul, Jennifer L. Kwan, Philip L.F. Johnson, Justin Lack, Eva P. Szymanski, Jana Lovell, Ladan Foruraghi, Cindy Palmer, Christa S. Zerbe, Janine R. Daub, Joie Davis, Dennis D. Hickstein, Katherine R. Calvo and Steven M. Holland. GATA2 at 14: genotype-phenotype correlations. Haematologica. 2026 Jan 15. doi: 10.3324/haematol.2024.287013 [Epub ahead of print]

Publisher's Disclaimer.
E-publishing ahead of print is increasingly important for the rapid dissemination of science. Haematologica is, therefore, E-publishing PDF files of an early version of manuscripts that have completed a regular peer review and have been accepted for publication.
E-publishing of this PDF file has been approved by the authors.
After having E-published Ahead of Print, manuscripts will then undergo technical and English editing, typesetting, proof correction and be presented for the authors' final approval; the final version of the manuscript will then appear in a regular issue of the journal.
All legal disclaimers that apply to the journal also pertain to this production process.

GATA2 at 14: genotype-phenotype correlations

Amy P. Hsu[1@], Subrata Paul[2], Jennifer L. Kwan[3], Philip L.F. Johnson[4], Justin Lack[2], Eva P. Szymanski[1*], Jana Lovell[1&], Ladan Foruraghi[1], Cindy Palmer[1], Christa S. Zerbe[1], Janine R. Daub[1], Joie Davis[1], Dennis D. Hickstein[5], Katherine R. Calvo[6], Steven M. Holland[1]

[1]Laboratory of Clinical Immunology and Microbiology, NIAID/NIH, Bethesda, MD, USA

[2] Integrated Data Sciences Section, Research Technologies Branch, NIAID/NIH, Bethesda, MD, USA

[3]Research Technologies Branch, NIAID/NIH, Bethesda, MD, USA

[4]Department of Biology, University of Maryland – College Park, College Park, MD, USA

[5]Immune Deficiency Cellular Therapy Program, NCI/NIH, Bethesda, MD, USA

[6]Department of Laboratory Medicine, Clinical Center, NIH, Bethesda, MD, USA

[@]Correspondence:

Amy P. Hsu, PhD

twins@niaid.nih.gov

Bldg 10 Rm 11S267, 10 Center Dr, Bethesda, MD 20892-1960

Phone: 301-761-7104

Fax: 301-480-4508

[*]Current affiliation Internal Medicine, University of Pennsylvania

[&]Current affiliation Heart and Vascular Institute, The Johns Hopkins School of Medicine, Baltimore, MD.

Authorship Contribution: A.P.H., E.S. and J.L. collected clinical data; A.P.H., S.P., and J.L.K analyzed data and made figures; A.P.H., S.P., J.L.K, J.L., and P.L.F.J. performed statistical analysis; L.F., C.P., C.S.Z., J.R.D., J.D., K.R.C, D.D.H. and S.M.H. provided clinical care and patient samples. K.R.C reviewed bone marrow biopsies, A.P.H. wrote the manuscript. A.P.H., K.R.C., and S.M.H. revised the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Acknowledgements: This research was supported in part by the Intramural Research Program of the NIH, National Institute of Allergy and Infectious Diseases, National Cancer Institute, and Clinical Center.

Abstract word count 240

Text word count 4714

Figures 6

Tables 2

References 45

Running Title: GATA2 Deficiency: Genotype-Phenotype Correlations

Data Sharing – Data is included as Supplemental Table 1; Code for analysis is available at https://github.com/spaul-genetics/GATA2_Report.

ClinicalTrials.gov identifier: NCT01905826

Funding: This research was funded in whole or in part by the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Cancer Institute, and Clinical Center, National Institutes of Health.

Abstract

GATA2 mutations cause adult-onset bone-marrow failure characterized by cytopenias, infections and increased risk of myeloid malignancy. We reviewed hospital records and referrals of 232 *GATA2* mutated individuals from 122 families to gather hematopoietic and syndromic features.

Mutations were categorized by GATA2 protein effect: mutation after the 2nd Zinc-finger (C-term, n=10); missense in the 2nd Zinc-Finger (ZF2, n=104); mutations producing stable truncated protein (Truncation, n=22); Null alleles (mRNA instability, large deletions, n=46); or Enhancer (n=50).

Regression models for symptom onset identified earlier onset and increased hazard ratios (HR) between Truncation, Null or ZF2 mutations (13 years, HR 5.00; 17 years, 3.60; 22 years, 2.23 respectively) and Enhancer. Commonly mutated ZF2 amino acids stratified: R396 or T354 presented earlier with increased hazard ratios (16 years, HR 2.96; 19 years, HR 2.16) versus R398 (34 years). Mutation groups with median onset <20 years had higher prevalence of cytopenias, infections, HPV, thromboses and hearing loss. Warts/anogenital HPV were common (93/232, 40.1%). Non-hematologic manifestations included hearing loss (45/232; 19.4%), lymphedema (30/232; 12.8%), and non-hematologic malignancies (46/232, 19.8%), the majority being HPV-related ano-genital cancers. Leukemia or lymphoma occurred in 25/232 patients (10.8%). Functional analysis of patient mutations showed impaired transactivation. Males had more monosomy 7 (17/73 males, 6/110 females, P=0.0016). Nontuberculous mycobacterial infection was the presenting symptom in 14/73 males vs 7/110 females (P=0.0151). GATA2 mutation type, and specific mutations, affect penetrance and expression. Sex is an independent risk for specific infectious presentations and cytogenetic abnormalities.

Introduction

GATA2 deficiency was described phenotypically in 2010 as a constellation of symptoms including infections, mono-, B, and NK cell cytopenias, and pulmonary manifestations with sporadic or dominant inheritance[1]. The following year the hematopoietic stem-cell transcription factor, *GATA2,* was identified as the causative gene[2] with phenotype extended to familial leukemia[3], loss of dendritic cells[4], lymphedema[5] and myelodysplastic syndrome (MDS)[6]. Non-hematopoietic findings included deafness and thrombotic events [7-11].

In the 14 years since GATA2 deficiency was identified, phenotypic characterization has expanded with *GATA2* mutations identified in registries of neutropenia[7] and pediatric[10] and adult[11] MDS. While early studies focused on unifying phenotypes, which defined the typical GATA2 patient seen in the clinic, more recent studies have catalogued somatic mutations affecting bone marrow hypoplasia, MDS and leukemia. A survey of 400 published *GATA2* mutations did not correlate clinical presentation with underlying mutation[12]. To address issues of penetrance and expression we took advantage of our large, longitudinal single-center cohort.

Methods

Patients referred to the National Institutes of Health (NIH) for genetic diagnosis or clinical management provided written informed consent prior to enrolling in IRB approved protocols. Chart review was performed for all records at NIH and from referring clinicians through June,

2020. Detailed descriptions of experimental methods used are provided in the Supplementary Appendix.

Results

Patient cohort

Our cohort consisted of 124 probands from 122 families, including two pairs of identical twins, referred to the NIH for genetic diagnostics or clinical management. All patients were confirmed to carry *GATA2* mutations; one patient had a full gene deletion, one had uniallelic *GATA2* expression (patient 23.I.3 previously reported[9,13]). The majority of mutations overlap with the 480 literature mutations recently reported with associated ACMG variant criteria[12]. At-risk relatives of mutation positive individuals, including all first-degree relatives or surviving descendants, were offered genetic testing. This led to the identification of 108 additional mutation-positive individuals totaling 232 individuals. Chart review and clinical history were used to identify initial symptoms, age at presentation, and additional clinical findings. There was a broad age range for initial symptom presentation (<1 year to 85 years); similarly, the initial presenting symptoms were varied, spanning hematopoietic, infections, and non-hematopoietic findings (Supplemental Figure 1). Of the 232 individuals, 104 are untransplanted, 73 are post-transplant, and 55 patients are deceased, 18 of whom underwent hematopoietic stem cell transplant (Supplemental Tables 1, 2; Supplemental Figure 2).

Classification of GATA2 mutations

GATA2 mutations are frequently private to a family, although several recurrent mutations occur at CpG dinucleotides, which undergo C>T transitions after demethylation. As a result, these mutations may arise sporadically and are not part of a larger population haplotype. Supporting the lack of a *GATA2* population haplotype, among 52 sets of parents of probands available for genetic testing, 21 probands had mutation-negative parents (21/52, 40.4%) including 4 with R396 and one with R398 mutations.

We classified the mutations' predicted effect on GATA2 protein (Figure 1A) and grouped patients accordingly. Mutations in the intronic enhancer element (n = 50) reduce transcription of the mutant allele by ~60%. However, these transcripts encode wild-type (WT) protein[13], yielding a decreased total amount of normal GATA2 protein. Since expression occurs from both alleles, the total GATA2 level should still be ~70% of wild-type. Null mutations (n = 46) included full-gene or large, partial-gene deletions and premature stop codons. Splicing or frameshift mutations causing nonsense-mediated decay (NMD) of the transcript result in loss of expression from the mutant allele; therefore, they are also classified as Null. These mutations have transcript and protein only from the intact WT allele, causing haploinsufficiency[13]. One family in our cohort carried a synonymous mutation at the last base of exon 5, c.1017G>T, p.L339L, which altered proper splicing, similar to other reported patients[14]. Transcript analysis demonstrated utilization of a cryptic intronic splice site causing a 16-base insertion. Since the mutant transcript had significantly lower expression than the wild-type allele (Supplemental Figure 3), the transcript was determined to be unstable, placing this family in the Null category. The remainder of the mutations are described with respect to the critical second zinc finger

(ZF2), which binds DNA during transcriptional activation of target genes. Truncations (n = 22) include stop codons, splice mutations or frameshifts occurring late enough in the transcript to prevent NMD, having stable mRNA and the potential for a protein lacking some or all of the ZF2. Truncation could also remove the nuclear localization signal (NLS) which occurs after the ZF2. Since there is evidence for chromatin binding even in the absence of the NLS[15], we treat Null and Truncation mutations as separate categories for this analysis. Two cases (c.1018-1G>A; c.1018- 50_1143+247del) in which cDNA analysis demonstrated in-frame exon skipping deleting most of the ZF2 are classified as Truncation due to the loss of the majority of ZF2. Missense changes or small, in-frame deletions wholly contained within the ZF2, are grouped together as ZF2 (n = 104). Lastly, C-terminal mutations (C-term, n = 10) include those missense or frameshift mutations occurring after the ZF2 region. The mechanism by which these C-term mutations alter GATA2 function is unresolved, however there are two families with multiple individuals carrying C-term mutations and phenotypes consistent with GATA2 deficiency. These last three mutation categories (Truncation, ZF2 and C-term) produce stable mRNA and would be predicted to have roughly half WT and half mutant GATA2 protein. Unique mutations from the cohort and the associated functional classifications are listed in Table 1 and shown in Figure 1B.

Aggregating patients by mutation category allowed group comparisons. We observed a striking difference between penetrance of Null and Truncation patients (95.5% each) compared to the ZF2 and enhancer patients (81.7% and 54%, respectively) (Table 2). Age variation between symptomatic and asymptomatic individuals did not account for presence of symptoms; the asymptomatic group had a higher median age (44.6 years) than the symptomatic group (21.0

years) providing sufficient time for phenotypes to develop (Supplemental Figure 4A). Comparing across mutation groups, the median age of asymptomatic individuals is the same or higher than the symptomatic patients (Supplemental Figure 4B). The C-term mutations showed an 80% penetrance, with the majority of these individuals coming from a single family[16,17]. Additional genetic factors within this family cannot be ruled out as a cause for the altered penetrance observed.

Next, we examined differences in age at first reported symptom by mutation category. For this analysis, we considered clinical presentations sufficiently abnormal to be treated or noted in a chart and excluded minor signs found during annual surveillance. Symptoms included both immune/hematologic (cytopenias, infections, abnormal bone marrows, hematologic malignancies) as well as non-hematologic manifestations (solid tumors, lymphedema, hearing issues or Erythema nodosum). Kaplan-Meier estimates for symptom onset showed clear differences across mutation categories, consistent with the observed differences in penetrance (Figure 2A, Supplemental Figure 5). The most penetrant mutation categories also had the lowest median ages at symptom onset: Truncation (13 years) and Null (17 years), ZF2 (22 years), C-term (31 years), and Enhancer mutations (45 years, with 35% of Enhancer mutation-positive individuals remaining asymptomatic into their 7th decades). Using a mixed-model Cox regression analysis, which accounted for relatedness among patients, we compared Enhancer mutations to the other mutation categories. Significant hazard ratios (HR) were found for Truncation (5.63, corrected $P<0.001$), Null (3.55, corrected $P<0.001$) and ZF2 mutations (2.24, corrected $P=0.006$) (Figure 2B). Pairwise comparisons across all mutation categories showed

significantly increased HR between Truncation and ZF2 (2.52, P=0.0487), while none of the remaining categories reached significance (Supplemental Table 2). No significant difference was observed between Null and Truncation mutations.

Given the large number of ZF2 mutation positive individuals, we identified four recurrently mutated sites, T354, R361, R396, and R398. There were 17 individuals with mutations at T354 (T354M n=15, 8 probands; T354K n=2, 1 proband); 14 with R361 mutations (R361C n=5, 3 probands; R361H n=5, 4 probands; R361P n=3, 1 proband; R361delRNAN n=1); 18 with R396 mutations (R396Q n=11, 6 probands; R396W n=6, 5 probands; R396G n=1 proband); and 33 with R398 mutations (R398W n=31, 12 probands; 1 proband each with R398L and R398Q) (Supplemental Figure 6). Kaplan-Meier estimates for age at symptom onset grouped by affected amino acid (Figure 1C, Supplemental Figure 7) showed striking differences among the groups: R396 and T354 presented in the second decade (16 and 19 years, respectively) reminiscent of the Truncation and Null mutation groups above. In contrast, the R361 and R398 groups had later median onsets (22 and 34 years, respectively). Applying a mixed-effect Cox regression analysis comparing R396, T354 and R361 to R398 demonstrated significantly increased HR for R396 and T354 (2.67, corrected P=0.004; 2.16, corrected P=0.028 respectively). R361 was not significantly different from R398 (HR 1.63, P=0.186) (Figure 2B).

Penetrance across recurrently mutated amino acids also mirrored symptom onset: 100% of individuals carrying R396 mutations had some disease manifestations while only 24/33 (72.7%) of R398 mutation carriers were symptomatic (Table 2). In fact, across all mutation groups there

was a strong, negative correlation between median age at first symptom onset and penetrance (R^2=0.8908, P=0.0001) (Figure 2D).

Functional assessment of GATA2 mutations

To assess the transactivation potential of GATA2 we transfected a luciferase GATA2-reporter construct[13] with decreasing amounts of GATA2-WT (Figure 3A) demonstrating a clear dose-dependent response to GATA2 levels. Reduction of GATA2-WT by 25%, similar to Enhancer mutations, produced significantly less luciferase than did WT (Figure 3A). Similarly, transfecting only 50%, akin to haploinsufficiency, decreased luciferase production (Figure 3A). Several patient variants were analyzed in the same system (Figure 3B). While the somatic, activating mutation, L359V, induced higher luciferase, all of the ZF2 mutations showed decreased transactivation compared to WT. The C-term mutation, S447R was not significantly different from wild-type (Figure 3B). Three rare variants selected from gnomAD, had normal (T347N, A364T) or slightly increased (G392R) activity.

Infectious phenotypes

The most common presenting symptom or sign within our cohort was infection (81/183 symptomatic individuals, 44.3%); cytopenias were the next most frequent (47/183, 25.7%) (Supplemental Table 3). However, over the course of disease, infections were observed in the majority (145/183 symptomatic individuals, 79.2%) while cytopenias were even more prevalent (161/183, 88.0%). As patients lost NK cells, viral infections became more pronounced[18]. Within our cohort there were 114 symptomatic individuals with viral infections (114/183, 62.3%).

Genital or extragenital HPV infections, including cutaneous warts, were seen in 94 patients; 22 patients had EBV, including three with chronic EBV; 19 had HSV, while 9 had CMV. Additional viral infections cited in records included H1N1 (n=2), VZV (n=8), measles (n=2) and one each of mumps, RSV, influenza (other than H1N1) and BK virus. Bacterial infections were seen in 74 patients (74/183, 40.4%), predominantly nontuberculous mycobacteria (NTM, 53/183, 29.0%), as well as staphylococcal and streptococcal infections in 5 patients each (2.7%). Fungal infections were seen in 44 patients (24.0%), most commonly *Candida* (n=18) and *Aspergillus* (n=12). Histoplasmosis (n=6), blastomycosis (n=2) and tinea (n=3) were also seen. Of the 145 patients with infections, 136 had documented cytopenias (93.8%) with the remaining 9 individuals being relatives with no lymphocyte phenotyping available. These findings are consistent with previous reports of co-occurrence of cytopenias and major complications of GATA2 deficiency[9].

Hematologic phenotypes

In 2016 GATA2 deficiency was recognized as a separate entity within myeloid neoplasms by the World Health Organization[19]. In view of the high rate of myeloid dysfunction and leukemia, we routinely conduct bone marrow surveillance on mutation positive individuals. Of 167 patients with bone marrow analysis, some abnormality was reported in 142 (142/167, 85.0%) including hypocellularity, atypical megakaryocytes, absent hematogones or cytopenias. Myelodysplastic syndrome (MDS) was documented in 102 patients (102/167, 61.1%), 116 of whom had bone marrow samples available for in-house hematopathologist review. MDS-MLD was the predominant classification (n = 43) followed by MDS-EB1 and MDS-SLD (4 each) and MDS-MPN

(n = 3), MDS-U, RCC, and CMML (2 each). Within the cohort, 26 patients had documented leukemia or lymphoma, some first identified during asymptomatic periods.

Non-hematopoietic phenotypes of GATA2 deficiency

Lymphedema with hearing loss and infections, first described as Emberger Syndrome, was recognized to be caused by *GATA2* mutations in 2011[5]. PROX1 is a critical transcription factor for lymphatic development with *Prox1* haploinsufficiency causing disruption of lymphovenous valves[20]. While it was suggested that *GATA2* null mutations were more common in lymphatic dysfunction and lymphedema[9,21], the Emberger cohort missense mutations (R361L, C373R, R396Q) display reduced binding to the *PROX1* -11kb element and significantly reduced DNA-binding and transactivation of the *PROX1* +4.5 regulatory region[22]. Twenty-nine individuals in our cohort (15.8% of 183 symptomatic individuals) exhibited some degree of lymphedema: 10 patients from the Null group, 3 from Truncation, 4 from Enhancer, 11 from ZF2 mutations and 1 from C-term. Of the 11 individuals with lymphedema and missense mutations, 6 carried mutations demonstrated to impact *PROX1* promoter binding[22] (4 R361, 1 each C373, R396).

Hearing loss in GATA2 deficiency was first described as part of Emberger syndrome in 2011[5] followed by identification in a broader group of *GATA2* patients in 2014[9], in which it was hypothesized to be related to aminoglycoside exposure. There were 45 individuals within our cohort with documented hearing loss (45/232, 19.4%); 8 documented by age 3 (8/232, 3.4%), and an additional 3 by age 10 (11/232, 4.8%), an order of magnitude larger than the CDC estimate of 5-7/1000 children[23]. Hearing loss after amikacin exposure is recognized but is

dependent on dosage and duration. As a result, we separated hearing loss by documented or suspected (after NTM infection) amikacin treatment. Congenital or bilateral hearing loss before onset of infections was seen in 8 individuals, while an additional 20 had sensorineural or conductive hearing loss or tinnitus *prior to* potential aminoglycoside exposure (28/232, 12.1%). An additional 17 (17/232, 7.4%) had been treated for NTM or *Pseudomonas,* with the onset of hearing loss after infection, possibly compatible with antibiotic exposure. These findings support a critical role for GATA2 in the development and maintenance of the inner ear. In fact, dominant *GATA3* mutations lead to hypothyroidism, deafness and renal insufficiency syndrome (HDR)[24]. There is clear overlap between *Gata2* and *Gata3* expression in the inner ear[25]. Examination of the Encyclopedia of DNA Elements (ENCODE) (encodeproject.org) track on the UCSC genome browser (genome.ucsc.edu) identified GATA2 binding to an intronic enhancer element within *GATA3* (Supplemental Figure 8A), suggesting direct regulation of *GATA3* by GATA2. Additionally, GATA2 binds to both the promoter and an intronic enhancer within *GDF6* (Supplemental Figure 8B) which, when deleted, disrupts cochlear development[26]. Together, these data suggest that some GATA2 transcriptional targets are critical for development and/or maintenance of the inner ear and may account for the hearing impairment seen with GATA2 deficiency.

Additional non-hematopoietic phenotypes included thrombotic events, identified in 15 individuals, 8 of whom had more than 1 incident, including deep vein thromboses, pulmonary embolism, stroke, renal infarct and portal vein thrombosis. Nine women reported miscarriages, 6 having more than 1.

Non-hematologic malignancies

Non-hematologic malignancies occurred in 46/232 cohort members (19.8%) including 5 individuals with 2 separate non-hematologic malignancies. Of these, the vast majority (34/46, 73.9%) had genital dysplasia or cancer. In addition, 12/46 (26.1%) had squamous cell carcinoma including skin (n = 5), oral mucosa (n = 2), or both (n = 1). Additionally, 4/46 (8.7%) had breast cancer, 4/46 (8.7%) had basal cell carcinomas, 2/46 (4.3%) had renal cancer, and we found one case each of lung cancer, desmoid tumor, leiomyosarcoma, metastatic melanoma and pancreatic cancer. The high rate of genital dysplasia/cancer and prevalence of squamous cell carcinoma is consistent with the high rate of HPV infection seen in GATA2 deficiency.

To determine if the rate of non-hematologic malignancies observed was disproportionately high, we compared our cohort to United States' cancer incidence reported in the National Cancer Institute's SEER database (seer.cancer.gov) for 2020. We grouped our cohort into the same 5-year age brackets used by SEER and compared the proportion of individuals in each interval with the specific malignancy by age. Since leukemic transformation is common in GATA2 deficiency, we first compared rates of leukemia between the two groups. As expected, we observed a much higher proportion of leukemia among our cohort (Figure 4A). Next, we compared the presence of any malignancy, either leukemia or non-hematopoietic. Consistent with the increased leukemia rate and our observation that nearly 20% of GATA2 patients were diagnosed with a non-hematopoietic malignancy, the GATA2 cohort demonstrated a higher proportion of malignancy than the general population (Figure 4B). Lastly, we examined the

recurrent malignancies - HPV-related anogenital and breast cancer in women (Figure 4C) and anogenital cancers in men (Figure 4D), again finding a much higher proportion among GATA2 patients than the general population. Across all comparisons, there appears to be an earlier age of onset among GATA2 patients, especially across the HPV-related anogenital malignancies. Together these findings reveal a previously unappreciated increased rate of non-hematopoietic malignancies among the GATA2 cohort and possibly an earlier age of onset as well, although larger cohort studies would be needed to confirm this.

Genotype-Phenotype correlations

Examination across the assembled cohort of GATA2 mutation-positive individuals showed a median age of symptom onset of 20 years (Supplemental Figure 9). We then divided the mutation groups between those with median symptom onset <20 years ("early" n=102, including Null, Truncation, and the recurrent ZF2 mutations at R396 and T354), or >20 years ("late" n=107, including recurrent ZF2 mutations at R361, R398, C-term and Enhancer). We performed comparisons between the two groups for presence and/or age of specific phenotype onset. This analysis found significant differences between the groups for events which occur early in GATA2 deficiency: cytopenias, infections, HPV (including warts), thrombotic events, hearing loss, and hearing loss before documented antibiotic exposure. In each case these phenotypes had a higher prevalence in the early onset group (Figure 5A). No significant differences in prevalence were observed for MDS, leukemia/lymphoma or non-hematologic malignancies between early and late onset groups. Both cytopenias and MDS occurred at significantly younger ages in the early compared to the late group (cytopenias - 18 years vs.

28.5 years, respectively, P=0.0052; MDS - 20 years vs. 31 years, respectively, P=0.0098) (Figure 5B); there was no difference in age of first infection or onset of lymphedema.

Thirty-eight patients had symptom onset recorded before 10 years; 25 of whom (65.8%) carried one of the "early" mutations classified above. The most frequent early-onset symptom was viral infection, usually warts (15/38 patients, 39.5%, one of whom had hearing loss as well). Each of those presenting in childhood with impaired hearing was diagnosed by age 3 years and carried a Truncation mutation (5/5). It is noteworthy that 4 of 5 patients presenting with lymphedema before age 10 had Null mutations; together these findings suggest effects of mutation type on specific pathways driving phenotypes. Larger cohorts will be required to validate these observations.

Sex and GATA2 Deficiency

We next asked if there were differences between age at symptom onset by sex. Examination of the 141 females (60.8%) and 91 males (39.2%) in our cohort found no sex difference in age at symptom onset (Figure 6A). Evaluation of presenting symptoms across 183 symptomatic individuals (110 females, 73 males) showed similar percentages of hematologic (40.9% vs 41.0%), non-hematologic (12.7% vs 17.8%) and infectious manifestations (46.4% vs 41.0%) between sexes. However, NTM infections were the presenting symptom among males 3-fold more than among females (14/73, 19.2% vs 7/110 6.4%, P=0.0096) (Figure 6B). Conversely, females had almost double the infections other than NTM as their presenting symptoms

compared to males (40.0% vs 21.9% P=0.0155) (Figure 6B), with HPV and other viral pathogens predominating (Figure 6C).

During disease progression, bone marrow stress may lead to clonal hematopoiesis and the acquisition of somatic mutations and cytogenetic abnormalities[27,28]. Among the 183 symptomatic patients, 23/72 (31.9%) males and 29/111 (26.1%) females developed gross chromosomal abnormalities. Consistent with previous reports on GATA2 deficiency[6] and myelodysplastic syndromes in general[28], isolated monosomy 7 (-7) (n=15) and trisomy 8 (+8) (n=23) were the most frequent cytogenetic variants. We also observed -7/+8 together (n=5) and +8 with +1 der(1;7)(q10;p10) (n=3) as well as additional isolated cytogenetic variants (Supplemental Table 4). Among patients exhibiting symptoms consistent with GATA2 deficiency, 17/72 males (23.6%) developed loss of chromosome 7 (isolated -7, +1der(1;7)(q10;p10), -7/+8) compared to 6/111 females (5.4%, P=0.0004, Fisher's Exact). In contrast, the majority of individuals with chromosome 8 abnormalities (+8; +8, +1(der); +8, -11) were female (20/111 vs 5/72) (P=0.0460, Fisher's Exact).

Discussion

Most previous studies in GATA2 deficiency have focused on patients with later-stage GATA2 deficiency[7,9-11,21,29], successfully describing disease manifestations. Importantly, patients with normal bone marrow lack somatic mutations leading to disease progression including MDS and leukemic transformation[30-32]. Our proactive search for, and inclusion of, all *GATA2*-mutated relatives allowed us to more fully explore penetrance and genotype/phenotype correlations

prior to somatic mutations or malignant transformation. In order to facilitate comparison, we classified mutations by reported residual GATA2 activity. Null and Truncation mutations had the earliest onset and highest penetrance. Enhancer mutations, which reduce but do not eliminate wild-type GATA2 protein from the mutant allele, were the least penetrant, with a median age of symptom onset of 45 years and a significant proportion remaining healthy into their seventh decade. The ZF2 missense mutations showed differences among the four recurrently mutated amino acids, with R396 and T354 having median ages of symptom onset similar to Truncation and Null mutations (16 and 19 years vs 13 and 17 years, respectively). In contrast, R361 and R398 mutations had median ages of onset a decade later (22 and 34 years, respectively).

Dickinson et al reported frameshift mutations associated with earlier presentation than substitution mutations[8]. In their cohort of 30 patients, 10 had early frameshifts predicted to cause NMD, 9 carried R398 mutations while 11 carried other ZF2 mutations. The median ages of presentation across their three groups were 17.6, 34 and 21.6 years, respectively. These are closely aligned with our findings of 17 years for patients with Null mutations and 35 years for those with R398 mutations.

Nonhematologic phenotypes reported in other cohorts were similar to ours. Lymphedema (23%) and deafness (9%) were described in a *GATA2*-mutated, pediatric MDS cohort[10]. A French and Belgian cohort[11,32] reported lymphedema in 15% but only 1 case of deafness (1.3%). In a follow-up of the same group plus patients from the UK[32], lymphedema was 15.8% but deafness was not independently reported. In comparison, among our symptomatic patients, 15.9% had

lymphedema whereas hearing loss occurred in 24.5%. While lymphedema incidence was consistent across cohorts, the differences in deafness prevalence may reflect ages studied, as we had several cases of adult-onset deafness; our median age at presentation or symptom-free assessment was 19 years. Also, we included later-onset hearing impairment, while the other two groups reported congenital deafness only. We also employed dedicated audiology assessment for all subjects regardless of reported symptoms, which likely increased sensitivity for mild hearing impairment.

Until now, no systematic review of non-hematologic malignancies in GATA2 deficiency has been reported. We identified a significant risk for HPV-driven anogenital cancers and an increased risk of breast cancers, in addition to the previously known risk of leukemic transformation in GATA2 deficient patients. Whether this reflects impaired NK surveillance, impaired T cell surveillance, or some other immune or non-immune defect remains to be seen. The very favorable response of anogenital HPV disease to HSCT suggests that restoration of immune function is ameliorative, at least for those malignancies. Given the effectiveness of the HPV vaccine, we typically vaccinate early, prior to exposure to the high-risk HPV strains and while patients are still able to mount an effective immune response. The role for additional cancer screening beyond general recommendations is undefined, but we suggest a high index of suspicion be maintained.

GATA2 is tightly regulated with selective pressures against damaging alleles. Among patients with haploinsufficiency due to deletions, NMD, truncating mutations that removed the ZF2, or missense mutations that directly impact binding to the WGATAR motif (R396[33]), 60/65

individuals were symptomatic, although 3 of the asymptomatic individuals are < 13 years. Missense mutations varied in age of onset and penetrance; those associated with early median age of onset, T354 and R396, have been shown to have severely reduced transactivation capacity[29,31]. Chong et al reported differential effects on disease depending on the specific GATA2 amino acid mutated[29]. They examined 94 reported individuals with either T354M, R396Q, or R398W for the cumulative incidence of MDS/AML or immunodeficiency: R396Q had the earliest disease onset and R398W the latest, consistent with our findings. While our in vitro assays were informative for decreased transactivation of ZF2 variants, a one-day transfection assay is unlikely to recapitulate all the nuances of life-long hematopoiesis, which may allow the eventual expression of even mild variants.

The latest median age of onset was within the intronic Enhancer group, in which only 24/50 individuals were symptomatic. Enhancer patients express some level of normal protein from the mutant allele[13], providing more WT GATA2 than any other category, but still insufficient for normal, life-long hematopoiesis. This variability across mutation types is concordant with differences in embryonic lethality between *Gata2*$^{-/-}$ mice (embryonic day 10.5[34], E10.5) and +9.5 intronic enhancer knockout mice (E13.5[35]) corresponding to different stages of hematopoietic development. *Gata2*$^{-/-}$ mice fail to transition from the aorto-gonad-mesonephros to fetal yolk sac/fetal liver, while +9.5$^{-/-}$ deletion impacts definitive hematopoiesis and vascular integrity. Heterozygous *Gata2*$^{+/-}$ mice had reduced numbers of hematopoietic stem cells that fared poorly in competitive transplantation assays[36]. These mouse models demonstrate differential requirements for GATA2 function during embryogenesis and may parallel the differences in disease onset observed among patients with different genotypes.

Phenotypic variability within mutation groups suggests additional as yet undiscovered factors, which could include *trans*-acting genetic variants impacting *GATA2* transcription or GATA2-targets. For missense changes, differing residual GATA2 activity may account for the specific phenotypic differences observed. Alternatively, environmental exposures could influence the course of disease in some patients[37,38]. Since these data were derived from patients and families referred to the NIH, in which at least one individual had symptomatic GATA2 deficiency, our data do not include individuals carrying *GATA2* variants of uncertain significance yet lacking clinical phenotypes. While mutation positive relatives of a deceased proband were included after genetic testing, our cohort does not include individual patients with early, fatal disease lacking a genetic diagnosis or their relatives. Additionally, the use primarily of DNA from whole blood does not provide screening for apparent mutation negative parents who could be germline mosaic, passing the mutation to the affected individual. There is also a small possibility that some singleton cases may carry somatic mutations yet have a high enough variant allele frequency to be mistaken for germline. This distinction cannot be determined without testing a germline control tissue such as hair follicles or skin fibroblasts. Understanding the impact of mutation type on disease onset and penetrance is critical to determining the best monitoring frequencies and management strategies.

Sex differences were observed within our GATA2 cohort. Consistent with studies of pediatric and adolescent MDS and leukemia[39,40], partial or complete monosomy-7 was seen more frequently in males than females. Additionally, while overall frequency of mycobacterial

infections was similar between males and females, males were more likely to have mycobacterial infections as their presenting symptom. HPV and other viral infections were more likely to be the presenting symptoms in females, possibly due to increased surveillance for women. Since most GATA2 patients maintain immunoglobulin levels even as B-cell numbers decline[11], vaccination prior to the loss of B-cells may likely provide some protection against HPV for those mutation-positive children vaccinated.

Almost a decade and a half after the initial clinical descriptions and identification of *GATA2* as a disease-causing gene, its phenotype now includes cytopenias, leukemia and mycobacterial illnesses[1-4], lymphedema[5], hearing loss, thrombotic events[9], as well as HPV related neoplasms[9]. Transplant cohort studies have shown that hematopoietic stem cell transplantation (HSCT) is curative for most hematologic phenotypes[41,42]. However, full resolution of HPV-related disease has been variable[42], and non-hematopoietic phenotypes including lymphedema and hearing loss have not be corrected by HSCT. Additionally, cases of donor-derived[43] and relapse[44] hematologic malignancies post-HSCT have been reported. Worse outcomes were associated with pre-HSCT disease severity and monosomy-7[45]. GATA2 haploinsufficiency is a complex syndrome with clear genotype/phenotype correlations. A genetic diagnosis empowers appropriate prophylaxis, successful hematopoietic transplants, and proactive family screening. The newer generation of GATA2 deficient patients will benefit from improved understanding of disease onset and progression, which will allow preemptive intervention prior to development of hematologic malignancies. It will be important to temper our prognoses and treatments with

these new understandings and to identify underlying molecular, cellular and environmental events that affect them.

References

1. Vinh DC, Patel SY, Uzel G, et al. Autosomal dominant and sporadic monocytopenia with susceptibility to mycobacteria, fungi, papillomaviruses, and myelodysplasia. Blood. 2010;115(8):1519-1529.
2. Hsu AP, Sampaio EP, Khan J, et al. Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. Blood. 2011;118(10):2653-2655.
3. Hahn CN, Chong C-E, Carmichael CL, et al. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. Nat Genet. 2011;43(10):1012-1017.
4. Dickinson RE, Griffin H, Bigley V, et al. Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. Blood. 2011;118(10):2656-2658.
5. Ostergaard P, Simpson MA, Connell FC, et al. Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). Nat Genet. 2011;43(10):929-931.
6. Calvo KR, Vinh DC, Maric I, et al. Myelodysplasia in autosomal dominant and sporadic monocytopenia immunodeficiency syndrome; diagnostic features and clinical implications. Haematologica. 2011;96(8): 1221-1225.
7. Pasquet M, Bellanne-Chantelot C, Tavitian S, et al. High frequency of GATA2 mutations in patients with mild chronic neutropenia evolving to MonoMac syndrome, myelodysplasia, and acute myeloid leukemia. Blood. 2013;121(5):822-829.

8. Dickinson RE, Milne P, Jardine L, et al. The evolution of cellular deficiency in GATA2 mutation. Blood. 2014;123(6):863-874.
9. Spinner MA, Sanchez LA, Hsu AP, et al. GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics, and immunity. Blood 2014;123(6):809-821.
10. Wlodarski MW, Hirabayashi S, Pastor V, et al. Prevalence, clinical characteristics, and prognosis of GATA2-related myelodysplastic syndromes in children and adolescents. Blood. 2016;127(11):1387-1397.
11. Donadieu J, Lamant M, Fieschi C, et al. Natural history of GATA2 deficiency in a survey of 79 French and Belgian patients. Haematologica. 2018;103(8):1278-1287.
12. Homan CC, Venugopal P, Arts P, et al. GATA2 deficiency syndrome: A decade of discovery. Hum Mutat. 2021;42(11):1399-1421.
13. Hsu AP, Johnson KD, Falcone EL, et al. GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. Blood. 2013;121(19):3830-3837.
14. Kozyra EJ, Pastor VB, Lefkopoulos S, et al. Synonymous GATA2 mutations result in selective loss of mutated RNA and are common in patients with GATA2 deficiency. Leukemia. 2020;34(10):2673-2687.
15. Silverio-Alves R, Kurochkin I, Rydstrom A, et al. GATA2 mitotic bookmarking is required for definitive haematopoiesis. Nat Commun. 2023;14(1):4645.
16. Mir MA, Kochuparambil ST, Abraham RS, et al. Spectrum of myeloid neoplasms and immune deficiency associated with germline GATA2 mutations. Cancer Med. 2015;4(4):490-499.

17. Haddox CL, Carr RM, Abraham RS, et al. Phenotypic heterogeneity associated with germline GATA2 haploinsufficiency: a comprehensive kindred study. Leuk Lymphoma. 2019;60(13):3282-3286.
18. Mace EM, Hsu AP, Monaco-Shawver L, et al. Mutations in GATA2 cause human NK cell deficiency with specific loss of the CD56(bright) subset. Blood. 2013;121(14):2669-2677.
19. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood. 2016;127(20):2391-2405.
20. Geng X, Cha B, Mahamud MR, et al. Multiple mouse models of primary lymphedema exhibit distinct defects in lymphovenous valve development. Dev Biol. 2016;409(1):218-233.
21. Kazenwadel J, Secker GA, Liu YJ, et al. Loss-of-function germline GATA2 mutations in patients with MDS/AML or MnoMAC syndrome and primary lymphedema reveal a key role for GATA2 in the lymphatic vasculature. Blood. 2012;119(5):1283-1291.
22. Kazenwadel J, Betterman KL, Chong CE, et al. GATA2 is required for lymphatic vessel valve development and maintenance. J Clin Invest. 2015;125(8):2979-2994.
23. Van Esch H, Groenen P, Nesbit MA, et al. GATA3 haplo-insufficiency causes human HDR syndrome. Nature. 2000;406(6794):419-422.
24. Zablotsky B, Black LI. Prevalence of children aged 3 - 17 with developmental disabilities, by urbanicity: United States, 2015-2018. Natl Health Stat Report. 2020;(139):1-7.

25. Lillevali K, Matilainen T, Karis A, Salminen M. partially overlapping expression of Gata2 and Gata3 during inner ear development. Dev Dyn. 2004;231(4):775-781.
26. Bademci G, Abad C, Cengiz FB, et al. Long-range cis-regulatory elements controlling GDF6 expression are essential for cochlear development. J Clin Invest. 2020;130(8):4213-4217.
27. Lindsley RC, Ebert BL. Molecular pathophysiology of myelodysplastic syndromes. Annu Rev Pathol. 2013;8:21-47.
28. Zahid MF, Malik UA, Sahail M, Hassan IN, Ali S, Shaukat MHS. Cytogenetic abnormalities in myelodysplastic syndromes: An overview. Int J Hematol Oncol Stem Cell Res. 2017;11(3):231-239.
29. Chong C-E, Venugopal P, Stokes PH, et al. Differential effects on gene transcription and hematopoietic differentiation correlate with GATA2 mutant disease phenotypes. Leukemia. 2018;32(1):194-202.
30. West RR, Calvo KR, Embree LJ, et al. ASXL1 and STAG2 mutations in GATA2 deficiency patients with bone marrow disease and myelodysplastic syndrome. Blood Adv. 2022;6(3):793-807.
31. McReynolds LJ, Yang Y, Wong HY, et al. MDS-associated mutations in germline GATA2 mutated patients with hematologic manifestations. Leuk Res. 2019;76:70-75.
32. Largeaud L, Collin M, Monselet N, et al. Somatic genetic alterations predict hematological progression in GATA2 deficiency. Haematologica. 2023;108(6):1515-1529.
33. Bates DL, Chen Y, Kim G, et al. 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(5):1292-1306.

34. Tsai F-Y, Keller G, Kuo FC, et al. An early haematopoietic defect in mice lacking the transcription factor GATA-2. Nature. 1994;371(6494):221-226.
35. Johnson KD, Hsu AP, Ryu MJ, et al. Cis-element mutated in GATA2-dependent immunodeficiency governs hematopoiesis and vascular integrity. J Clin Invest. 2012;122(10):3692-3704.
36. Rodrigues NP, Janzen V, Forkert R, et al. Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. Blood. 2005;106(2):477-484.
37. Nisse C, Haguenoer JM, Grandbastien B, et al. Occupational and environmental risk factors of the myelodysplastic syndromes in the North of France. Br J Haematol. 2001;112(4):927-935.
38. Avgerinou C, Giannezi I, Theodoropoulou S, et al. Occupational, dietary, and other risk factors for myelodysplastic syndromes in Western Greece. Hematology. 2017;22(7):419-429.
39. Hasle H, Arico M, Basso G, et al. Myelodysplastic syndrome, juvenile myelomonocytic leukemia, and acute myeloid leukemia associated with complete or partial monosomy 7. Leukemia. 1999;13(3):376-385.
40. Hasle H, Alonzo TA, Auvrignon A, et al. Monosomy 7 and deletion 7q in children and adolescents with acute myeloid leukemia; An international retrospective study. Blood. 2007;109(11):4641-4647.
41. Nichols-Vinueza DX, Parta M, Shah NN, et al. Donor source and post-transplantation

cyclophosphamide influence outcome in allogeneic stem cell transplantation for GATA2 deficiency. Br J Haematol. 2022;196(1):169-178.

42. Parta M, Cole K, Avila D, et al. Hematopoietic cell transplantation and outcomes related to Human Papillomavirus disease in GATA2 deficiency. Transplant Cell Ther. 2021;27(5):435.
43. Galera P, Hsu AP, Wang W, et al. Donor-derived MDS/AML in families with germline GATA2 mutation. Blood. 2018;132(18):1994-1998.
44. Sicre de Fontbrune F, Chevillon F, Fahd M, et al. Long-term outcome after allogeneic stem cell transplantation for GATA2 deficiency: An analysis of 67 adults and children from France and Belgium. Br J Haematol. 2024 Aug 19. doi:10.1111/bjh.19691. [Epub ahead of print].
45. Bortnick R, Wlodarski M, de Haas V, et al. Hematopoietic stem cell transplantation in children and adolescents with GATA2-related myelodysplastic syndrome. Bone Marrow Transplant. 2021;56(11):2732-2741.

Table 1. GATA2 mutations and classifications

# patients	# families	Mutation cDNA	Mutation protein	Mutation Category	Early/Late
1	1	lg deletion	lg deletion	Null	Early
4	4	uniallelic expression	uniallelic	Null	Early
3	1	c.1-200_871+527del_ins7	p.M1Kfs*35	Null	Early
2	1	c.58C>T	p.Q20*	Null	Early
1	1	c.229+1G>A		Null	Early
1	1	c.243_244delAinsGC	p.G82Rfs*102	Null	Early
1	1	c.247C>T	p.Q83*	Null	Early
1	1	c.248delA	p.Q83Rfs*35	Null	Early
1	1	c.302delG	p.G101Afs*16	Null	Early
1	1	c.317_318delCT	p.S106Cfs*77	Null	Early
1	1	c.417dupT	p.V140Cfs*44	Null	Early
1	1	c.586_593dup	p.G199Lfs*21	Null	Early
1	1	c.615_618delAGAG	p.E206Tfs*10	Null	Early
1	1	c.680_83del4	p.S227fs*	Null	Early
1	1	c.769_778dup	p.Y260fs*24	Null	Early
3	1	c.793_802del10	p.F265Efs	Null	Early
1	1	c.802G>T	p.G268*	Null	Early
1	1	c.803delG	p.G268fs	Null	Early
1	1	c.842delA	K281Sfs*45	Null	Early
1	1	c.871+2_3insT		Null	Early
1	1	c.898dupG	p.A300Gfs*83	Null	Early
1	1	c.921dupG	p.R308Afs*75	Null	Early
1	1	c.941_951dup	p.A318Tfs*12	Null	Early
3	2	c.988C>T	p.R330*	Null	Early
1	1	c.989_992dup	p.L332Tfs*53	Null	Early
1	1	c.996delC	p.I333Sfs*53	Null	Early
6	5	c.1009C>T	p.R337*	Null	Early
3	1	c.1017G>T	p.L339L (reduced exp)	Null	Early
1	1	c.1017+2T>G	reduced exp	Null	Early
2	1	c.1017+512del28		Enhancer	Late
47	10	c.1017+572C>T		Enhancer	Late
1	1	c.1017+573G>A		Enhancer	Late
1	1	c.1018-50_1143+247del	p.340_381del	Truncation	Early

1	1	c.1018-1G>A	p.340_381del	Truncation	Early
1	1	c.1018-1_1023del7	p.S340Pfs*44	Truncation	Early
2	2	c.1021delG	p.A341Pfs*46	Truncation	Early
1	1	c.1024_1025insG	p.A342Gfs*41	Truncation	Early
2	2	c.1024_1025insGCCG	p.A342Gfs*41	Truncation	Early
5	1	c.1040_1041CC>AT	p.T347N	ZF2	
1	1	c.1045T>C	p.C349R	ZF2	
15	8	c.1061C>T	p.T354M	ZF2	Early
2	1	c.1061C>A	p.T354K	ZF2	Early
1	1	c.1078T>A	p.W360R	ZF2	
5	4	c.1081C>T	p.R361C	ZF2	Late
5	4	c.1082G>A	p.R361H	ZF2	Late
3	1	c.1082G>C	p.R361P	ZF2	Late
1	1	c.1083_1094del12	p.R361delRNAN	ZF2	Late
6	3	c.1084C>T	p.R362*	Truncation	Early
2	1	c.1099insG	p.D367Gfs*15	Truncation	Early
2	2	c.1113C>G	p.N371K	ZF2	
2	2	c.1114G>A	p.A372T	ZF2	
3	1	c.1116_1130del15	p.C373del5	ZF2	
3	2	c.1123C>T	p.L375F	ZF2	
2	1	c.1128C>A	p.Y376*	Truncation	Early
1	1	c.1150delA	p.R384fs	Truncation	Early
3	1	c.1159_1160dupAC	p.M388Pfs*1	Truncation	Early
1	1	c.1160C>A	p.T387N	ZF2	
4	1	c.1163T>C	p.M388T	ZF2	
1	1	c.1168_1170del	K390del	ZF2	
6	6	c.1186C>T	p.R396W	ZF2	Early
1	1	c.1186C>G	p.R396G	ZF2	Early
10	6	c.1187G>A	p.R396Q	ZF2	Early
31	12	c.1192C>T	p.R398W	ZF2	Late
1	1	c.1193G>A	p.R398Q	ZF2	Late
1	1	c.1193G>T	p.R398L	ZF2	Late
1	1	c.1277C>G	p.S426C	C-term	Late
7	1	c.1339A>C	p.S447R	C-term	Late
2	1	c.1348_*74del169	p.G450Efs*34	C-term	Late

Table 2. Penetrance by Mutation Class or Amino Acid

Mutation Class	Individuals (#)	Symptomatic (#)	Penetrance %
Enhancer	50	27	54
Null	44*	42	95.5
Truncation	22	21	95.5
ZF2	104	85	81.7
C-term	10	8	80

* 2 individuals < 1 year omitted from analysis

Amino Acid	Individuals (#)	Symptomatic (#)	Penetrance %
T354	17	14	82.4
R361	13*	11	84.6
R396	17	17	100.0
R398	33	24	72.7

* 1 individual < 1 year omitted from analysis

Figure 1. GATA2 cohort mutation classifications and locations. A. Mutations were grouped into categories based on the effect of the mutation on mRNA (Enhancer and Null mutations) or GATA2 protein (Truncation, ZF2, and C-term). Wild-type allele depicted on top with amino acid numbers of defined domains. Enhancer mutations cause reduced levels of wild-type transcript and protein from the mutant allele. ZF2 mutations are small missense or in-frame deletions wholly within ZF2. Truncation mutations predicted to have stable protein but delete all or part of ZF2. C-term mutations occur after ZF2. Null mutations lack transcript and protein from the mutant allele. Predicted GATA2 functional level from mutant allele (left) or mutant plus WT alleles (right) B. Mutation diagram. Top shows genomic *GATA2* organization, middle is linear protein, bottom is enlargement of ZF2. Mutations are shaped by mutation type and color coded by mutation classification. Numbers within symbols indicate independent probands with mutation.

Figure 2. Differential age of symptom onset by GATA2 mutation. A. Kaplan-Meier plot of symptom onset (age at presenting symptom) across mutation groups. Table below indicates number of at-risk individuals at each 5 year age increment. B. Hazard ratios of mutation groups compared to Enhancer (top) or recurrently mutated AA compared to R398 (bottom). C. Kaplan-Meier plot of symptom onset for patients with mutations at one of 4 recurrently mutated ZF2 amino acids. Table below indicates number of at-risk individuals at each 5 year age increment. D. Correlation between median age of symptom onset and penetrance by mutation category or recurrently mutated amino acid. (R^2 = 0.8908 , P = 0.0001. P value and R^2 calculated using simple linear regression).

Figure 3. GATA2 transactivation is dose dependent and decreased by patient mutations. A. Effect of decreasing amounts of GATA2-WT on luciferase production in transfected HEK cells. Differing amounts of GATA2-WT were transfected into HEK cells. After 20 hours, cells were lysed and renilla-normalized luciferase activity was measured. B. GATA2 induced luciferase by empty vector (EV grey), somatic, activating (L359V red), patient (blue), or gnomAD (white) variants normalized to WT. Data are the results of 5 to 13 independent experiments of each variant; within each experiment, samples were tested in triplicate, averaging the results. P-value calculated using Unpaired t-test comparing to GATA2-WT 500 ng (A) or Kruskal-Wallis with Dunn's multiple comparison test (B). * $P<0.05$, ** $P<0.01$, *** $P<0.001$, **** $P<0.0001$. Error bars represent mean + SEM

Figure 4. Increased prevalence of malignancies among GATA2 cohort. Patients were grouped by age into 5 year bins and the proportion of individuals with the specific malignancy was plotted for each age (pink) compared to the SEER database incidence (blue). A. Leukemia, B. Any malignancy, C. Female anogenital malignancy (filled circles) or breast cancer (open circles), D. Male anogenital malignancy.

Figure 5. Increased prevalence of specific phenotypes in Early mutations. A. Mutation categories or recurrently mutated amino acids were grouped by median age <20 years "Early" or >20 years "Late". Specific GATA2-related phenotypes occur more frequently in the "Early" mutation group compared to "Late", shown in red. B. Age of onset of specific phenotypes among patients in "Early" versus "Late" mutation groups. Cytopenias and MDS appear earlier in

"Early" patients (P = 0.0052, P=0.0098 respectively). No difference was observed in age of 1st infection or occurrence of lymphedema. Error bars – median with 95% confidence interval; q values calculated using Mann-Whitney with two-stage linear step-up procedure of Benjamini, Krieger, and Yekutieli, FDR <0.05.

Figure 6. Sex differences in presenting symptoms. A. Kaplan-Meier plot of symptom onset (age at presenting symptom) comparing sexes. B. Patients presenting with infection as the first symptom split by sex and nontuberculous mycobacterial infection (NTM) or non-NTM infections. P-value calculated using Fisher's Exact with Benjamini, Krieger, Yekutieli two-stage step-up method for 5% FDR C. Breakdown of presenting infections among patients with infection as the first symptom split by sex.

Fig 1

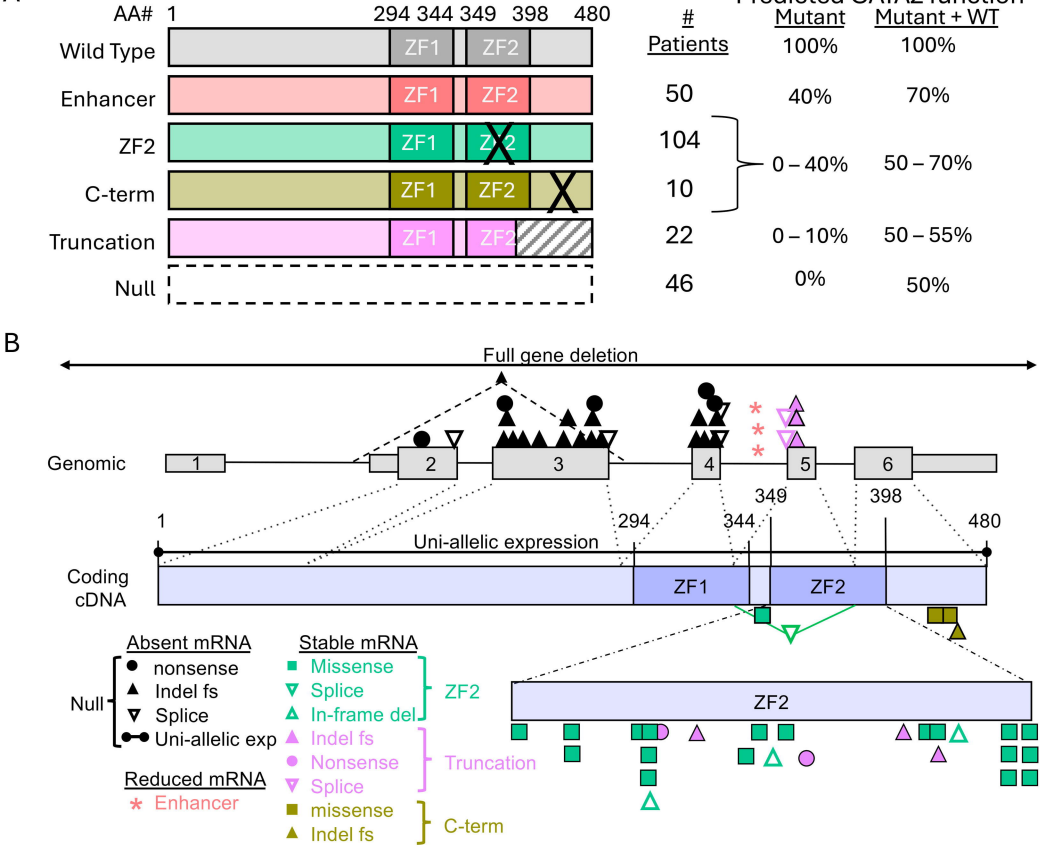
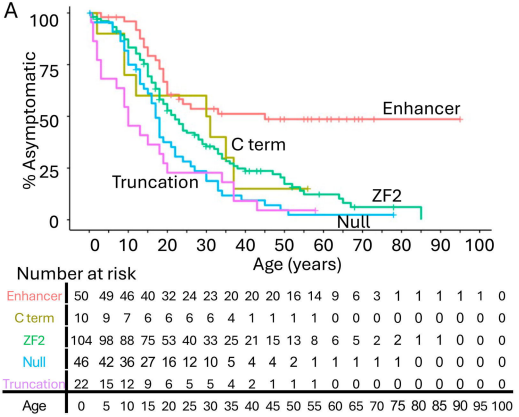
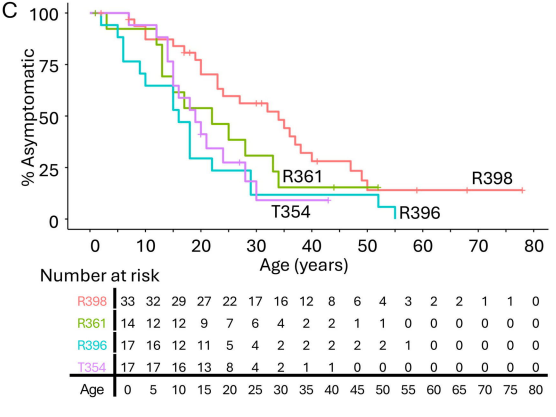

A
AA# 1
294 344 349 398 480
Wild Type
Enhancer
ZF2
C-term
Truncation
Null
ZF1
ZF2
Predicted GATA2 function
Patients
Mutant
Mutant + WT
100%
100%
50
40%
70%
104
10
0 – 40%
50 – 70%
22
0 – 10%
50 – 55%
46
0%
50%
B
Full gene deletion
Genomic
1
2
3
4
5
6
1
294
344
349
398
480
Uni-allelic expression
Coding cDNA
ZF1
ZF2
Absent mRNA
Null
nonsense
Indel fs
Splice
Uni-allelic exp
Stable mRNA
Missense
Splice
In-frame del
ZF2
Indel fs
Nonsense
Splice
Truncation
Reduced mRNA
Enhancer
missense
Indel fs
C-term
ZF2

Fig 2

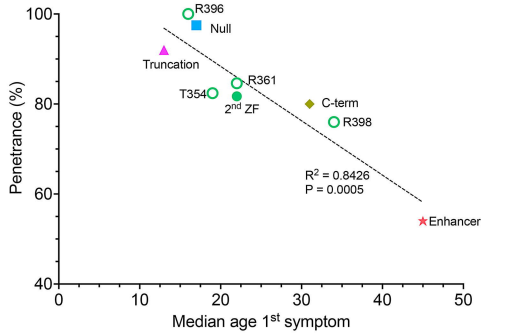


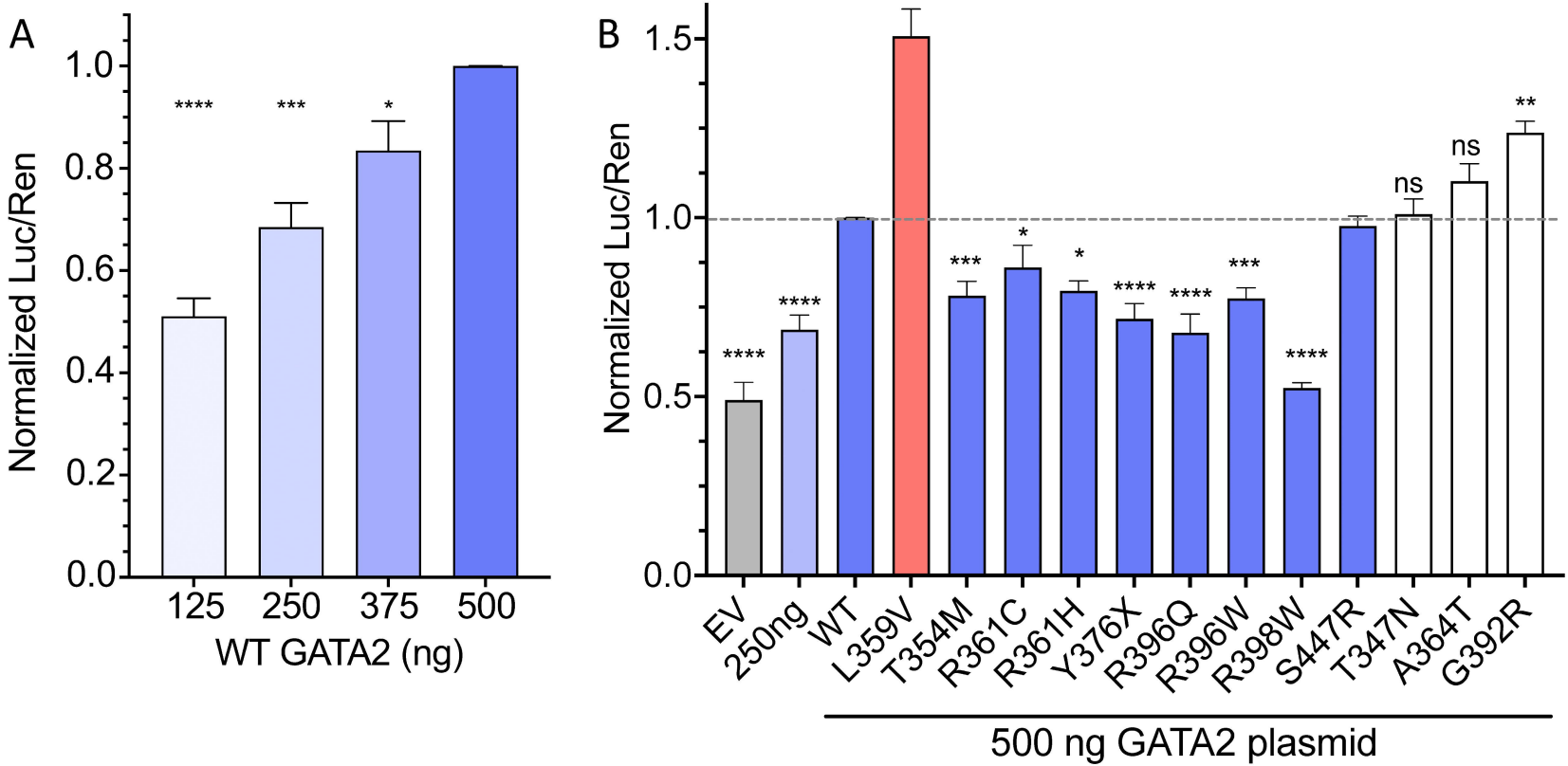


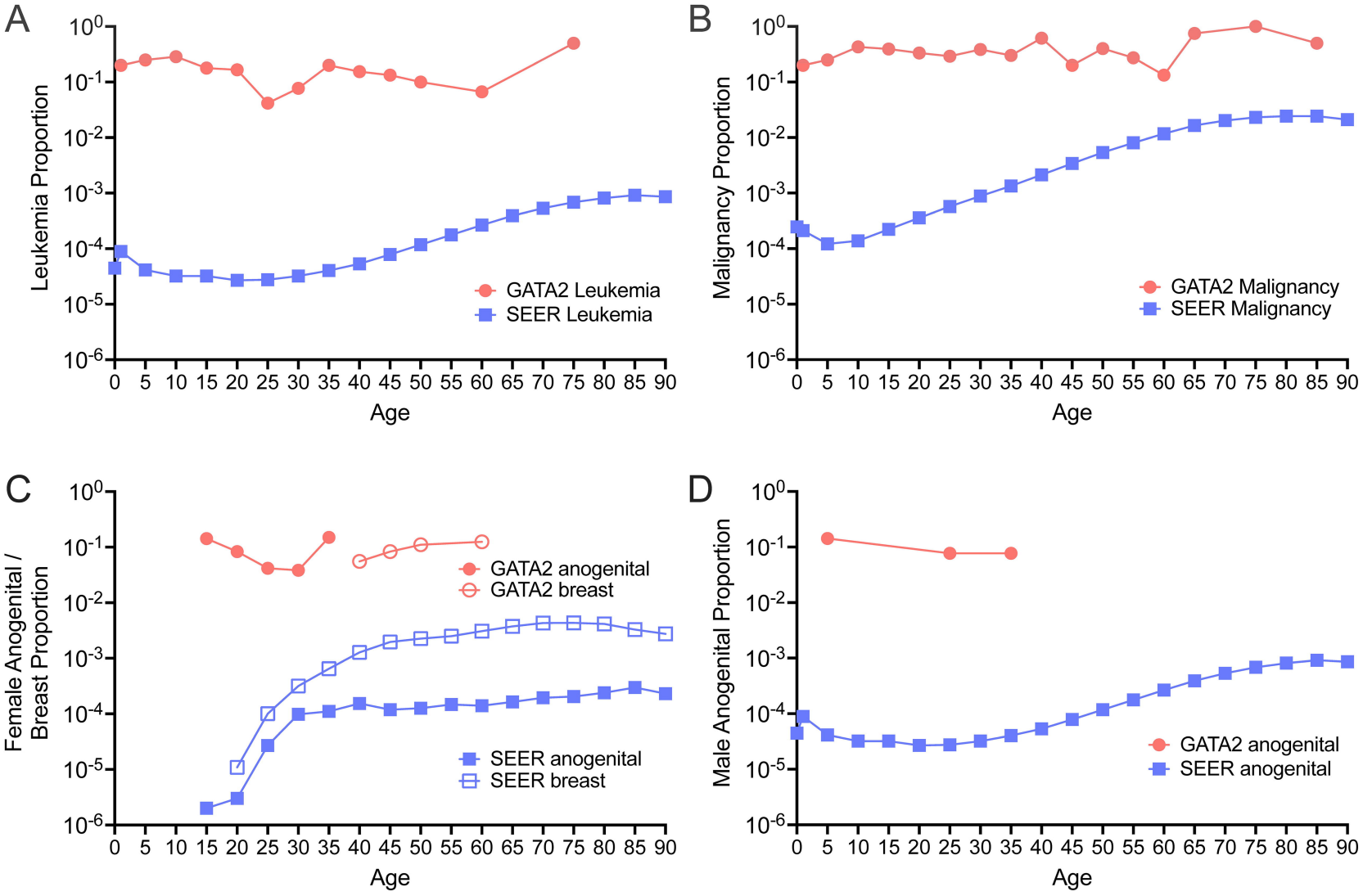
B

Mutation Group vs Enhancer	Hazard Ratio	Confidence Interval	P value
C term	2.23	0.77 - 6.43	0.139
ZF2	2.24	1.25 - 3.99	**0.006**
Null	3.55	1.86 - 6.76	**<0.001**
Truncation	5.63	2.57 - 12.36	**<0.001**
Amino Acid vs R398	**Hazard Ratio**	**Confidence Interval**	**P value**
R361	1.63	0.79 - 3.35	0.186
R396	2.67	1.36 - 5.26	**0.004**
T354	2.16	1.09 - 4.31	**0.028**

D



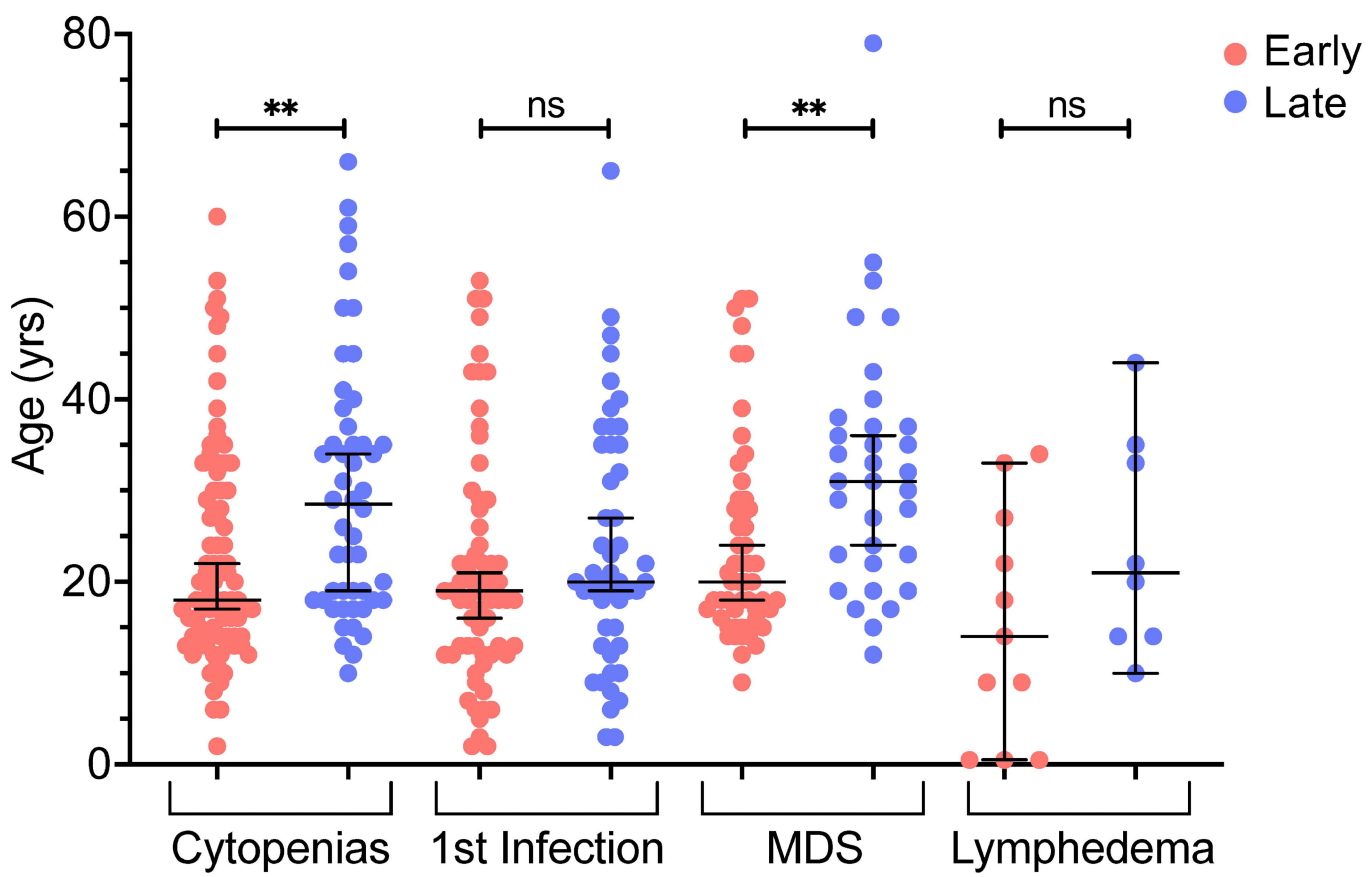

A
Normalized Luc/Ren
1.0
0.8
0.6
0.4
0.2
0.0
****
***
*
125
250
375
500
WT GATA2 (ng)
B
Normalized Luc/Ren
1.5
1.0
0.5
0.0
****
****
****
***
*
****
****
***
****
ns
ns
**
EV
250ng
WT
L359V
T354M
R361C
R361H
Y376X
R396Q
R396W
R398W
S447R
T347N
A364T
G392R
500 ng GATA2 plasmid

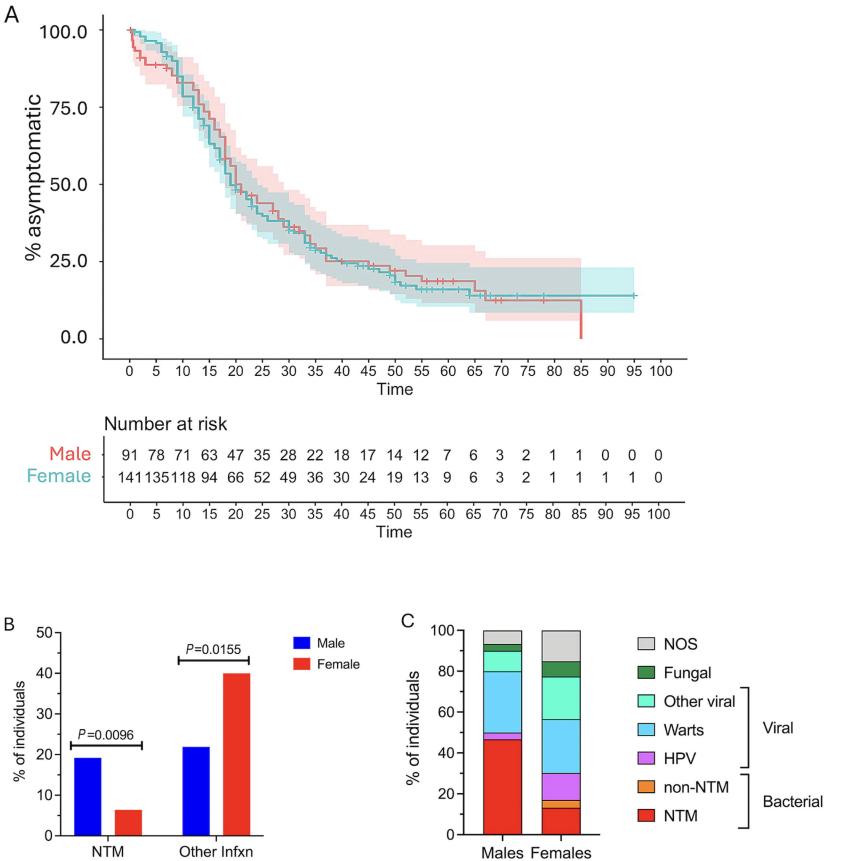

A
Leukemia Proportion
GATA2 Leukemia
SEER Leukemia
Age
B
Malignancy Proportion
GATA2 Malignancy
SEER Malignancy
Age
C
Female Anogenital / Breast Proportion
GATA2 anogenital
GATA2 breast
SEER anogenital
SEER breast
Age
D
Male Anogenital Proportion
GATA2 anogenital
SEER anogenital
Age

A

Symptom	Adj P value	"Early"	"Late"
Cytopenia	0.0007	85/97 (87.6%)	51/100 (51.0%)
Infections	0.0169	73/102 (71.6%)	56/107 (52.3%)
HPV (including warts)	0.029	49/102 (48.0%)	33/107 (30.8%)
HPV (excluding warts)	0.168	26/102 (25.5%)	19/107 (17.8%)
MDS	0.0944	58/85 (68.2%)	35/65 (53.8%)
Leukemia or lymphoma	0.5546	11/102 (10.8%)	13/107 (12.1%)
Thrombotic event	0.0382	11/102 (10.8%)	3/107 (2.8%)
Hearing loss	0.0382	21/102 (20.6%)	10/107 (9.3%)
Hearing loss pre-amikacin	0.0258	13/102 (12.7%)	3/107 (2.8%)
Nonhematologic malignancy	0.437	21/102 (20.6%)	18/107 (16.8%)
Lymphedema	0.2352	15/102 (14.7%)	10/107 (9.3%)

B




A
100.0
75.0
50.0
25.0
0.0
% asymptomatic
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Time
Number at risk
Male 91 78 71 63 47 35 28 22 18 17 14 12 7 6 3 2 1 1 0 0 0
Female 141 135 118 94 66 52 49 36 30 24 19 13 9 6 3 2 1 1 1 1 0
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Time
B
50
40
30
20
10
0
% of individuals
P=0.0096
P=0.0155
Male
Female
NTM
Other Infxn
C
100
80
60
40
20
0
% of individuals
Males Females
NOS
Fungal
Other viral
Warts
HPV
non-NTM
NTM
Viral
Bacterial

Supplementary files to accompany Hsu, et al. GATA2 at 14: Genotype – Phenotype Correlations

Supplemental methods

Data were extracted to a standard form with defined fields for recognized symptoms, age of onset for symptoms, and free-text recording of uncategorized findings. We defined "presenting symptoms" as the first documented GATA2 disease manifestation or atypical illness noted by a physician or recorded in clinical records including lymphedema, hearing impairment, cytopenias, infections, warts, abnormal bone marrow findings, leukemia and non-hematologic malignancy. Lymphedema included reference to lymphedema or documented hydrocele. Hearing impairment included documented congenital deafness, sensorineural or conductive hearing loss or tinnitus. Whenever possible, bone marrow samples were performed or reviewed at our institution by a single hematopathologist, in other cases we relied on outside reports.

Mutations are reported using NM_001145661.2 and NP_001139133.1 as reference genomic and protein sequences respectively. For patients with mutations potentially affecting mRNA processing or stability (premature stop codons or insertions/deletions, and those near splice sites), full-length cDNA analysis was performed when possible to evaluate the effect of the mutation. cDNA amplification of GATA2 was performed as previously reported[1].

Penetrance was defined as the number of mutation positive individuals over the age of 1 year with GATA2 deficiency-compatible clinical symptoms divided by the number of mutation positive individuals. Due to our study selection criteria in which each family has at least one

proband, our penetrance estimates are biased upward relative to a random sample but are directly comparable between mutation categories.

Kaplan-Meier analysis for disease onset and mixed effect Cox regression were performed using R (version 4.2.2)[2] with the survival (version 3.4.0)[3,4] and coxme (version 2.2-18.1)[5] packages. Disease onset was considered to be the age at presence of presenting symptom as defined above and included in Supplemental Table 1. The Cox proportional hazards model is commonly used in survival analysis to model the hazard function which denotes the instantaneous rate of the occurrence of an event at a time given that the subject survived up to that point. Including random effects in the Cox proportional hazards model enables us to incorporate account for familial structure in our data and the resulting non-independence between samples. The kinship matrix (kinship2 R package version 1.9.6)[6] was used to correct for relatedness as a random effect. Code for analysis is available at https://github.com/spaul-genetics/GATA2_Report. Fisher's exact test was calculated for 2 x 2 contingency analyses followed by correction for multiple testing with two-stage step-up method of Benjamini, Krieger and Yekutieli controlling FDR < 0.05.

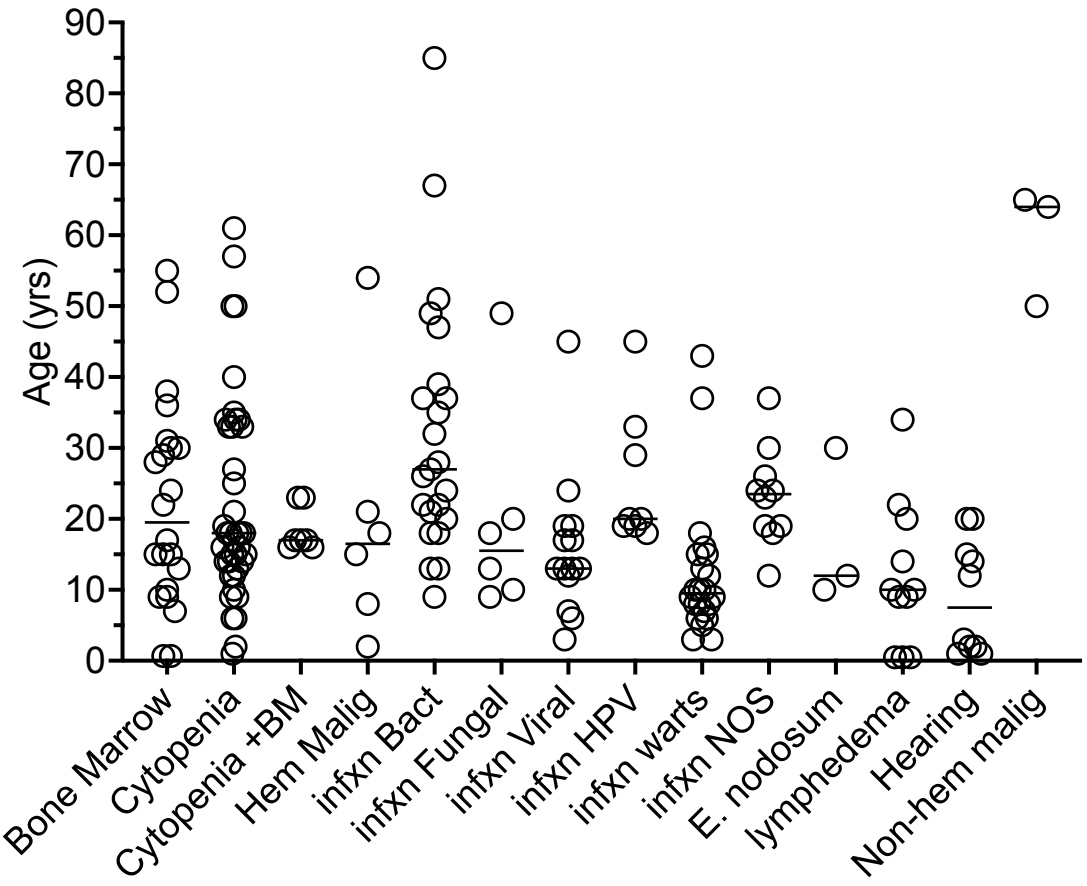
National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) data[7] were obtained from seer.cancer.gov. We investigated the rates of cancers from the SEER 12 registries by age and cancer type for all age groups. Since SEER data is binned in discrete age categories (0-1, 1-4, then 5-year increments), we standardized the GATA2 patients using the age at last patient encounter or transplant. Because GATA2 patients were enrolled over several decades,

and cancer screening processes improved, leading to increased incidence over time, the last year of GATA2 case collection (2020) was selected as the comparator for rates. This choice of operator is expected to bias any estimates downward as an effect of the larger denominator of the most recent data. Since SEER data is patient agnostic, we included each different cancer occurrence among GATA2 patients as a separate event. Data are presented as the proportion of individuals within the age bracket with the specific cancer diagnosis. For leukemia comparisons, “Lymphocytic”, “Myeloid and monocytic”, and “Other” leukemias counts were combined; for female anogenital cancers, “Anus, Anal Canal and Anorectum”, “Rectum”, “Cervix Uteri”, and “Vulva” counts were combined; for male anogenital cancers, “Anus, Anal Canal and Anorectum”, “Rectum”, and “Penis” counts were combined.

Supplementary References

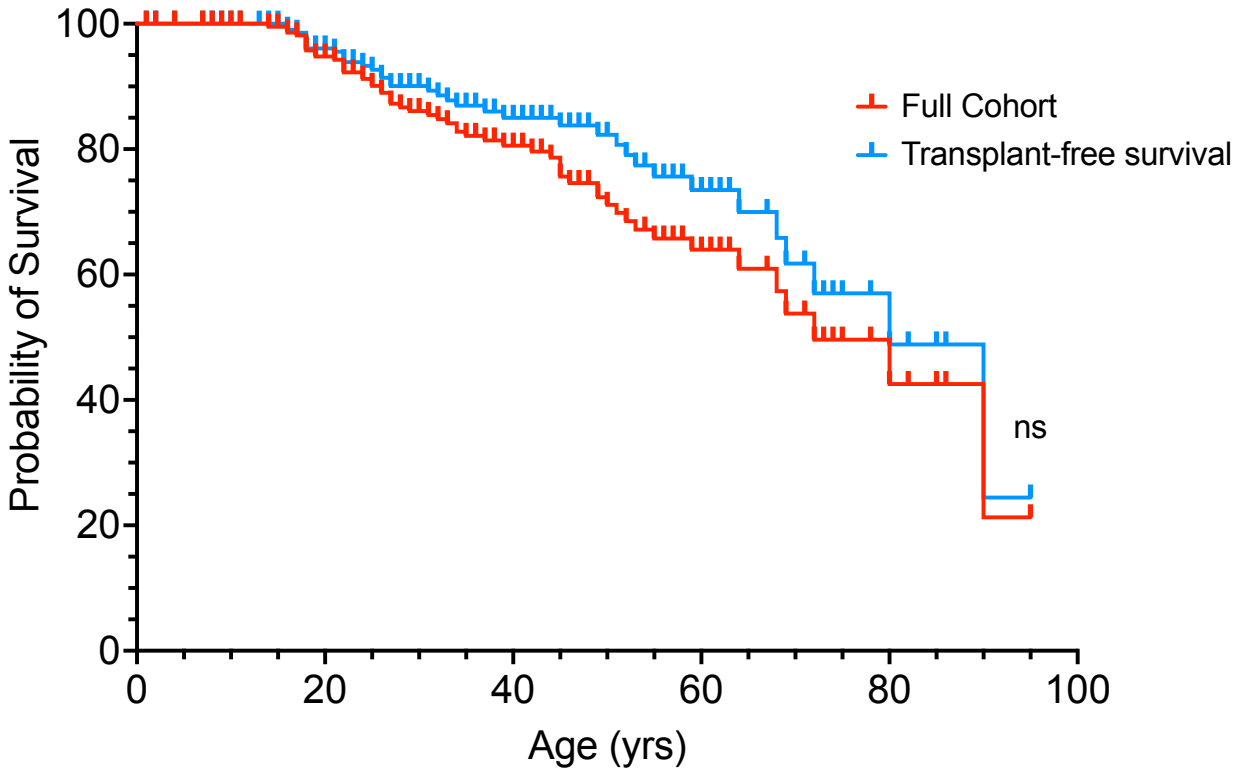
1. Hsu AP, Johnson KD, Falcone EL, et al. GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. Blood. 2013 May 9;121(19):3830-7.
2. R Core Team. R: A language and environment for statistical computing. 2022. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
3. Therneau T. A package for survival analysis in R. R package version 3.4-0. 2022. https://CRAN.R-project.org/package=survival.
4. Therneau TM, Grambsch PM. Modleing survival data : extending the Cox model. 2000. Springer, New York. ISBN 0-387-98784-3.
5. Therneau TM. coxme: Mixed Effects Cox Models. 2022. R package version 2.2-18.1, https://CRAN.R-project.org/package=coxme.
6. Sinnwell J, Therneau T. kinship2: Pedigree Functions. 2022. R package version 1.9.6, https://CRAN.R-project.org/package=kinship2.
7. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER Research Data, 12 Registries, Nov 2024 Sub (1992-2022) – Age-adjusted rates, Linked to county attributes- Time Dependent 1990-2023). National Cancer Institute, DCCPS, Surveillance Research Program, released April 2025, based on the November 2024 submission.

Supplemental Figure 1



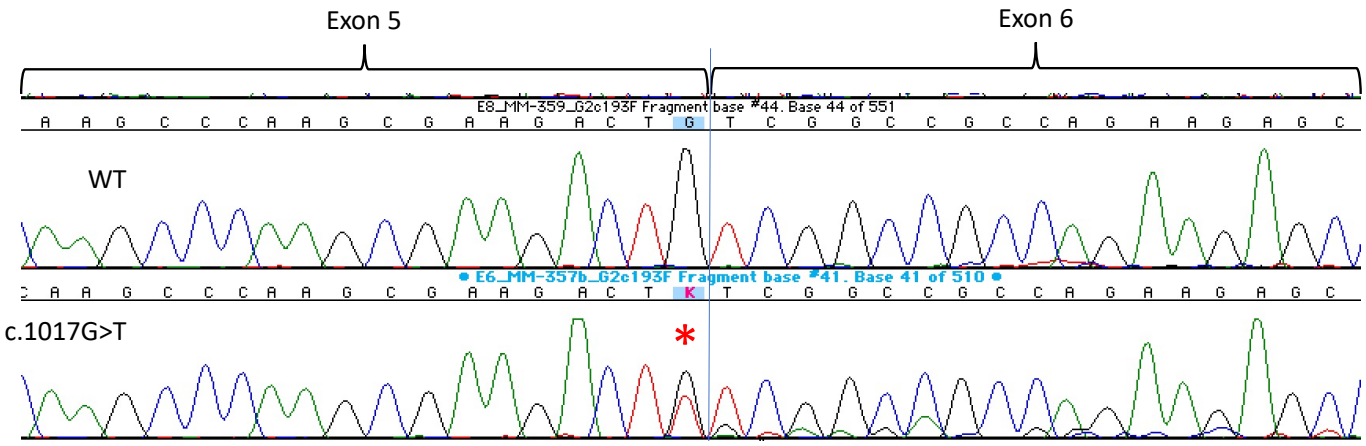
Presenting symptoms across GATA2 cohort. Summary of first symptom observed and age of occurrence.

Supplemental Figure 2.



Kaplan Meier survival curve of GATA2 patients. Red line, survival all patients regardless of transplant status. Blue line, survival censoring patients at age of transplant.

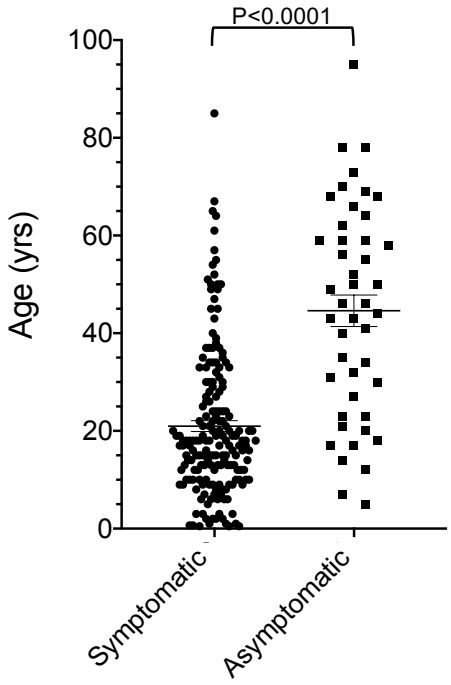
Supplemental Figure 3.



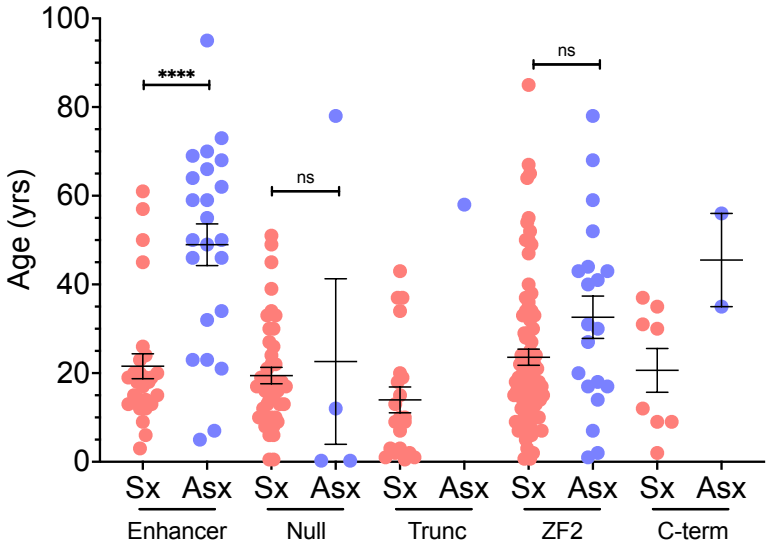
Reduced GATA2 cDNA from c.1017G>T allele. Wild-type sequence shown in top row with appropriate splicing from exon 5 to exon 6. cDNA from c.1017G>T patient peripheral blood mononuclear cells showing the mutation at the last base of exon 5. The mutant transcript is seen in small secondary peaks.

Supplemental Figure 4.

A



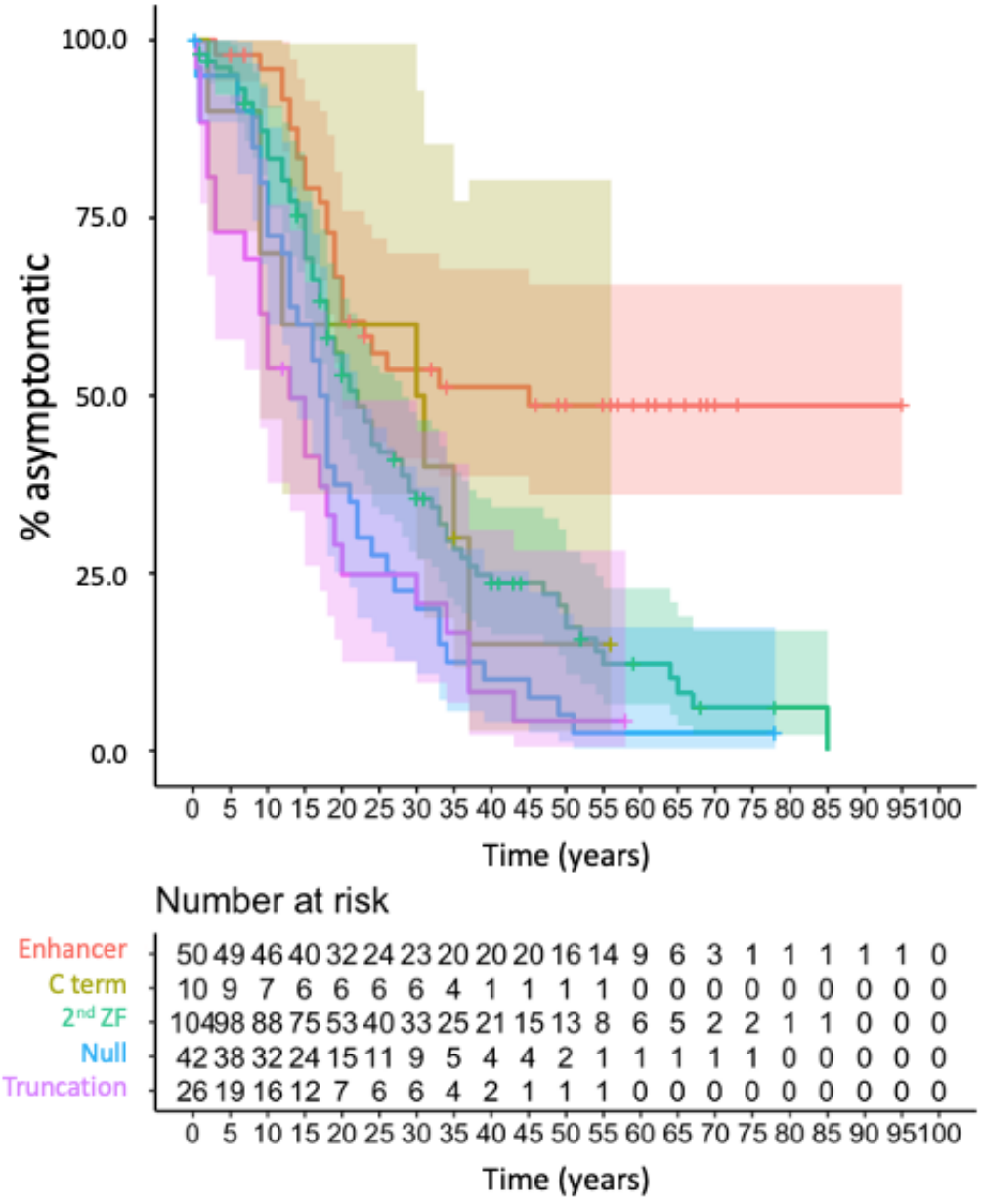
B



Age variation between Symptomatic and Asymptomatic *GATA2* mutation positive individuals.

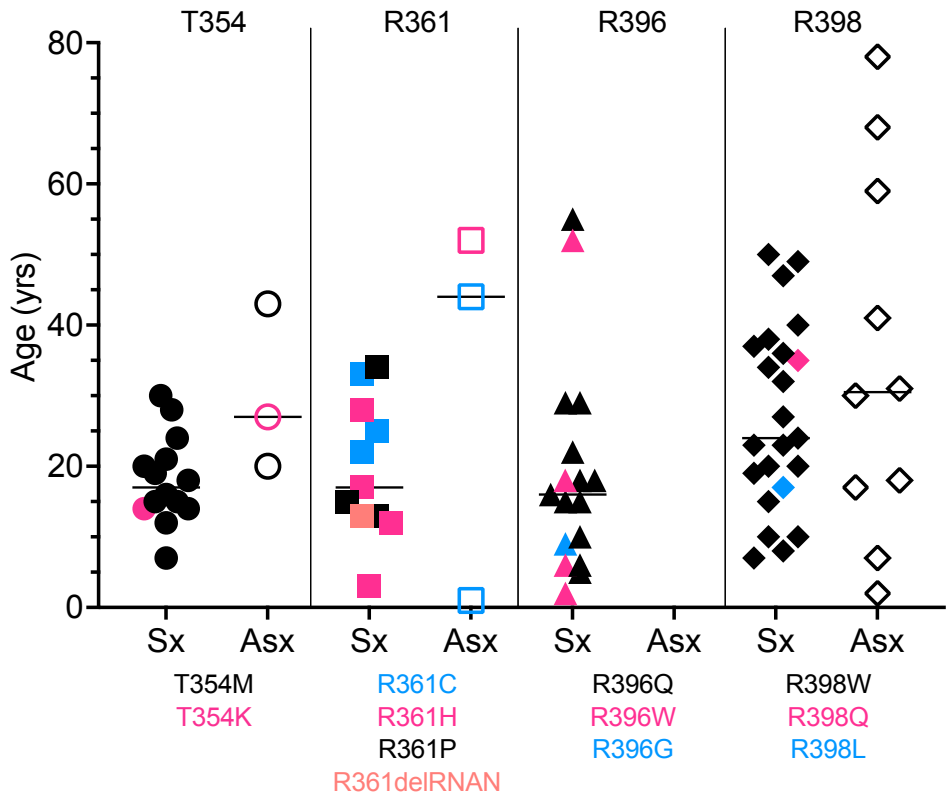
A. Age at symptom onset (Symptomatic) or last clinical visit (Asymptomatic) for GATA2 mutation positive individuals. B. Age at symptom onset as in A with individuals subdivided by mutation group. Bars are median +/- SEM. P<0.0001, Unpaired, two-tailed Mann-Whitney test.

Supplemental Figure 5



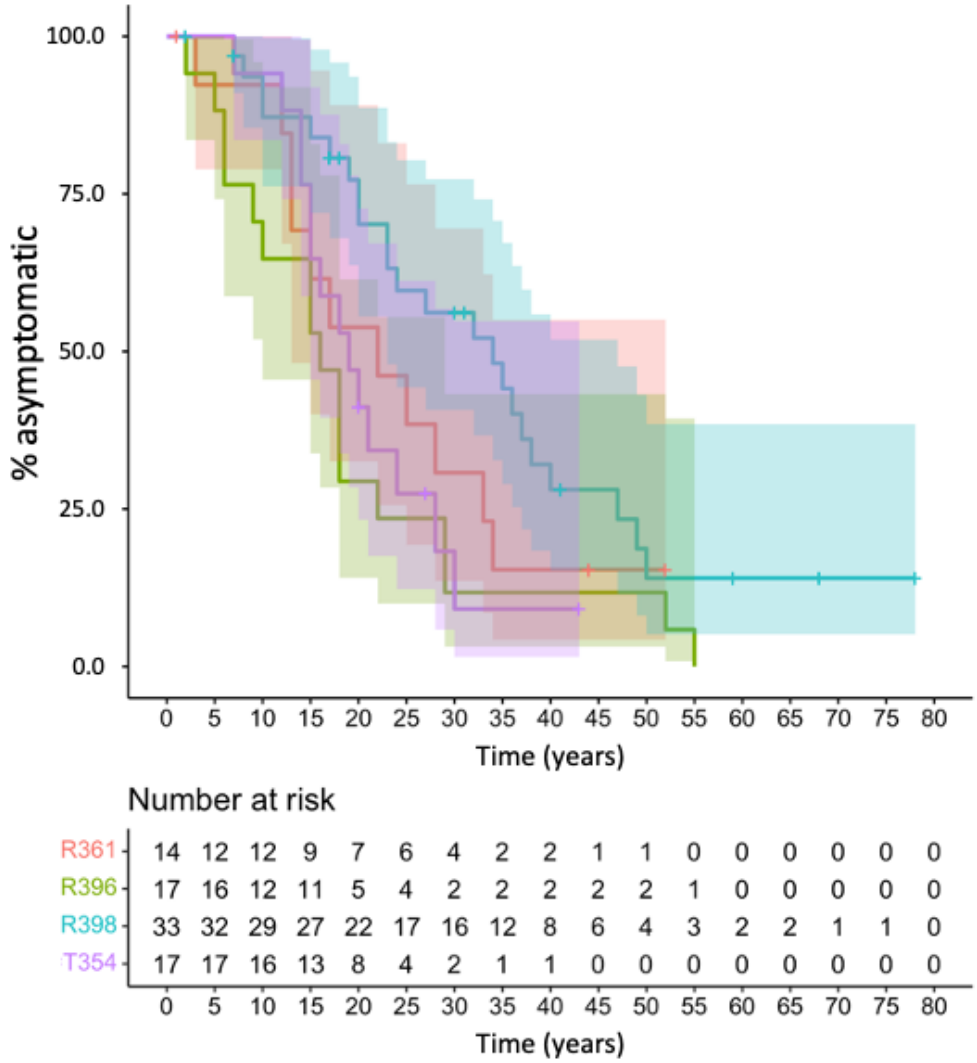
Kaplan-Meier plot of symptom onset across mutation groups. Shading represents 95% confidence intervals. Table below specifies number of at-risk individuals at each 5 year age increment.

Supplemental Figure 6



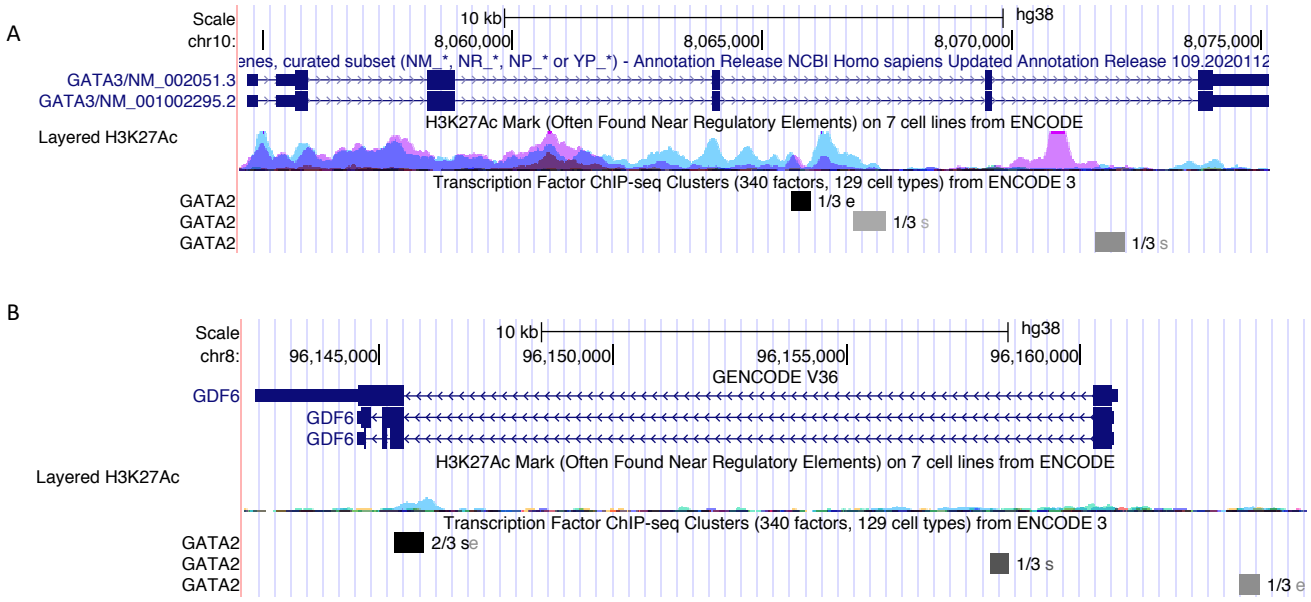
Individuals with mutations at recurrently mutated amino acids. Patients with mutations at one of 4 recurrently mutated amino acids, T354, R361, R396, and R398 are shown, separated by those having symptoms consistent with GATA2 deficiency, (symptomatic – Sx) or remaining asymptomatic (Asx). Age is age of symptom onset for symptomatic patients or age at last evaluation (Asx).

Supplemental Figure 7



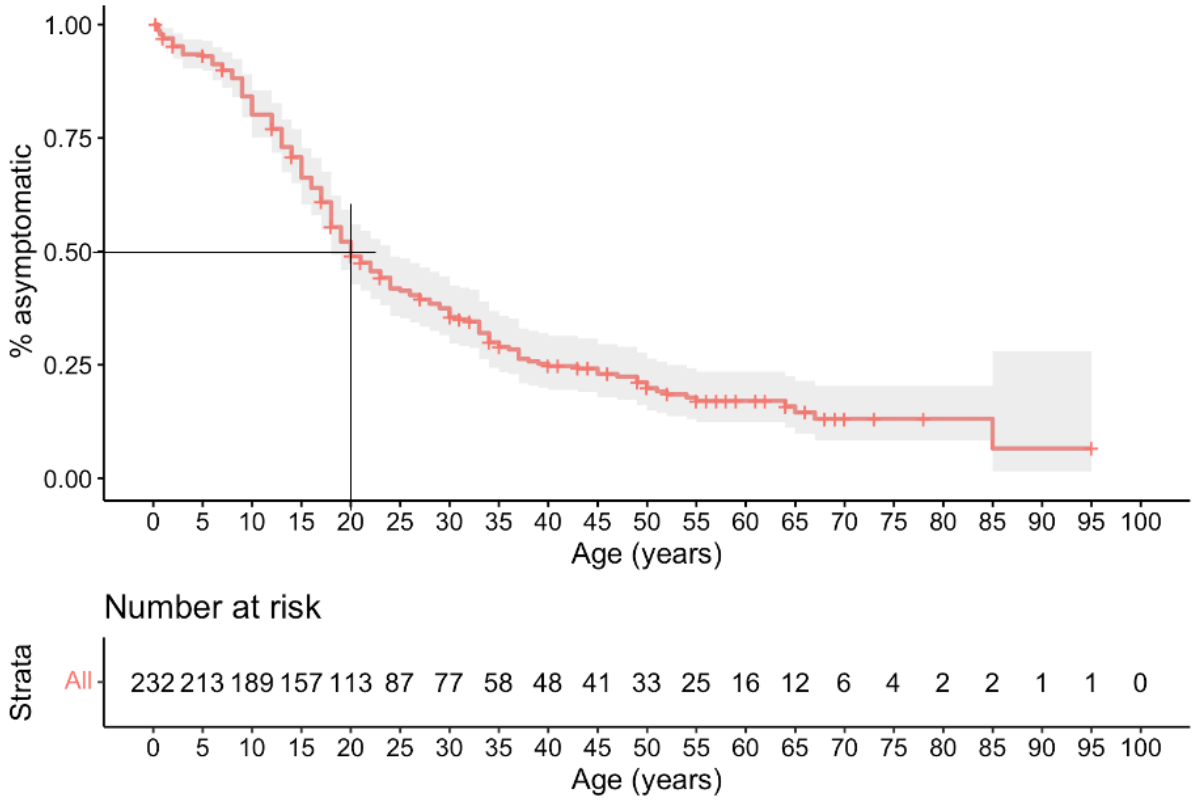
Kaplan-Meier plot of symptom onset by recurrently mutated amino acid. Plot represents individuals with mutations at one of the 4 recurrently mutated, 2nd ZF amino acids, shading indicates 95% confidence interval. Table below specifies number of at-risk individuals at each 5-year age bracket.

Supplemental Figure 8



GATA2 transcriptionally regulates genes critical for inner ear development. A. GATA2 binds to intronic elements within the *GATA3* locus. Shown are two transcripts of GATA3 reflecting exons (boxes) and introns (line with arrows indicating direction of transcription. H3K27 marks are taken from the ENCODE data set (www.encode.org). The binding sites are experimentally determined as part of the ENCODE data set and occur within a region of H3K27 histone acetylation, a marker of regulatory elements. B. GATA2 binds to an intronic element within *GDF6*. Similar to A, there are multiple recognized *GDF6* transcripts, GATA2 binding sites are experimentally determined as part of the ENCODE data set.

Supplemental Figure 9



Median age of symptom onset across entire cohort is (n = 232) is 20 years.

Supplemental Table 1. Data table for all patients included in the study documenting mutation, major symptoms, age at initial symptom onset, bone marrow findings, and current or latest status of patients. This is a separate Excel file.

Supplemental Table 2. Patient status at last encounter

Status	Number	Age last encounter	Median Age
Alive	104	1 - 95	39.0
Alive / Transplanted	91	10-60	28.00
Deceased	55	14-90	31.00

Supplemental Table 3. Pairwise comparison of mutation categories

Mutation category comparison	**HR**	**p.value (Adjusted)**	**p.value (Unadjusted)**	**Lower CI**	**Upper CI**
Truncation / Enhancer	5.635	**0.000155**	0.0000159	1.889	16.807
Null / Enhancer	3.547	**0.00114**	0.000121	1.445	8.707
Truncation / 2nd ZF	2.52	**0.0487**	0.00619	1.003	6.329
2nd ZF / Enhancer	2.236	**0.0497**	0.00634	1.001	4.998
Null / 2nd ZF	1.586	0.36	0.0684	0.795	3.163
Truncation / C-term	2.531	0.473	0.101	0.539	11.875
C-term / Enhancer	2.226	0.576	0.139	0.509	9.74
Truncation / Null	1.589	0.718	0.209	0.581	4.341
Null / C-term	1.593	0.896	0.367	0.389	6.516
2nd ZF / C-term	1.004	1	0.993	0.259	3.895

Supplemental Table 4. Cytogenetic abnormalities identified by sex

Cytogenetics	Males (n = 92)	Females (n = 111)
-7	11	4
-7, +8	4	1
+8, +1 der(1;7)(q10;p10)	2	1
Chr 7 abnormalities	17	6
+8	4	19
+8, +1(der)	1	0
+8, -11	0	1
Chr 8 abnormalities	5	20
dup 1q	0	1
t(3;18)(q26;q23)	0	1
del 13q	0	1
del Y	1	0

Supplemental Table 5. Frequency of infections and cytopenias

Symptomatic (n = 183)	**Presenting (n; %)**	**Occurrence (n; %)**
Infection	81; 44.3	145; 79.2
Cytopenia	47; 25.7	161; 88.0
Viral		114; 62.3
Genital/ extragenital HPV		94; 51.4
EBV		22; 12.0
Bacterial		74; 40.4
NTM		53; 29.0
staphylococcal		5; 2.7
streptococcal		5; 2.7
Fungal		44; 24.0
Candida		18; 9.8
Aspergillus		12; 6.6

Incidence of infection or cytopenias as presenting symptoms or total occurrence among 183 symptomatic GATA2 patients. HPV – human papilloma virus; EBV – Ebstein Barr virus; NTM – nontuberculous *Mycobacteria*.

Supplemental Table 6. Bone marrow abnormalities and classifications

Bone Marrow (n = 167)	**Occurrence (n; %)**
Any abnormality	142; 85.0
MDS	102; 61.1
In-house review	116; 69.5
WHO MDS	61; 52.6
MDS-MLD	43
MDS-EB1	4
MDS-SLD	4
MDS-MPN	3
MDS-U	2
RCC	2
CMML	2
MDS-EB2	1

Bone marrow findings from clinical records of 167 patients; 116 bone marrows were reviewed by a hematopathologist, 61 patients met the WHO criteria for MDS. Identified subtypes are listed in order of prevalence. MDS – myelodysplastic syndrome; WHO MDS – MDS according to World Health Organization criteria; MDS-MLD – MDS with multilineage dysplasia; MDS-EB1 – MDS with excess blasts; MDS-SLD – MDS with single lineage dysplasia; MDS-MPN – myelodysplastic/myeloproliferative neoplasm; MDS-U – unclassifiable MDS; RCC – refractory cytopenia in childhood; CMML – chronic myelomonocytic leukemia; MDS-EB2 – MDS with excess blasts.