REGULAR ARTICLE

blood advances

Non-del(5q) myelodysplastic syndromes–associated loci detected by SNP-array genome-wide association meta-analysis

Kathy L. McGraw,[1] Chia-Ho Cheng,[2] Y. Ann Chen,[2] Hsin-An Hou,[3] Björn Nilsson,[4] Giulio Genovese,[5-7] Thomas Cluzeau,[8] Andrea Pellagatti,[9] Bartlomiej P. Przychodzen,[10] Mar Mallo,[11] Leonor Arenillas,[12] Azim Mohamedali,[13] Lionel Adès,[14] David A. Sallman,[1] Eric Padron,[1] Lubomir Sokol,[1] Chimene Moreilhon,[8] Sophie Raynaud,[8] Hwei-Fang Tien,[3] Jacqueline Boultwood,[9] Benjamin L. Ebert,[15] Francesc Sole,[11] Pierre Fenaux,[14] Ghulam J. Mufti,[13] Jaroslaw P. Maciejewski,[10] Peter A. Kanetsky,[16] and Alan F. List[1]

[1]Department of Malignant Hematology and [2]Department of Biostatistics and Bioinformatics, Moffitt Cancer Center, Tampa, FL; [3]Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; [4]Department of Laboratory Medicine, Section of Hematology and Transfusion Medicine, Lund University, Lund, Sweden; [5]Stanley Center for Psychiatric Research and [6]Program in Medical and Population Genetics, Broad Institute of the Massachusetts Institute of Technology and Harvard, Cambridge, MA; [7]Department of Genetics, Harvard Medical School, Boston, MA; [8]Cote d'Azur University, Centre Hospitalier Universitaire of Nice, Nice, France; [9]Bloodwise Molecular Haematology Unit, Nuffield Division of Clinical Laboratory Sciences, Radcliffe Department of Medicine, University of Oxford and Oxford Biomedical Research Centre Haematology Theme, Oxford, United Kingdom; [10]Cleveland Clinic, Taussig Cancer Institute, Cleveland, OH; [11]Institut de Recerca Contra la Leucèmia Josep Carreras, Institut Catala d'Oncologia–Hospital GermansTrias i Pujol, Universitat Autonòma de Barcelona, Badalona, Barcelona, Spain; [12]Laboratori de Citologia Hematòlogica, Servei de Patologia, Hospital del Mar, Barcelona, Spain; [13]Department of Haematological Medicine, Kings College London, London, United Kingdom; [14]Senior Hematology Department, Hopital Saint Louis, Paris, France; [15]Departments of Medicine and Medical Oncology, Harvard Medical School and Brigham and Women's Hospital, Boston, MA; and [16]Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL

Key Points

- We report here the findings from the first known MDS genome-wide association study and meta-analysis identifying 8 unique loci.
- Genes harboring suggested MDS-associated loci, including *EYA2*, are innate immune regulators and may have clinical and biological relevance.

Myelodysplastic syndromes (MDS) are hematopoietic stem cell malignancies. Known predisposing factors to adult MDS include rare germline mutations, cytotoxic therapy, age-related clonal hematopoiesis, and autoimmune or chronic inflammatory disorders. To date, no published studies characterizing MDS-associated germline susceptibility polymorphisms exist. We performed a genome-wide association study of 2 sample sets (555 MDS cases vs 2964 control subjects; 352 MDS cases vs 2640 control subjects) in non-del(5q) MDS cases of European genomic ancestry. Meta-analysis identified 8 MDS-associated loci at 1q31.1 (*PLA2G4A*), 3p14.1 (*FAM19A4*), 5q21.3 (*EFNA5*), 6p21.33, 10q23.1 (*GRID1*), 12q24.32, 15q26.1, and 20q13.12 (*EYA2*) that approached genome-wide significance. Gene expression for 5 loci that mapped within or near genes was significantly upregulated in MDS bone marrow cells compared with those of control subjects ($P < .01$). Higher *PLA2G4A* expression and lower *EYA2* expression were associated with poorer overall survival ($P = .039$ and $P = .037$, respectively). Higher *PLA2G4A* expression is associated with mutations in *NRAS* ($P < .001$), *RUNX1* ($P = .012$), *ASXL1* ($P = .007$), and *EZH2* ($P = .038$), all of which are known to contribute to MDS development. *EYA2* expression was an independently favorable risk factor irrespective of age, sex, and Revised International Scoring System score (relative risk, 0.67; $P = .048$). Notably, these genes have regulatory roles in innate immunity, a critical driver of MDS pathogenesis. EYA2 overexpression induced innate immune activation, whereas EYA2 inhibition restored colony-forming potential in primary MDS cells indicative of hematopoietic restoration and possible clinical relevance. In conclusion, among 8 suggestive MDS-associated loci, 5 map to genes upregulated in MDS with functional roles in innate immunity and potential biological relevance to MDS.

Submitted 30 August 2019; accepted 15 October 2019. DOI 10.1182/bloodadvances.2019000922.

Data-sharing requests should be forwarded to Kathy McGraw (kathy.mcgraw@moffitt.org).

The full-text version of this article contains a data supplement.

Introduction

Myelodysplastic syndromes (MDS) are senescence-dependent stem cell malignancies with ineffective hematopoiesis, cytological dysplasia, and a propensity for progression to acute myeloid leukemia (AML). Nongenetic predisposing factors include previous chemotherapy, autoimmune or chronic inflammatory disorders, and age-related clonal hematopoiesis.[1,2] Our knowledge of inherited susceptibility to MDS is limited. Familial MDS is more common among younger patients and associates with other bone marrow failure syndromes, including Fanconi anemia, dyskeratosis congenita, and Shwachman-Diamond syndrome.[3] Germline mutations in the *RUNX1*, *ANKRD26*, *DDX41, CEBPA*, *ETV6*, *SRP72, TERT*, or *GATA2* genes are found in a portion of, but not all, familial cases. To date, there are no known common susceptibility loci for MDS. Single nucleotide polymorphism-array (SNP-A) genotyping can identify genetic abnormalities not resolved by classical metaphase cytogenetics or florescent in situ hybridization, as well as germline polymorphisms.[4,5] However, to our knowledge, there are no published studies to date with satisfactory power to characterize possible germline polymorphisms associated with MDS susceptibility. Here, we provide results of the first genome-wide association study (GWAS) using a meta-analysis approach to identify MDS-associated genetic susceptibility loci.

Materials and methods

Samples

Genotype data sources and institutions are summarized in supplemental Table 1. Each contributing center locally performed genotyping, and our analysis was restricted to cases genotyped on the Affymetrix Genome-Wide Human SNP Array 6.0 platform. In total, we securely received 1363 Affymetrix raw CEL files for processing. No other clinical or phenotypic data were available. Comparable control samples for performing association analyses were obtained from 3 control series (discussed in the following sections) from dbGaP (https://www.ncbi.nlm.nih.gov/gap).

GENEVA Diabetes Study (NHS/HPFS) The study performed genome-wide association analysis from the Nurses' Health Study (NHS) and the Health Professionals' Follow-up Study (HPFS) to identify genetic type 2 diabetes susceptibility loci. This study was part of the Gene Environment Association Studies initiative (GENEVA; http://www.genevastudy.org) funded by the trans-National Institute of Health Genes, Environment, and Health Initiative. We extracted data from individuals in data sets version 1 and 2 (n = 3149) as control subjects. The age (mean ± standard deviation) of these control subjects was 57.09 ± 7.66 years.

GAIN: Genome-Wide Association Study of Schizophrenia This is a case-control study aimed at identifying schizophrenia susceptibility loci. Data were collected under the Genetic Association Information Network (GAIN) at the Broad Institute. For our analyses, individuals of European descent from data set version 2 (n = 1378) were used. The mean age of these control subjects was 51.19 ± 16.98 years.

Molecular Genetics of Schizophrenia, non-GAIN This study was part of the Molecular Genetics of Schizophrenia GWAS, performed at the Broad Institute, and case and control subject data were not collected under the GAIN (non-GAIN). We used data from 1297 non-GAIN control subjects for the current study. The mean age of these control subjects was 49.82 ± 15.74 years.

Genotyping, quality control, case, and marker restriction

To reduce potential batch effects across data sets, raw genotyping files for all case and control subjects were combined, and genotypes were simultaneously called by using Birdseed v2. Quality control (QC) at both the sample- and marker-level was performed. We excluded markers with genotype missingness surpassing 5%, minor allele frequency ≤0.01, and those not in Hardy-Weinberg equilibrium ($P \leq 1 \times 10^{-12}$). For sample-level quality control, related individuals were excluded based on an identity-by-descent score >0.1875. Samples with >5% of genotyping missingness or extreme heterozygosity (<5% or >40%) were excluded. Analysis was restricted to individuals of European ancestry, determined by principal component analysis using EIGENSTRAT (supplemental Figure 1A). Analysis was restricted to non-del(5q) cases determined by copy number variant evaluation using Affymetrix Chromosome Analysis Suite as previously described.[6] For samples in the final analytic data set that passed quality control and restriction filters, principal components were reestimated for use as covariates in logistic regression models. The total number of case and control subjects passing quality control is provided in Table 1.

Imputation

Genotype data passing quality control as described earlier was submitted to the Michigan Imputation Server (https://imputation-server.sph.umich.edu/index.html) for imputation using the resources of the Haplotype Reference Consortium.[7] We received imputed genotypes on 39 117 104 markers and analyzed 23 278 269 with adequate imputation quality indicated by an info score >0.4. Imputed SNPs with a minor allele frequency <0.01 were excluded from further analysis, resulting in 7 515 136 markers in 6512 samples.

Genome-wide association using a meta-analysis approach

To address heterogeneity arising from multiple sources of case samples and genotype data, and supported by the limited analytic benefit of pooled vs summary analytic approaches,[8,9] we opted to partition MDS cases into 2 analytic subsets. Subset 1 consisted solely of MDS cases from the Cleveland Clinic (n = 555 after QC), which represented >60% of all contributed cases. For these cases, control subjects were selected from the GENEVA data set as the comparison group (N = 2965 after QC). Subset 2 consisted of all other MDS cases (n = 352 after QC) aggregated across the remaining 5 contributing centers; control subjects from both the GAIN and non-GAIN data sets served as a comparison group (N = 2640 after QC) for these cases.

Logistic regression was conducted separately for Subsets 1 and 2 to determine marker associations with MDS disease status, assuming an additive genetic model. We adjusted for genetically determined sex and the first 4 principal components. For Subset 1, the genomic inflation factor, λ, was 1.05; for Subset 2, λ was 1.00 (supplemental Figure 1B-C). Summary odds ratios (ORs) and corresponding 95% CIs were estimated by using fixed effects

Table 1. Summary of MDS associated poymorphisms

Cytoband	Gene neighborhood	SNP	Position	Group	Controls, n	Cases, n	Minor allele	Major allele	MAF case	MAF control	OR (95% CI)	*P*	P_{het}	I^2
1q31.1	PLA2G4A	rs6683416	1.87E+08	Sample set 1	2965	555	A	C	0.20	0.25	0.74 (0.63-0.86)	1.87E-04	.94	0
				Sample set 2	2640	352			0.23	0.25	0.73 (0.57-0.93)	1.03E-02		
				Combined	5605	907					0.73 (0.64-0.84)	5.88E-06		
3p14.1	FAM19A4	rs34539210	68959905	Sample set 1	2965	555	G	T	0.20	0.24	0.72 (0.61-0.85)	1.17E-04	1.00	0
				Sample set 2	2640	352			0.21	0.23	0.72 (0.56-0.93)	1.19E-02		
				Combined	5605	907					0.72 (0.63-0.83)	4.18E-06		
5q21.3	EFNA5	rs341274	1.07E+08	Sample set 1	2965	555	A	G	0.27	0.22	1.31 (1.13-1.53)	3.75E-04	.50	0
				Sample set 2	2640	352			0.25	0.23	1.44 (1.14-1.84)	2.42E-03		
				Combined	5605	907					1.35 (1.19-1.53)	3.78E-06		
6p21.33	NA	rs1634783	31283604	Sample set 1	2965	555	A	G	0.45	0.50	0.78 (0.68-0.89)	2.76E-04	.33	0
				Sample set 2	2640	352			0.43	0.50	0.69 (0.56-0.85)	5.33E-04		
				Combined	5605	907					0.76 (0.67-0.84)	8.46E-07		
10q23.1	GRID1	rs7099032	87632795	Sample set 1	2965	555	C	A	0.20	0.25	0.73 (0.62-0.85)	9.63E-05	.76	0
				Sample set 2	2640	352			0.20	0.24	0.76 (0.59-0.98)	3.26E-02		
				Combined	5605	907					0.74 (0.64-0.84)	9.14E-06		
12q24.32	NA	rs2947170	1.28E+08	Sample set 1	2965	555	T	A	0.12	0.07	1.61 (1.29-2.00)	1.73E-05	.61	0
				Sample set 2	2640	352			0.08	0.07	1.80 (1.25-2.57)	1.47E-03		
				Combined	5605	907					1.65 (1.37-1.99)	1.03E-07		
15q26.1	NA	rs4404050	92136977	Sample set 1	2965	555	T	C	0.37	0.33	1.27 (1.10-1.46)	8.62E-04	.39	0
				Sample set 2	2640	352			0.39	0.32	1.42 (1.15-1.76)	1.41E-03		
				Combined	5605	907					1.31 (1.17-1.47	5.85E-06		
20q13.12	EYA2	rs1206818	45705792	Sample set 1	2965	555	A	C	0.28	0.23	1.36 (1.18-1.58)	4.12E-05	.40	0
				Sample set 2	2640	352			0.30	0.23	1.53 (1.22-1.93)	2.96E-04		
				Combined	5605	907					1.41 (1.24-1.59)	6.66E-08		

CI, confidence interval; het, heterogeneity; MAF, minor allele frequency.

meta-analysis implemented in PLINK. SNP markers brought forward for further scrutiny included those with the following: (1) evidence of good genotype clustering; (2) low heterogeneity (I^2 <35) across the 2 subsets; (3) $P \leq 1.0 \times 10^{-6}$; and (4) a regional plot with linkage disequilibrium patterning of proximal SNPs suggestive of a true positive hit (supplemental Figure 2). The meta-analysis Q-Q plot ($\lambda = 0.99$) is provided in supplemental Figure 1D. Q-Q and Manhattan plots were generated by using R; regional plots were generated by using LocusZoom.[10]

Gene expression profiling

Gene expression analysis was conducted on data from 3 sources. The first was the publicly available National Center for Biotechnology Information Gene Expression Omnibus (GEO) data set,[11] GDS3795, which includes gene expression profiling on $CD34^+$ cells from 183 MDS cases and 17 healthy donors. The second data set was unsorted cells from the National Taiwan University Hospital (NTUH), including 295 MDS cases (213 fulfilled the criteria of the World Health Organization [WHO] classification), 300 AML cases, and 20 healthy bone marrow transplant donors profiled with the HumanHT-12 v4 Expression BeadChip (Illumina Inc., San Diego, CA).[12] The NTUH study was approved by the institutional review board of the NTUH (NTUH 20150709RINA), with written informed consent obtained from all participants in accordance with the Declaration of Helsinki. Cohort demographic characteristics are provided in supplemental Table 2. We also used BloodSpot, a database of gene expression profiles and transcriptional programs for healthy and malignant hematopoiesis (http://servers.binf.ku.dk/bloodspot/), to compare gene expression in normal hematopoiesis.[13] Analysis was restricted to the HemaExplorer database and only cell populations that may be most relevant to MDS, specifically, stem, progenitor, and myeloid populations. Diagrams represent probe sets with highest mean expression.

Gene expression was analyzed either by using the independent Student *t* test (National Center for Biotechnology Information database) or the Mann-Whitney *U* test (NTUH cohort) to compare continuous variables and medians of distributions, respectively. Overall survival was measured from the date of diagnosis to date of death from any cause or date of last follow-up. Survival curves were estimated by using the Kaplan-Meier method, and comparisons based on gene expression levels were tested by using the log-rank test. $P < .05$ was considered statistically significant. Statistical analyses were performed by using SPSS 18 software (IBM Corporation, Armonk, NY) and StatsDirect (Cheshire, United Kingdom).

Expression quantitative trait loci lookup

To investigate the potential functional mechanism of our putative susceptibility loci, we examined known expression quantitative trait loci (eQTLs) in Blood eQTL and Genotype-Tissue Expression (*GTEx*) portals (https://www.gtexportal.org/home) for these SNPs and proxies. Both cis- and trans-eQTL databases were downloaded from Blood eQTL (https://genenetwork.nl/bloodeqtlbrowser).[14] Data used for this analyses were obtained from the GTEx Portal on August 15, 2018. Proxies were identified in PLINK by using linkage disequilibrium $r^2 > 0.4$ and 1000 kb range from all SNPs imputed or genotyped from our 2 studies (supplemental Table 3).

Transfection and western blotting

THP-1 cells were stably transfected with pcDNA3.1 Flag-Eya2 vector purchased from Addgene (plasmid #49264; Cambridge, MA) using an Amaxa Nucleofector (Lonza, Walkersville, MD) with the manufacturer's THP-1–optimized transfection protocol. Cells were selected by using 320 μg/mL of Hygromycin B Gold and harvested for western blotting, performed as previously described.[15]

Colony-forming capacity assay

Bone marrow mononuclear cells were isolated by using Ficoll-Paque (GE Healthcare, Pittsburgh, PA). Five hundred thousand cells per milliliter were plated in MethoCult (Stemcell Technologies, Vancouver, BC, Canada) with complete cytokines supporting myeloid and erythroid colony growth with 1% penicillin/streptomycin, and 3% final volume autologous bone marrow plasma. One milliliter was plated per well ± the EYA2 inhibitor (MLS000544460) and incubated at 37°C in a humidified chamber with 5% carbon dioxide for 14 days, when the number of erythroid (burst forming unit-erythroid and colony forming unit-erythroid) and myeloid (colony-forming unit-granulocyte, macrophage) were counted. The number of replicates is indicated, and averages are represented as means ± standard error.

Results

Meta-analysis of our 2 case-control subsets (supplemental Table 1) identified 8 independent loci that were suggestively associated with non-del(5q) MDS ($P \leq 1.0 \times 10^{-6}$ for each); no genetic markers reaching classical genome-wide significance were identified ($P \leq 5.0 \times 10^{-8}$) (Figure 1; Table 1; supplemental Figure 2). Five of these loci were located within or proximal to genes on 1q31.1 (rs6683416, *PLA2G4A*), 3p14.1 (rs34539210, *FAM19A4*), 5q21.3 (rs341274, *EFNA5*), 10q23.1 (rs7099032, *GRID1*), and 20q13.12 (rs1206818, *EYA2*); the remaining 3 loci (rs1634783 at 6p21.33, rs2947170 at 12q24.32, and rs4404050 at 15q26.1) were located in intergenic regions.

We conducted in silico analyses (including eQTL analysis), performed a literature review of published GWAS, and searched for functional activity on all 8 MDS-associated loci and proxies (supplemental Table 2) and the 5 genes housing these loci. Gene expression analysis was also performed on the 5 relevant genes. In vitro functional analysis was additionally performed on *EYA2*, the gene housing the most statistically significant loci (rs1206818; $P = 6.66 \times 10^{-8}$). A study flow diagram is provided in supplemental Figure 3, and a summary of eQTL and gene expression findings for each loci is given in supplemental Table 4. Highlighted here are results from our in silico and in vitro analyses for each locus in order of chromosome number.

Rs6683416 (OR, 0.7341; $P = 5.88 \times 10^{-6}$) maps within an intron of Phospholipase A2 Group IVA *(PLA2G4A)*. In normal hematopoietic cells, *PLA2G4A* expression is highest in common myeloid progenitors and megakaryocyte-erythroid progenitors (supplemental Figure 4A,F; supplemental Table 5). Gene expression levels of *PLA2G4A* were significantly higher in $CD34^+$ selected MDS bone marrow cells (GDS3795) compared with those of healthy control subjects (level; $P = 8.77 \times 10^{-5}$) (Figure 2A). Similarly, *PLA2G4A* expression was significantly increased in WHO-defined MDS unsorted bone marrow cells of patients compared with those of

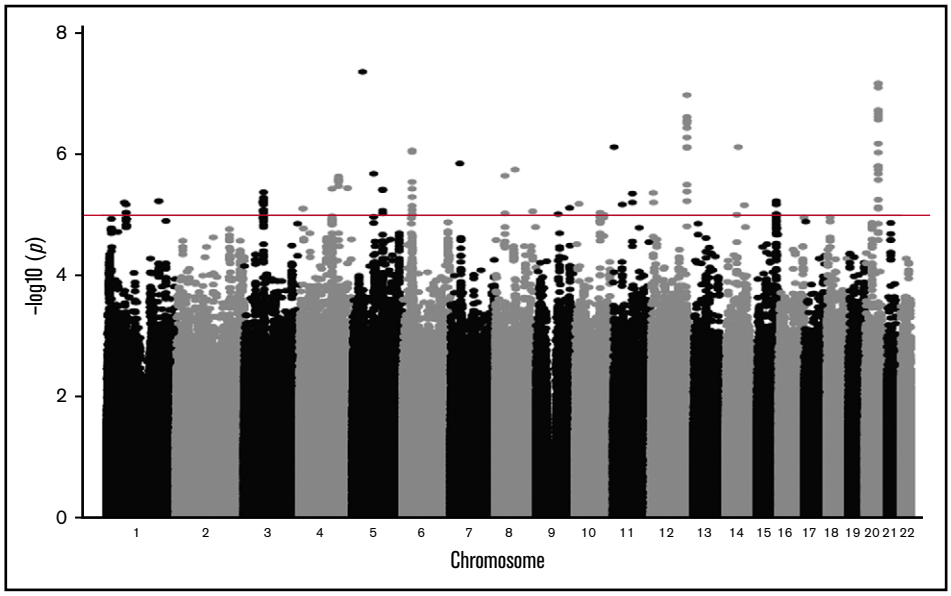

Figure 1. Manhattan plot. The Manhattan plot is an overview of markers with corresponding *P* values of our meta-analysis. Hits with a *P* value $< 1 \times 10^{-6}$ (red line) were investigated further.

healthy donors (NTUH, level; $P < .001$). Interestingly, *PLA2G4A* expression levels were significantly higher in unsorted bone marrow cells from AML cases than in either MDS cases (levels; $P < .001$) or healthy donors (levels; $P < .001$) (Figure 2B). Furthermore, expression increased with higher risk MDS subtypes, including refractory anemia with excess blasts-1 (RAEB1), RAEB2, and RAEB in transformation (-T) (Figure 2C). Higher *PLA2G4* expression was more often associated with cases harboring somatic mutations in *NRAS* ($P < .001$), *RUNX1* ($P = .012$), *ASXL1* ($P = .007$), or *EZH2* ($P = .038$) genes, well-described MDS mutations with prognostic relevance. Cases with higher *PLA2G4A* expression also had significantly shorter overall survival (OS) compared with cases with lower expression ($P = .039$) (Figure 2D). GTEx eQTL browser analysis revealed several known whole-blood eQTL *PLA2G4A* (supplemental Table 6). In our data set, these eQTL loci have meta-analysis *P* values that range from .10 to .36.

Rs34539210 (OR, 0.72; $P = 4.18 \times 10^{-6}$) maps proximal to the Family With Sequence Similarity 19 Member A4 gene (*FAM19A4*). In normal hematopoietic cells, *FAM19A4* is more highly expressed in bone marrow promyelocytes and myelocytes vs immature myeloid precursors (supplemental Figure 4 B,F; supplemental Table 5). GDS3795 GEO data exhibit little to no gene expression in primary MDS $CD34^+$ bone marrow cells. However, examination of unsorted cells (NTUH) revealed significantly increased expression of *FAM19A4* in WHO-defined MDS compared with both healthy donors and AML cases (both, $P < .0001$) (Figure 3A). Although we observed no differences in expression based on subtype, increased expression was associated with *KRAS* gene mutations ($P = .045$). Blood eQTL analysis identified several proxies to rs34539210 that are known cis-eQTLs (supplemental Table 7) for *ARL6IP5,* ADP Ribosylation Factor Like GTPase 6 Interacting Protein 5. ARL6IP5 (also known as *JWA*) is significantly upregulated in MDS $CD34^+$ cells compared with those from healthy donors ($P < .001$) (supplemental Table 8).

Rs341274 (OR, 1.35; $P = 3.78 \times 10^{-6}$) is located on chromosome 5 within an intron of the Ephrin A5 gene (*EFNA5*). In normal hematopoietic cells, *EFNA5* is more highly expressed in bone marrow myelocytes compared with any more immature cell types (supplemental Figure 4 C,F; supplemental Table 5). According to GDS3795, there is relatively little or no expression in MDS $CD34^+$ cells. In the NTUH data, there was significantly higher expression in unsorted bone marrow cells from WHO-defined MDS cases compared with those from AML cases and from heathy control subjects (both, $P < .0001$) (Figure 3B) in nonsorted bone marrow cells. We found no association, however, between the expression of *EFNA5* with MDS subtype, specific somatic gene mutations, or OS.

Rs1634783 (OR, 0.76; $P = 8.46 \times 10^{-7}$) maps to chromosome 6 (6p21.33) and is not associated with known genes. However, Blood eQTL analysis revealed that rs1634783 proxies were known cis-eQTL to several genes, including *IER3*, *HLA-B*, *HLA-C*, and *C6orf48* (supplemental Table 7). Furthermore, GTEx eQTL analysis revealed several eQTL for both rs1634783 itself and several proxies, the genes of which are listed in supplemental Table 9 and include *CCHCR1*, *HCG27*, *HLA-C*, *HLA-S*, *NOTCH4*, *POU5F1*, *PSORS1C3*, *XXbac-BPG181B23.7*, *XXbac-BPG248L24.12*, and *XXbac-BPG299F13.14*. Per GDS3795, *HCG27*gene expression was significantly higher in MDS cases vs healthy donors ($P < .01$), and *CCHR1* and *C6orf348* approached significance ($P = .11$ and $P = .09$, respectively) (supplemental Table 8). From NTUH data, *CCHCR1*, *HLA-C*, *NOTCH4*, *POU5F1*, and *PSORS1C3* have increased expression in MDS vs healthy donors. In contrast, MDS unsorted bone marrow cells have lower *HCG27* and *IER3* expression compared with healthy donors. There was no difference in *HLA-B* and *C6orf48* expression between unsorted MDS and healthy donor cells (supplemental Table 10).

Rs7099032 (OR, 0.74; $P = 9.14 \times 10^{-6}$) maps to chromosome 10 within an intron of Glutamate Ionotropic Receptor Delta Type Subunit 1 (*GRID1*). In normal hematopoietic cells, *GRID1* is more highly expressed in bone marrow myelocytes vs more immature hematopoietic cells (supplemental Figure 4D,F; supplemental Table 5). Consistent with this maturation-restricted expression pattern, there was little gene expression in primitive $CD34^+$ cells in GDS3795 and no differences in expression between MDS and

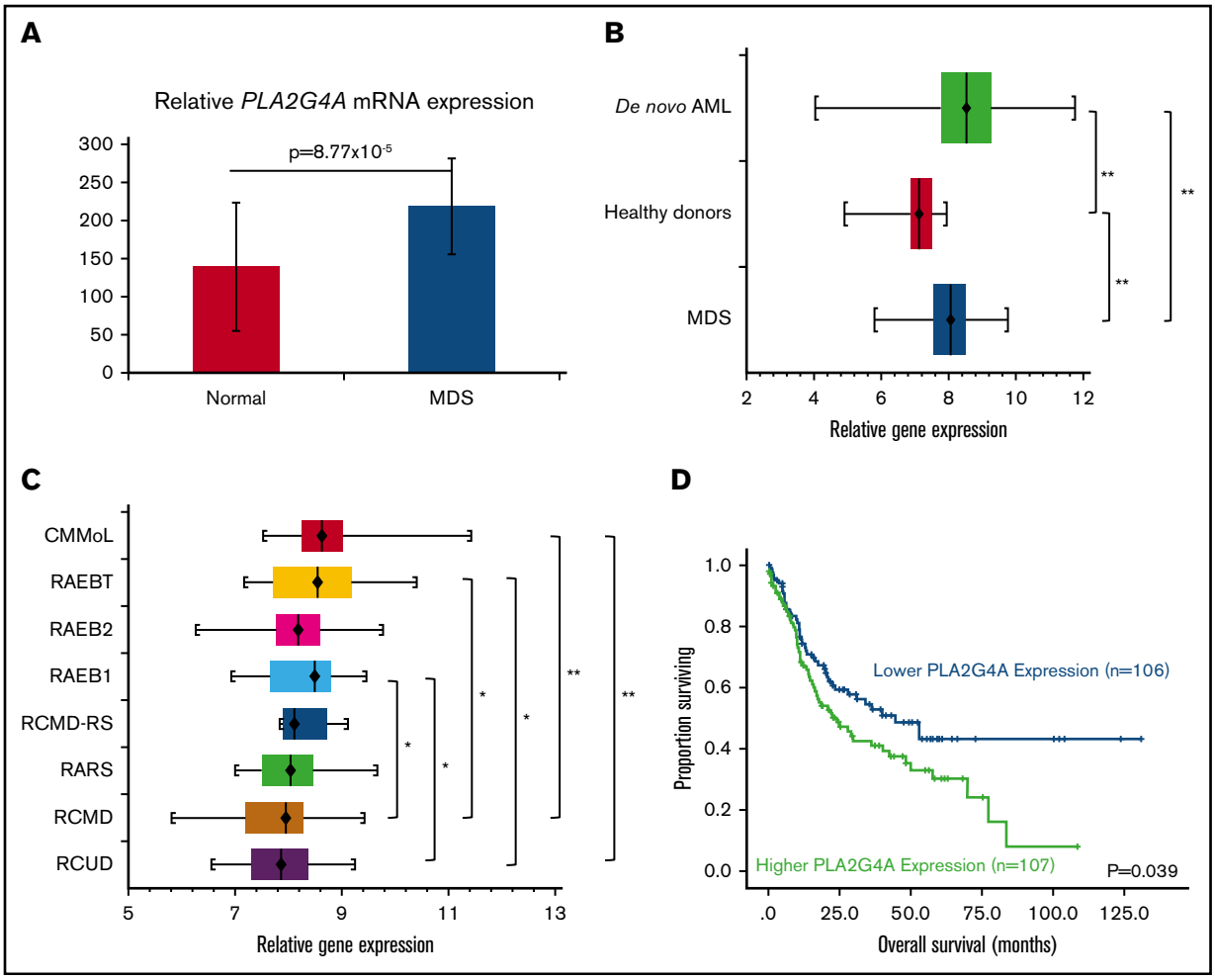

Figure 2. *PLA2G4A* expression in MDS and impact on survival. (A) Relative *PLA2G4A* gene expression using the National Center for Biotechnology Information GEO data set, GDS3795 showing significantly higher gene expression in MDS case subjects compared with control subjects. (B) Relative *PLA2G4A* gene expression using the NTUH independent data set confirming significantly higher *PLA2G4A* gene expression in MDS case subjects compared with control subjects. (C) Higher *PLA2G4A* gene expression is associated with higher risk disease subtypes in RAEB1, RAEB2, and RAEB-T. (D) Kaplan-Meier plot showing that lower *PLA2G4A* gene expression is associated with improved overall survival. Error bars represent mean ± standard error. *$P < .005$, **$P < .0001$.

normal control subjects. However, in nonsorted bone marrow cells from the NTUH data set, there was significantly greater expression in WHO-defined MDS cases compared with both healthy control subjects and AML cases (both, $P < .0001$) (Figure 3C). There was no difference in expression according to MDS subtype, mutation type, or OS.

Rs1206818 (OR, 1.41; $P = 6.66 \times 10^{-8}$) was the most highly significant hit in our meta-analysis and maps to chromosome 20 within an intron of the Eyes absent homolog 2 (*EYA2*). In normal hematopoietic cells, *EYA2* is more highly expressed in bone marrow myelocytes compared with immature hematopoietic cells (supplemental Figure 4E-F; supplemental Table 5). MDS $CD34^+$ bone marrow cells do not exhibit differential expression vs normal donors in GDS3795; however, nonsorted cells from NTUH display significantly higher *EYA2* expression in MDS cases vs normal donors ($P < .0001$). Interestingly, AML cases have *EYA2* expression far greater than that of MDS cases ($P < .0001$) (Figure 4A) despite higher expression in lower risk MDS subtypes vs higher risk MDS subtypes (Figure 4B). We found no relation between *EYA2* expression and specific somatic gene mutations. However, OS was significantly increased in MDS cases with higher *EYA2* expression ($P = .037$) (Figure 4C). Importantly, *EYA2* was an independently favorable risk factor irrespective of age, sex, and Revised International Prognostic Scoring System score (relative risk, 0.665; $P = .048$). EYA members are sensors of cytosolic DNA, which can serve as a danger-associated molecular pattern (DAMP) to trigger inflammasome activation.[16,17] Based on our previous findings that innate immune activation, specifically NLRP3 inflammasome activation, drives MDS pathogensis,[18] we interrogated whether *EYA2* is functional in MDS. To address this question, we overexpressed *EYA2* in monocytic THP-1 cells. *EYA2* expression induced caspase-1 and pro–IL-1β cleavage and increased phosphorylation of NF–κB, indicating priming and activation of the inflammasome innate immune axis (Figure 4D). We previously found that inhibition of the inflammasome by a small molecule inhibitor, MCC950, restored hematopoiesis in MDS bone marrow progenitors, as evidenced by improved colony-forming capacity.

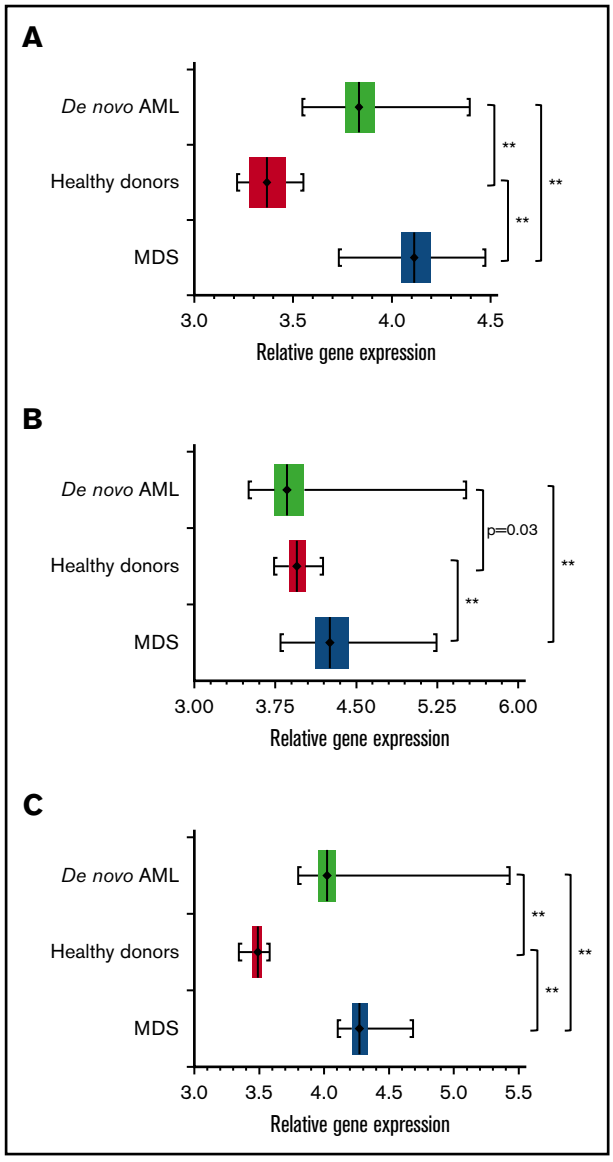

Figure 3. *FAM19A4*, *EFNA5*, and *GRID1* relative gene expression. Relative gene expression of *FAM19A4* (A), *EFNA5* (B), and *GRID1* (C) using the NTUH data set showing significantly higher gene expression in MDS case subjects compared with control subjects. Error bars represent mean ± standard error. $^{**}P < .0001$.

Here, we treated 3 primary MDS bone marrow samples with the EYA2 inhibitor MLS000544460, and scored colony recovery. We observed a significant, concentration-dependent increase in erythroid colony-forming capacity, indicating improved effective hematopoiesis (Figure 4E).

Discussion

Our investigations are the first to identify suggested non-del(5q) MDS-associated germline genetic variations. To our knowledge, no previous MDS GWASs have been reported. We found no overlap between our loci with those reported by Astle et al,[19] who used GWAS to describe blood cell phenotypes, or loci associated with human red blood cell biology as reported by van der Harst et al.[20]

The majority of the 8 loci identified in the current study map in or proximal to *PLA2G4A*, *FAM19A4*, *EFNA5*, *GRID1*, and *EYA2*, with potential functional consequences for MDS pathobiology. Expression of each of these genes was significantly increased in MDS vs healthy donors (NTUH data). Furthermore, higher *PLA2G4A* and *FAM19A4* gene expression was associated with common, myeloid-specific, prognostically adverse, somatic gene mutations, meriting further investigation. Lastly, lower *PLA2G4A* and higher *EYA2* expression was associated with significantly prolonged OS, demonstrating prognostic relevance. Importantly, these genes are involved in innate immune modulation. We showed that somatic gene mutations common to MDS (regardless of class) and other DAMPs direct NLRP3 inflammasome activation and caspase-1–driven cell death (pyroptosis), potentiating the chronically inflamed bone marrow milieu in MDS.[18]

PLA2G4A encodes cytosolic phospholipase A2 (cPLA2) with pleotropic functions, including membrane phospholipid hydrolysis in the presence of free Ca^{2+}.[21] Increased *PLA2G4A* is reported in prostate cancer, AML, and tumor necrosis factor–induced leukemic cell cytotoxicity.[22-24] Gerstung et al[25] investigated differential gene expression associated with common somatic gene mutations and chromosomal abnormalities in bone marrow $CD34^{+}$ cells of patients with MDS. These data showed that *PLA2G4A* gene expression is significantly ($P = .038$) increased in association with *U2AF1* mutations (0.527 increased expression), warranting investigation of the loci identified here with *U2AF1* mutations. Importantly, cPLA2 regulates NLRP3-inflammasome.[18,26] cPLA2α activation decreases inflammasome activation, which is indispensable for pyroptotic cell death and β-catenin activation, suggesting potential disease-related biological relevance. Importantly, iPLA2 is required for S100A9 phosphorylation by MAPK, which is necessary for proinflammatory effects and intracellular scaffold functions.[27] Notably, S100A9 is the major cell-extrinsic and cell-intrinsic inflammasome-activating DAMP in MDS that is overexpressed in both MDS progenitors and the expanded population of medullary myeloid-derived suppressor cells in these patients.[18]

FAM19A4 encodes a cytokine whose function is not fully characterized; however, expression is upregulated in monocytes and macrophages after endotoxin exposure, suggesting a role in inflammatory signaling.[28] *FAM19A4* methylation described in varied solid tumors suggests potential tumor suppressor function,[29,30] which should be further explored in the context of MDS because hypermethylation contributes to signaling aberrancy in these neoplasms. The relation to more mature cell types and *KRAS* mutations identified here should be explored in future studies. Notably, chromosome 3 abnormalities are associated with unfavorable prognosis in MDS, and 3p abnormalities specifically account for ~15% of cases harboring a change in this chromosome.[31] Interrogation of the SNP allele frequency with respect to these abnormalities may have clinical implications.

Although little is known about the function of Efna5 outside of neuronal differentiation, the ephrin receptor–binding pathway has been implicated in gastric adenocarcinoma, implicating a possible role in oncogenesis.[32] Per Gerstung et al,[25] *EFNA5* is significantly ($P = .008$) upregulated in association with del(5q) abnormalities (increased by 0.056), which is the most commonly found chromosomal abnormality observed in MDS and carries distinct clinical and biological features. Furthermore, Efna5 signals in part

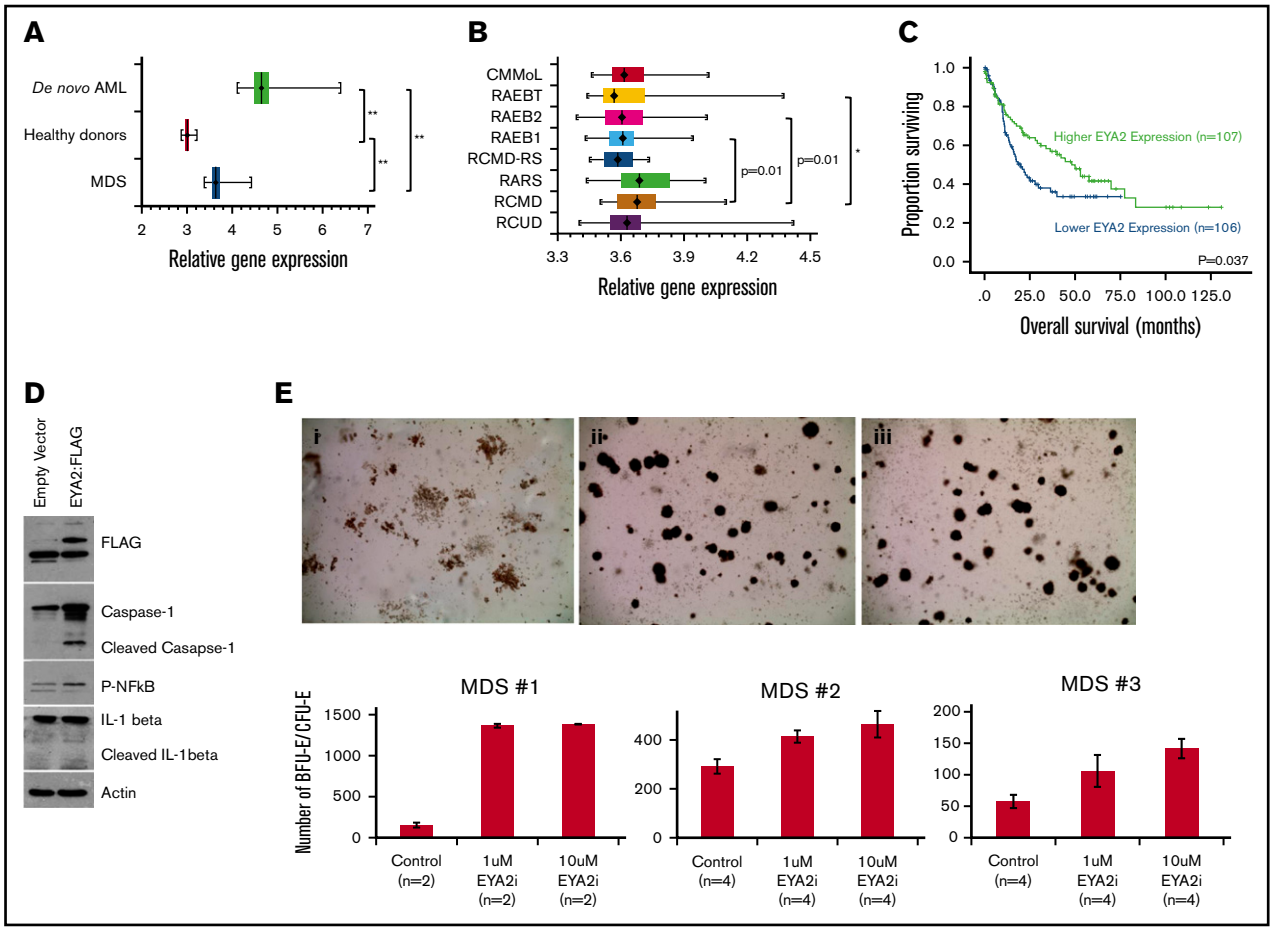

Figure 4. *EYA2* expression and impact on survival and hematopoiesis. (A) Relative gene expression of *EYA2* using the NTUH data set showing significantly higher *EYA2* expression in MDS case subjects compared with control subjects. (B) *EYA2* gene expression is higher in lower risk disease subtypes, including RARS and RCMD. (C) Kaplan-Meier plot showing that higher *EYA2* expression is associated with improved OS. (D) Western blot showing activation of innate immune signaling after EYA2 transfection. (E) Representative micrographs (40×) of colony-forming capacity assays of control (i) or EYA2 inhibitor–treated (ii, 1 μM; iii, 10 μM) MDS bone marrow mononuclear cells showing significantly greater erythroid colonies in treated samples with accompanying quantitation for each of 3 independent samples (mean ± standard error). *$P < .005$, **$P < .0001$.

through the JNK pathway, a key mediator of inflammatory signaling in MDS.[33-35]

Gerstung et al[25] showed that *GRID1* expression is significantly ($P = .013$) upregulated (by 0.023) in association with *DNMT3A* mutations, a commonly found MDS driver mutation that should be explored in the context of the loci identified here. *GRID1* introns encode microRNAs critical for unfolded protein response (UPR).[36] Of particular importance, endoplasmic reticulum (ER) stress arising from unresolved UPR activates the NLRP3 inflammasome and pyroptotic cell death.[37] *U2AF1*-S34F mutants promoted splicing at a distal poly(A) site in autophagy-related factor 7 (ATG7) messenger RNA, thereby generating a longer and poorly translated transcript.[38] Subsequent ATG7 depletion disrupted autophagy to initiate ER stress. UPR and ER stress are often found as a result of somatic gene mutation and are currently being investigated in MDS.

Gerstung et al[25] reported that JAK2 mutations are associated with increased (0.008) *EYA2* gene expression ($P = .051$). JAK2 mutations are found both in MDS and myeloproliferative neoplasms. Of note, interstitial deletion of part of the long arm of chromosome 20 [del(20q)] is a common chromosome abnormality in MDS associated with favorable prognosis, and *EYA2* resides within this commonly deleted region.[39,40] Importantly, we explored the functional activity of *EYA2*, a member of the Eyes absent gene family (*EYA*) in the context of MDS. Of particular interest, *EYA2* contains a unique haloacid dehydrogenase family, tyrosine phosphatase, in the EYA-domain that dephosphorylates H2AX, potentially influencing DNA damage repair**.** The N-terminal region of EYA2 contains a serine/threonine phosphatase involved in the regulation of innate immune response.[16] EYA members are sensors of cytosolic DNA that can serve as DAMPs to trigger inflammasome activation.[16,17] We confirmed activation of innate immune signaling and, using an EYA2 pharmacological inhibitor, we reported concentration-dependent improvement in colony-forming capacity in primary MDS specimens. It is important to note that although increased *EYA2* expression was associated with increased OS, protein inhibition has potential beneficial clinical utility. EYA2 inhibition is able to restore hematopoiesis, particularly erythropoiesis, demonstrable by an increased colony-forming capacity that may result in hematologic improvement.

Hematologic improvement, particularly erythropoiesis improvement, may significantly decrease transfusion dependence, which directly correlates with reduction in survival and rising risk of leukemia transformation.[41] These data suggest potential clinical utility underscoring the biological relevance of *EYA2* and offer a target for therapeutic investigation that may modify NLRP3 inflammasome-mediated pyroptosis in lower risk patients with MDS.

eQTL interrogation of SNP proxies identified several affected genes. Importantly, we used Blood eQTL and GTEx on whole blood only. MDS is a stem cell malignancy characterized by malignant clonal expansion, and analysis of whole blood is therefore not a true representation of diseased tissue. Nonetheless, it is the closest tissue from which there is available content to search. Interestingly, there are some associations of genes identified by using eQTL analysis and MDS. For example, expression of IER3, an NF–κB signaling interacting protein that can regulate cell death, is dysregulated in MDS due in part to commonly observed translocations.[42] Furthermore, *IER3*-deficient mice develop an MDS phenotype after radiation exposure, suggesting a critical role in pathogenesis.[43] Here, *IER3* expression was decreased in unsorted MDS bone marrow cells vs those from healthy donors, and the role of our loci should be explored in this context. Furthermore, we found eQTL for *ARL6IP75* (also *JWA*) and significant upregulation of this gene in MDS $CD34^+$ cells vs healthy donors. *JWA* is heavily involved in myeloid cell differentiation and leukemia initiation,[44] and functional loci variations are associated with increased leukemic risk.[45]

Although we have yet to explore the functional consequences of all loci identified here, the linkage to a known pathogenetic driver of MDS, namely innate immune activation, suggests important clinical and biological implications of this study; these implications are supported by our data on differential gene expression patterns in MDS case subjects vs healthy control subjects, survival prognostication, associations with known prognostically adverse somatic gene mutations, and demonstrated functional data. Validation of these MDS-associated loci in other studies enrolling ample numbers of cases with MDS is needed; however, these findings provide valuable insight into MDS predisposition, biology, and possible therapy targets. Future studies will elucidate the functional impact of the specific polymorphisms and also investigate possible polymorphism associations in del(5q) MDS. In conclusion, we have provided the first evidence of genetic loci linked to MDS predisposition and new pathway targets that may have pathobiological importance in MDS.

Acknowledgments

This work was supported in part by the Cancer Center Support Grant P30CA076292 to the H. Lee Moffitt Comprehensive Cancer Center and Research Institute. F.S. funding included a grant from the Instituto de Salud Carlos III, Ministerio de Economia y Competividad, Spain (PI/14/00013, PI/17/0575), 2017 SGR288 (GRC) Generalitat de Catalunya, support from CERCA Programme/Generalitat de Catalunya, Fundació Internacional Josep Carreras, and funding from "la Caixa" Foundation. A.P. and J.B. are supported by Bloodwise (grant 13042). The NTUH research was partially supported by the grants MOST 100-2314-B-002-057-MY3 and 106-2314-B-002-231 from the Ministry of Science and Technology, Taiwan.

Authorship

Contribution: K.L.M., P.A.K., Y.A.C., and A.F.L. conducted the study, analyzed data, and wrote the manuscript; C.-H.C., H.-A.H., and G.G. performed statistical analysis; T.C., A.P., B.P.P., M.M., L. Arenillas, A.M., L. Adès, C.M., S.R., B.N., J.B., B.L.E., F.S., P.F., G.J.M., and J.P.M. contributed cases; and D.A.S., E.P., L.S., and H.-F.T. analyzed data.

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

ORCID profiles: K.L.M., 0000-0001-9184-5399; G.G., 0000-0003-3066-5575; A.P., 0000-0002-6122-0221; L. Arenillas, 0000-0002-9020-8766; J.B., 0000-0002-4330-2928; F.S., 0000-0002-3251-2161; P.A.K., 0000-0002-5567-9618.

Correspondence: Kathy L. McGraw, Moffitt Cancer Center, 12902 Magnolia Dr, Tampa, FL 33612; e-mail: kathy.mcgraw@moffitt.org.

References

1. Genovese G, Kähler AK, Handsaker RE, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med*. 2014;371(26):2477-2487.
2. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med*. 2014;371(26):2488-2498.
3. Bannon SA, DiNardo CD. Hereditary predispositions to myelodysplastic syndrome. *Int J Mol Sci*. 2016;17(6):E838.
4. Tiu RV, Gondek LP, O'Keefe CL, et al. Prognostic impact of SNP array karyotyping in myelodysplastic syndromes and related myeloid malignancies. *Blood*. 2011;117(17):4552-4560.
5. Makishima H, Rataul M, Gondek LP, et al. FISH and SNP-A karyotyping in myelodysplastic syndromes: improving cytogenetic detection of del(5q), monosomy 7, del(7q), trisomy 8 and del(20q). *Leuk Res*. 2010;34(4):447-453.
6. Yoshizato T, Nannya Y, Atsuta Y, et al. Genetic abnormalities in myelodysplasia and secondary acute myeloid leukemia: impact on outcome of stem cell transplantation. *Blood*. 2017;129(17):2347-2358.
7. McCarthy S, Das S, Kretzschmar W, et al; Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279-1283.

8. Lin DY, Zeng D. Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data. *Genet Epidemiol*. 2010;34(1):60-66. 10.1002/gepi.20435

9. Lin DY, Zeng D. On the relative efficiency of using summary statistics versus individual-level data in meta-analysis. *Biometrika*. 2010;97(2):321-332.

10. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26(18):2336-2337.

11. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*. 2013;41(database issue):D991-D995. 10.1093/nar/gks1193

12. Chuang MK, Chiu YC, Chou WC, Hou HA, Chuang EY, Tien HFA. A 3-microRNA scoring system for prognostication in de novo acute myeloid leukemia patients. *Leukemia*. 2015;29(5):1051-1059.

13. Bagger FO, Sasivarevic D, Sohi SH, et al. BloodSpot: a database of gene expression profiles and transcriptional programs for healthy and malignant haematopoiesis. *Nucleic Acids Res*. 2016;44(D1):D917-D924.

14. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet*. 2013;45(10):1238-1243.

15. McGraw KL, Fuhler GM, Johnson JO, et al. Erythropoietin receptor signaling is membrane raft dependent. *PLoS One*. 2012;7(4):e34477.

16. Okabe Y, Sano T, Nagata S. Regulation of the innate immune response by threonine-phosphatase of Eyes absent. *Nature*. 2009;460(7254):520-524.

17. Liu X, Sano T, Guan Y, Nagata S, Hoffmann JA, Fukuyama H. Drosophila EYA regulates the immune response against DNA through an evolutionarily conserved threonine phosphatase motif. *PLoS One*. 2012;7(8):e42725.

18. Basiorka AA, McGraw KL, Eksioglu EA, et al. The NLRP3 inflammasome functions as a driver of the myelodysplastic syndrome phenotype. *Blood*. 2016;128(25):2960-2975.

19. Astle WJ, Elding H, Jiang T, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell*. 2016;167(5):1415-1429.e19.

20. van der Harst P, Zhang W, Mateo Leach I, et al. Seventy-five genetic loci influencing the human red blood cell. *Nature*. 2012;492(7429):369-375.

21. Kramer RM, Sharp JD. Structure, function and regulation of Ca2+-sensitive cytosolic phospholipase A2 (cPLA2). *FEBS Lett*. 1997;410(1):49-53.

22. Patel MI, Singh J, Niknami M, et al. Cytosolic phospholipase A2-alpha: a potential therapeutic target for prostate cancer. *Clin Cancer Res*. 2008;14(24):8070-8079.

23. Runarsson G, Feltenmark S, Forsell PK, Sjöberg J, Björkholm M, Claesson HE. The expression of cytosolic phospholipase A2 and biosynthesis of leukotriene B4 in acute myeloid leukemia cells. *Eur J Haematol*. 2007;79(6):468-476.

24. Wu YL, Jiang XR, Newland AC, Kelsey SM. Failure to activate cytosolic phospholipase A2 causes TNF resistance in human leukemic cells. *J Immunol*. 1998;160(12):5929-5935.

25. Gerstung M, Pellagatti A, Malcovati L, et al. Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. *Nat Commun*. 2015;6(1):5901.

26. Sokolowska M, Chen LY, Liu Y, et al. Prostaglandin E2 inhibits NLRP3 inflammasome activation through EP4 receptor and intracellular cyclic AMP in human macrophages. *J Immunol*. 2015;194(11):5472-5487.

27. Schenten V, Plançon S, Jung N, et al. Secretion of the phosphorylated form of S100A9 from neutrophils is essential for the proinflammatory functions of extracellular S100A8/A9. *Front Immunol*. 2018;9:447.

28. Wang W, Li T, Wang X, et al. FAM19A4 is a novel cytokine ligand of formyl peptide receptor 1 (FPR1) and is able to promote the migration and phagocytosis of macrophages. *Cell Mol Immunol*. 2015;12(5):615-624.

29. Steenbergen RD, Ongenaert M, Snellenberg S, et al. Methylation-specific digital karyotyping of HPV16E6E7-expressing human keratinocytes identifies novel methylation events in cervical carcinogenesis. *J Pathol*. 2013;231(1):53-62.

30. Dmitriev AA, Rosenberg EE, Krasnov GS, et al. Identification of novel epigenetic markers of prostate cancer by NotI-Microarray Analysis. *Dis Markers*. 2015;2015:241301.

31. Sallman D, Al Ali NH, Padron E, Lancet J, List A, Komrokji R. Chromosomal 3 abnormalities in myelodysplastic syndromes: implications on prognosis and response to treatment. *Blood*. 2014;124(21):4632.

32. Wang Z, Dai J, Hu N, et al. Identification of new susceptibility loci for gastric non-cardia adenocarcinoma: pooled results from two Chinese genome-wide association studies. *Gut*. 2017;66(4):581-587.

33. Noh H, Lee H, Park E, Park S. Proper closure of the optic fissure requires ephrin A5-EphB2-JNK signaling. *Development*. 2016;143(3):461-472.

34. Czibere A, Prall WC, Zerbini LF, et al. The nonsteroidal anti-inflammatory drug Exisulind selectively induces apoptosis via JNK in secondary acute myeloid leukemia after myelodysplastic syndrome. *Cell Cycle*. 2005;4(6):812-817.

35. Oben KZ, Alhakeem SS, McKenna MK, et al. Oxidative stress-induced JNK/AP-1 signaling is a major pathway involved in selective apoptosis of myelodysplastic syndrome cells by Withaferin-A. *Oncotarget*. 2017;8(44):77436-77452.

36. Bartoszewski R, Brewer JW, Rab A, et al. The unfolded protein response (UPR)-activated transcription factor X-box-binding protein 1 (XBP1) induces microRNA-346 expression that targets the human antigen peptide transporter 1 (TAP1) mRNA and governs immune regulatory genes. *J Biol Chem*. 2011;286(48):41862-41870.

37. Lerner AG, Upton JP, Praveen PV, et al. IRE1α induces thioredoxin-interacting protein to activate the NLRP3 inflammasome and promote programmed cell death under irremediable ER stress. *Cell Metab*. 2012;16(2):250-264.

38. Park SM, Ou J, Chamberlain L, et al. U2AF35(S34F) promotes transformation by directing aberrant ATG7 Pre-mRNA 3′ end formation. *Mol Cell*. 2016; 62(4):479-490.

39. Douet-Guilbert N, Basinko A, Morel F, et al. Chromosome 20 deletions in myelodysplastic syndromes and Philadelphia-chromosome-negative myeloproliferative disorders: characterization by molecular cytogenetics of commonly deleted and retained regions. *Ann Hematol*. 2008;87(7): 537-544.

40. Braun T, de Botton S, Taksin AL, et al. Characteristics and outcome of myelodysplastic syndromes (MDS) with isolated 20q deletion: a report on 62 cases. *Leuk Res*. 2011;35(7):863-867.

41. Melchert M, List AF. Management of RBC-transfusion dependence. *Hematology Am Soc Hematol Educ Program*. 2007;398-404.

42. Steensma DP, Neiger JD, Porcher JC, et al. Rearrangements and amplification of IER3 (IEX-1) represent a novel and recurrent molecular abnormality in myelodysplastic syndromes. *Cancer Res*. 2009;69(19):7518-7523.

43. Ramsey H, Zhang Q, Brown DE, Steensma DP, Lin CP, Wu MX. Stress-induced hematopoietic failure in the absence of immediate early response gene X-1 (IEX-1, IER3). *Haematologica*. 2014;99(2):282-291.

44. Huang S, Shen Q, Mao WG, et al. JWA, a novel signaling molecule, involved in all-trans retinoic acid induced differentiation of HL-60 cells. *J Biomed Sci*. 2006;13(3):357-371.

45. Shen Q, Tang WY, Li CP, et al. Functional variations in the JWA gene are associated with increased odds of leukemias. *Leuk Res*. 2007;31(6):783-790.