

The RUNX1 database (RUNX1db): establishment of an expert curated RUNX1 registry and genomics database as a public resource for familial platelet disorder with myeloid malignancy

Familial platelet disorder with associated myeloid malignancy (FPD-MM, OMIM:601399)^{1,2} is a rare cancer predisposition syndrome caused by pathogenic germline variants in *RUNX1*.³ Despite research dating back over two decades, many challenges remain in improving outcomes for individuals with FPD-MM.⁴ Firstly, the syndrome may go unrecognized due to poor recognition of family history and/or access to appropriate genetic testing. Secondly, intentional screening or incidental detection (e.g., tumour-sequencing) of *RUNX1* variants requires access to expert interpretation. Thirdly, after diagnosis, the relative rarity of the disorder inhibits the collation of sizeable local cohorts, making identification of commonalities in disease course and/or outcome highly challenging. To help overcome these significant challenges, we have developed an interactive public web-based international collaborative database for *RUNX1*: RUNX1db (<https://runx1db.runx1-fpd.org>). RUNX1db is a centralized repository for germline *RUNX1* variant information, associated next-generation sequencing (NGS) data, and expert-curated variant information (both germline and somatic).

We recently identified, from publications, 140 different families with germline *RUNX1* variants.⁴ While being a rich resource, historically reported variants are largely not classified according to the American College of Medical Genetics and Genomics/Association for Molecular Pathology (ACMG/AMP) guidelines, only established in 2015.⁵ Additionally, the Clinical Genome Resource myeloid malignancy variant curation expert panel (ClinGen MM-VCEP) recently created guidelines specific for classification of germline *RUNX1* variants.⁶ Gene-specific guidelines, while important, add additional complexity to the curation of identified variants. Making

available expert knowledge to accurately classify these germline variants prevents both missing pathogenic variants or the misattribution of benign variants as causative in families.^{7,8} Additionally, variants identified through clinical services and research studies don't always make it into the public domain due to constraints associated with the reporting of variants through publication or variant repositories. To address some of these challenges, we updated curated variants from publications and undertook an international survey of colleagues, identifying unpublished variants. This study identified an additional 119 families (259 in total), with 164 unique variants. These included ten new variants not previously described

Table 1. RUNX1database genomics cohort demographics.

Parameters		RUNX1 database cohort
Germline Mutations	Total Individuals/Unique Mutations	120/47
Age	All samples: Median (range), years	40 (1-76)
	Malignancy Samples: Median (range), years	43 (3-69)
	Pre-leukemic: Median (range), years	34 (1-76)
Gender	Males: Total/Malignancy	51/23
	Females: Total/Malignancy	67/29
Malignancy Subtype	Total Samples/Individuals	62/48
	AML: #Individuals	29
	MDS: #Individuals	14
	MDS/MPN: #Individuals	2
	MPN: #Individuals	2
	ALL: #Individuals	4
	AL: #Individuals	1
Pre-Leukemic	(#Samples/#Individuals)	65/56

AML: acute myeloid leukemia; MDS: myelodysplastic syndrome; MDS/MPN: myelodysplastic syndrome/myeloproliferative neoplasm; MPN: myeloproliferative neoplasm; ALL: acute lymphoblastic leukemia; AL: acute leukemia.

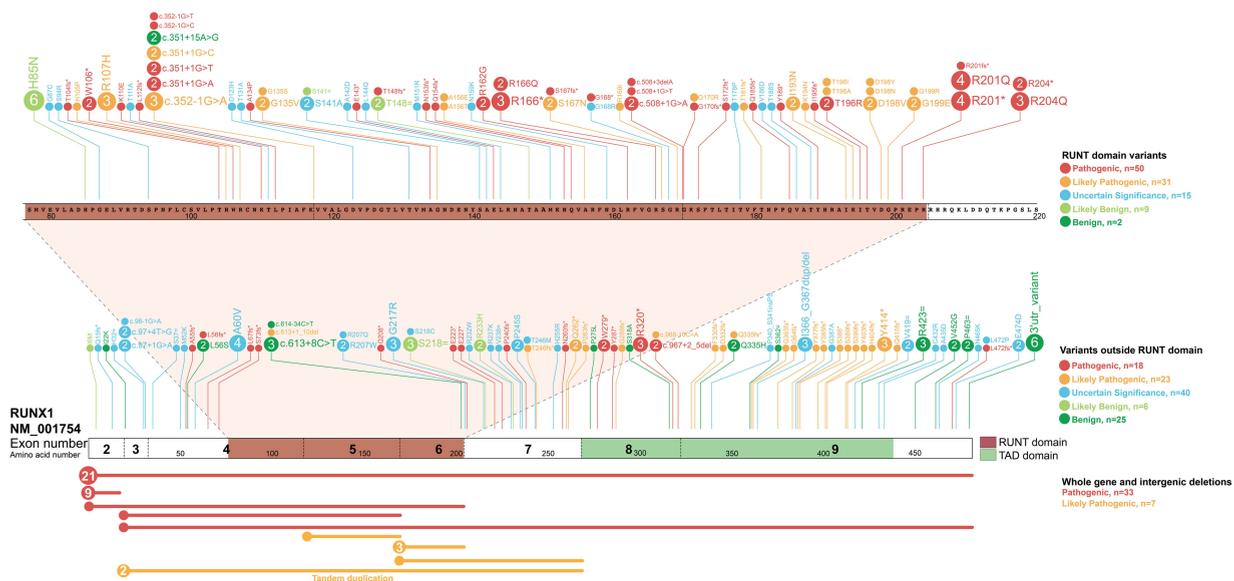


Figure 1. Registry of germline *RUNX1* mutations. Germline *RUNX1* variants currently included in the RUNX1db registry are visualised using the ProteinPaint web application (<https://pecan.stjude.cloud/home>).¹² Variants (displayed as protein changes where possible) are colour-coded according to pathogenicity classification as determined by the MM-VCEP *RUNX1*-specific recommendations. The number of probands for each variant is indicated within the circle where the number is greater than one. All variants are annotated to RUNX1c; NM_001754.4; LRG_482.

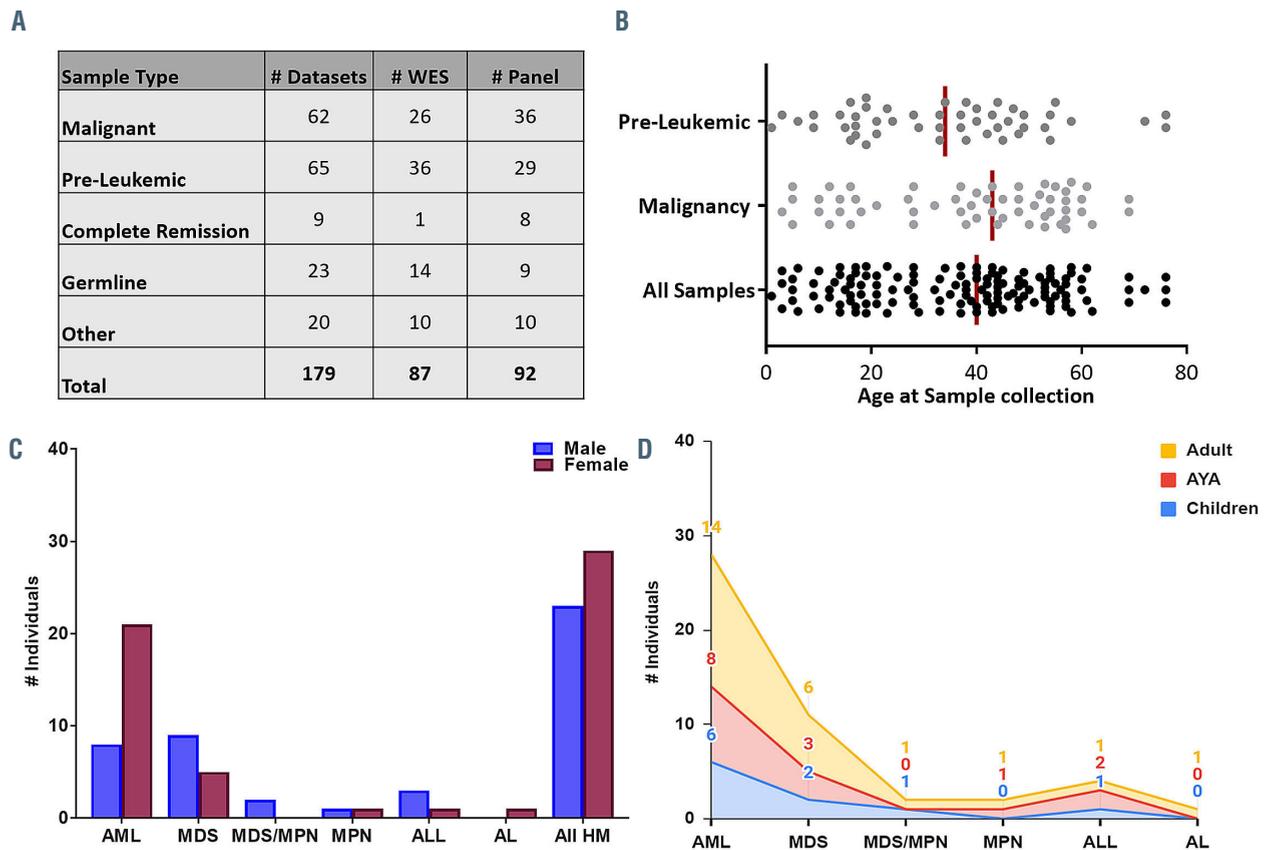


Figure 2. *RUNX1* database genomics cohort demographic. (A) Breakdown of the number and types of NGS samples currently stored in the *RUNX1*db. Pre-Leukemic: thrombocytopenia, asymptomatic Other: includes post-transplant/post-treatment and saliva samples. Both WES and panel data is analysed and stored in the database. (B) Scatter plot displaying the age of the individual when each sample was collected. Major *RUNX1*db cohorts (malignancy and pre-leukemic samples) are displayed. The median age for each cohort is represented by the vertical line. Clinical demographics of the malignancy cohort is shown with the number of individuals with different types of FPD-MM malignancy presentation and the (C) gender and (D) age distribution; Adult ≥ 40 years, AYA=15-39 years, children ≤ 14 years. AML: Acute myeloid leukemia; MDS: myelodysplastic syndromes; MDS/MPN: myelodysplastic syndrome/myeloproliferative Neoplasm overlap; MPN: Myeloproliferative Neoplasm; ALL: acute lymphoblastic leukemia; AL: acute undifferentiated leukemia.

(Figure 1 and Table S1). Using this data, we created the first comprehensive *RUNX1* germline registry and performed expert curation of all variants according to the *RUNX1*-specific ACMG classification rules (ALB, CNH, LAG, LM, CDD MM-VCEP members). The registry represents the largest collection of curated and clinically classified *RUNX1* germline variants to date, providing a unique clinical resource for researchers, clinical genomics laboratories, and haematologists (Figure 1, Table S1). Utilizing this resource, we have identified 97 pathogenic/likely pathogenic *RUNX1* variants, with 54 located within the RUNT domain (RHD)(75% of RHD variants), of which 24 are missense mutations. Only one pathogenic missense variant is observed outside of the RHD, suggesting the RHD is highly intolerant to genetic-variation. Most commonly observed pathogenic germline *RUNX1* variations are whole-gene deletions (21 probands), deletion of exons 1-2 (9 probands), and mutation of amino acid p.Arg201 within the RHD (8 probands)(Table S1). Accessibility and update-ability of this information is available through a live-webportal which hosts the registry (<https://runx1db.runx1-fpd.org/classification/classifications>). Each curated variant has links to patient-phenotypic information and the current clinical classification, including the evidence for each ACMG code assessed and links to external clinical databases, including ClinVar and associated publications. Importantly, expert crowdsourcing allows the real-time updating of the database through user profile accounts. Newly-identified variants can be

easily added to the database and are automatically annotated with over 137 parameters required for accurate classification (e.g., population frequency, pathogenicity predictions). These parameters populate a classification tool that guides users stepwise through the ACMG classification of new variants (or updating current classifications with new information). Once curated and classified, collated information can be exported as an automated classification report summary, flagged for expert-review, shared with other users, and uploaded to ClinVar.

In addition to a germline *RUNX1* variant registry, *RUNX1*db has the capacity to house NGS datasets, creating the first international genomics cohort of this rare disease. This initiative intends to enable researchers to answer questions about FPD-MM beyond germline variant detection. For example, family members, heterozygous for *RUNX1* mutations, can have varying clinical presentations indicating variable penetrance and expressivity. In almost all cases, germline *RUNX1* carriers present with thrombocytopenia and qualitative platelet defects, and progression to hematologic malignancies (HM) is incompletely penetrant with variable age of onset ranging from early childhood to late adulthood.² Patients develop myeloid malignancies most frequently, and T-cell and, more rarely, B-cell acute lymphoblastic leukaemia (ALL).⁴ Currently, there is no way to predict which individuals will progress to myelodysplastic syndrome (MDS), acute myeloid leukaemia (AML), or other HM. Accumulation of somatic mutations and additional

germline modifier variants are mechanisms proposed to contribute to this heterogeneity.⁴ NGS technology is widely used for surveillance and diagnosis of HM,⁴ accumulating large amounts of data often not utilized beyond *RUNX1* variant detection. Individual laboratories often only have small numbers of patients with deleterious *RUNX1* germline variants, which makes asking larger questions about commonalities of genotype-phenotype, disease progression, monitoring, treatment and outcome, difficult.⁹ To accumulate the data required to make evidence-based clinical decisions in FPD-MM, a dedicated resource utilizing the collective wealth of NGS data generated from research and diagnostic laboratories internationally is ideal in standardizing and collating disease-specific clinical and genomics data. The database has also been designed for the accumulation, sharing and curation of genomics data acquired from individuals with germline *RUNX1* mutations both pre- and post-malignancy progression. We have collated 179 NGS datasets, both whole-exome sequencing (WES) and HM gene panel data, from 19 distinct research centres worldwide. This includes NGS from 60 FPD-MM families and 120 individuals, making it the largest FPD-MM NGS dataset (Figure 2). The dataset includes individuals ranging in age from 1-76 years, malignancy phenotypes of AML, MDS, myelodysplastic syndrome/myeloproliferative neoplasm (MDS/MPN), ALL, and pre-leukemic phenotypes including thrombocytopenia and asymptomatic carriers (Table 1). Detailed clinical information for each patient and associated samples are stored on the database and can be updated, enabling specific phenotypic-genotypic cohort studies to be performed on the clinical spectrum of FPD-MM. Additionally, the database can be updated easily with new NGS data as available, including longitudinal datasets from serial testing of individual patients. The database allows for a comprehensive, unbiased and customizable review of all *RUNX1* germline datasets with all raw sequencing data being analyzed through a standardized bioinformatics pipeline. This is designed to identify both somatic and germline variants and is available on the database as variant level data (VCF, Figure S1). Using the integrated VariantGrid (<https://github.com/SACGF/variantgrid>) genomics analysis software, we have curated a panel of somatic variants for each dataset (including all malignancy and pre-leukemic samples), prioritizing the identification of potentially pathogenic variants in HM (2,643 variants, 167 samples). Standard filtering criteria were adapted for identifying somatic variants (*Online Supplementary Figure S1*). Variants that passed all filtering criteria were subsequently manually curated. Variants classified as having no clinical significance (benign/likely benign) according to ACMG/AMP guidelines, were excluded. Remaining variants were either classified as 1) Clinically relevant, 2) Possibly relevant, or 3) Unknown relevance (*Online Supplementary Figure S1*).^{10,11} Curated somatic variant data is available through the interactive-oncplot on the database homepage or variant page. Shared in real-time with the scientific community, this curated dataset has already allowed the selection of secondary mutations to model FPD-MM disease and therapy in vitro and in animals. Importantly, investigators can interrogate the data to answer additional research questions as the software provides a fully automated annotation of variants and allows non-bioinformaticians to filter, sort, analyze, and curate genetic variants stored in the database via a graphical interface (*Online Supplementary Figure S2*).

This project serves as a model for data accumulation for rare cancer predisposition syndromes. The adoption

of a single database that serves as a repository for patient demographic and clinical data, a mutational germline registry, and patient genomics data, which can be interrogated as a large cohort are essential components for the diagnosis and treatment of patients with a rare-disorder such as FPD-MM. This resource is especially useful in FPD-MM, where the genetic cause is well established but variability in clinical presentation and disease development render diagnosis challenging. The aggregation of multiple families, individuals, and disease stages into a centralized database where all data undergo rigorous quality control using a single bioinformatics analysis strategy will aid in the exploration and discovery of the molecular progression of the disorder. The harmonized interpretation of genomic variants is imperative to understanding the mutational profile of a malignancy, which is achieved through a curated list of variants displayed for each sample. Institutional, national, and international ethics and data sharing guidelines may initially limit contributions to initiatives like this that are supported by patient advocates but need to be overcome, given the importance of the work. We envision that information from this database will guide precision-based approaches to patient care plans with reasonable surveillance and adequate counselling and, eventually, the application of new targeted therapies and interventions prior to malignancy development for germline *RUNX1* carriers. With the continued accumulation of data and clinical information, this type of gene-specific database can provide the basis to developing evidence-based clinical decisions such as when to watch and wait and when to apply more aggressive therapies such as stem cell transplantation. Finally, we hope that this database will serve as a model from which similar efforts will emerge for other HMs, benefiting all our patients and families.

Claire C. Homan,^{1,2} Sarah L. King-Smith,^{1,2}
David M. Lawrence,^{1,2,3} Peer Arts,^{1,2} Jinghua Feng,^{2,3}
James Andrews,^{1,2} Mark Armstrong,^{1,2} Thuong Ha,^{1,2}
Julia Dobbins,^{1,2} Michael W. Drazer,⁴ Kai Yu,⁵ Csaba Bödör,⁶
Alan Cantor,⁷ Mario Cazzola,^{8,9} Erin Degelman,¹⁰
Courtney D. DiNardo,¹¹ Nicolas Duployez,^{12,13} Remi Favier,¹⁴
Stefan Fröhling,^{15,16} Jude Fitzgibbon,¹⁷ Jeffery M. Klco,¹⁸
Alwin Krämer,¹⁹ Mineo Kurokawa,²⁰ Joanne Lee,²¹
Luca Malcovati,^{8,9} Neil V. Morgan,²² Georges Natsoulis,²³
Carolyn Owen,¹⁰ Keyur P. Patel,¹¹ Claude Preudhomme,^{12,13}
Hana Raslova,²⁴ Hugh Rienhoff,²⁵ Tim Ripperger,²⁵
Rachael Schulte,²⁶ Kiran Tawana,²⁷ Elvira Velloso,^{28,29}
Benedict Yan,²¹ Paul Liu,⁵ Lucy A. Godley,⁴
Andreas W. Schreiber,^{2,3,30} Christopher N. Hahn,^{1,2,31}
Hamish S. Scott,^{1,2,30,31} and Anna L. Brown^{1,2,31} on behalf of
the *RUNX1* international data-sharing consortium.

¹Department of Genetics and Molecular Pathology, SA Pathology, Adelaide, SA, Australia; ²Centre for Cancer Biology, SA Pathology and University of South Australia, Adelaide, SA, Australia; ³Australian Cancer Research Foundation (ACRF) Cancer Genomics Facility, Centre for Cancer Biology, SA Pathology, Adelaide, SA, Australia; ⁴Section of Hematology/Oncology, Departments of Medicine and Human Genetics, Center for Clinical Cancer Genetics, and The University of Chicago Comprehensive Cancer Center, The University of Chicago, Chicago, IL; ⁵National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892; ⁶HCEMM-SE Molecular Oncohematology Research Group, 1st Department of Pathology and Experimental Cancer Research, Semmelweis University, Budapest, Hungary; ⁷Division of Hematology/Oncology, Boston Children's Hospital and Dana Farber Cancer Institute, Harvard Medical School, Boston, MA; ⁸Department of Molecular Medicine, University of Pavia, Pavia, Italy; ⁹Department of Hematology Oncology, Fondazione IRCCS Policlinico San Matteo,

Pavia, Italy; ¹⁰Division of Hematology and Hematological Malignancies, Foothills Medical Centre, Calgary, AB, Canada; ¹¹Department of Leukemia, University of Texas MD Anderson Cancer Center, Houston, TX; ¹²Laboratory of Hematology, Biology and Pathology Center, Centre Hospitalier Régional Universitaire de Lille, Lille, France; ¹³Jean-Pierre Aubert Research Center, INSERM, Universitaire de Lille, Lille, France; ¹⁴Assistance Publique-Hôpitaux de Paris, Armand Trousseau children's Hospital, Paris, France; ¹⁵Department of Translational Medical Oncology, National Center for Tumor Diseases (NCT) and German Cancer Research Center (DKFZ), Heidelberg, Germany; ¹⁶German Cancer Consortium (DKTK), Heidelberg, Germany; ¹⁷Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, London, UK; ¹⁸St Jude Children's Research Hospital, Memphis, Tennessee, United States; ¹⁹Clinical Cooperation Unit Molecular Hematology/Oncology, German Cancer Research Center (DKFZ) and Dept. of Internal Medicine V, University of Heidelberg, Heidelberg, Germany; ²⁰Department of Hematology & Oncology, Graduate School of Medicine, The University of Tokyo, Japan; ²¹Department of Haematology-Oncology, National University Cancer Institute, National University Health System, Singapore; ²²Institute of Cardiovascular Sciences, College of Medical and Dental Sciences, University of Birmingham, UK; ²³Imago Biosciences, Inc., San Francisco, CA, USA; ²⁴Institut Gustave Roussy, Université Paris Sud, Equipe Labellisée par la Ligue Nationale Contre le Cancer, Villejuif, France; ²⁵Department of Human Genetics, Hannover Medical School, Hannover, Germany; ²⁶Department of Pediatrics, Division of Pediatric Hematology and Oncology, Monroe Carell Jr. Children's Hospital, Vanderbilt University Medical Center, Nashville, TN, USA; ²⁷Department of Haematology, Addenbrooke's Hospital, Cambridge, CB2 0QQ; ²⁸Service of Hematology, Transfusion and Cell Therapy and Laboratory of Medical Investigation in Pathogenesis and Directed Therapy in Onco-Immuno-Hematology (LIM-31) HCFMUSP, University of Sao Paulo Medical School, Sao Paulo, Brazil; ²⁹Genetics Laboratory, Hospital Israelita Albert Einstein, Sao Paulo, Brazil. ³⁰School of Biological Sciences, University of Adelaide, Adelaide, SA, Australia and ³¹School of Medicine, University of Adelaide, Adelaide, SA, Australia

^oClinical Genome Resources Variation Myeloid Malignancy Expert Panel committee members

Correspondence:

ANNA L. BROWN - anna.brown@sa.gov.au

doi:10.3324/haematol.2021.278762

Received: March 22, 2021.

Accepted: July 2, 2021.

Pre-published: July 8, 2021.

Disclosures: the authors declare no competing financial interests.

Contributions: CCH and ALB designed the research, wrote the manuscript, collected NGS and clinical data, curated NGS data, performed ACMG RUNX1 variant classification and analyzed the data. SLK collected and curated NGS data, and performed ACMG RUNX1 variant classification. PA, JD collected and curated NGS data. PA designed Figure 1. MA, TH, JF, AWS designed and performed bioinformatics analysis. DML, JA designed the database and VariantGrid software. MWD, KY, CB, AC, MC, ED, ND, RF, SF, JF, JMK, AK, MK, JL, NVM, GN, CO, KPP, CP, HR, TR, RS, KT, EV, BY, PL, CDD, LAG, LM, ALB contributed NGS data, clinical patient information and scientific insight. CDD, LAG, LM, ALB as members of the MM-VCEP advised on RUNX1 variant classification. ALB, CNH, HSS conceived and designed the study. All authors critically reviewed and approved the manuscript.

Acknowledgments: the authors would also like to thank the RUNX1 Research Program for their support in helping to facilitate the development of the database and fostering collaborations. We also thank the patients and their family members for their willingness to participate in

this study and the RUNX1 international data-sharing consortium for their valuable contributions. This project is also proudly supported by funding from the Leukaemia Foundation of Australia, and project grants APP1145278 and APP1164601 from the National Health and Medical Research Council of Australia. This work was produced with the financial and additional support of Cancer Council SA's Beat Cancer Project on behalf of its donors and the State Government of South Australia, through the Department of Health (PRF Fellowship to HSS). PA is supported by a fellowship from The Hospital Research Foundation. Part of this project was undertaken whilst PA was holding a Royal Adelaide Hospital Mary Overton Early Career Fellowship. LM is supported by the Associazione Italiana per la Ricerca sul Cancro (AIRC) (Accelerator Award Project 22796; 5x1000 Project 21267; Investigator Grant 2017 Project 20125).

LAG is supported by the Cancer Research Foundation. KY and PL are supported by the Division of Intramural Research, National Human Genome Research Institute, NIH. TR is supported by a grant from the European Hematology Association (EHA) and BMBF MyPred (01GM1911B). CB is supported by the EU's Horizon 2020 Research and Innovation Program, under grant agreement No. 739593. The RUNX1 international data-sharing consortium includes all co-authors and others, including Michael Doube (Masaryk University, Czechia), Stephen Langabeer (St. James's Hospital, Ireland), Koneti Rao (Sol Sherry Thrombosis Research Center, USA), Josée Hébert (Université de Montréal, Canada), Lauren M. Bear (Massachusetts General Hospital, USA), Timothy A. Graubert (Massachusetts General Hospital, USA), Akiko Shimamura (Harvard Medical School, USA), Peter Ganly (Canterbury District Health Board, NZ), Marc H.G.P. Raaijmakers (Erasmus Medical Center Cancer Institute, Netherlands), Peter J.M. Valk (Erasmus Medical Center Cancer Institute, Netherlands), Paula Heller (Instituto de Investigaciones Médicas (IDIM) Alfredo Lanari, Argentina).

Funding: this work is supported by a grant from the RUNX1 Research Program.

References

- 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. [doi:10.1182/blood-2016-03-6621](https://doi.org/10.1182/blood-2016-03-6621).
- Brown AL, Hahn CN, Scott HS. Secondary leukemia in patients with germline transcription factor mutations (RUNX1, GATA2, CEBPA). *Blood*. 2020;136(1):24-35. [doi:10.1182/blood-2019-07-8701](https://doi.org/10.1182/blood-2019-07-8701).
- Song WJ, Sullivan MG, Legare RD, et al. Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat Genet*. 1999;23(2):166-175. [doi:10.1038/10189](https://doi.org/10.1038/10189).
- Brown AL, Arts P, Carmichael CL, et al. RUNX1-mutated families show phenotype heterogeneity and a somatic mutation profile unique to germline predisposed AML. *Blood Adv*. 2020;4(6):1131-1144. [doi:10.1182/bloodadvances.2020.014701](https://doi.org/10.1182/bloodadvances.2020.014701).
- Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-424. [doi:10.1038/gim.2015.30](https://doi.org/10.1038/gim.2015.30).
- Luo X, Feurstein S, Mohan S, et al. ClinGen Myeloid Malignancy Variant Curation Expert Panel recommendations for germline RUNX1 variants. *Blood Adv*. 2019;3(20):2962-2979. [doi:10.1182/bloodadvances.2019.014701](https://doi.org/10.1182/bloodadvances.2019.014701).
- Brown AL, Hahn C, Hiwase D, Godley LA, Scott HS. Correct application of variant classification guidelines in germline RUNX1 mutated disorders to assist clinical diagnosis. *Leuk Lymphoma*. 2020;61(1):246-247. [doi:10.1080/10420152.2020.1781111](https://doi.org/10.1080/10420152.2020.1781111).
- Feurstein S, Zhang L, DiNardo CD. Accurate germline RUNX1 variant interpretation and its clinical significance. *Blood Adv*. 2020;4(24):6199-6203. [doi:10.1182/bloodadvances.2020.014701](https://doi.org/10.1182/bloodadvances.2020.014701).
- Bellissimo DC, Speck NA. RUNX1 mutations in inherited and sporadic leukemia. *Front Cell Dev Biol*. 2017;5:111. [doi:10.3389/fcell.2017.00111](https://doi.org/10.3389/fcell.2017.00111).
- Branford S, Wang P, Yeung DT, et al. Integrative genomic analysis reveals cancer-associated mutations at diagnosis of CML in patients with high-risk disease. *Blood*. 2018;132(9):948-961. [doi:10.1182/blood-2017-08-8000](https://doi.org/10.1182/blood-2017-08-8000).
- Li MM, Datto M, Duncavage EJ, et al. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: a joint consensus recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *J Mol Diagn*. 2017;19(1):4-23. [doi:10.1016/j.jmoldiag.2016.11.002](https://doi.org/10.1016/j.jmoldiag.2016.11.002).
- Zhou X, Edmonson MN, Wilkinson MR, et al. Exploring genomic alteration in pediatric cancer using ProteinPaint. *Nat Genet*. 2016;48(1):4-6. [doi:10.1038/ng.3542](https://doi.org/10.1038/ng.3542).