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

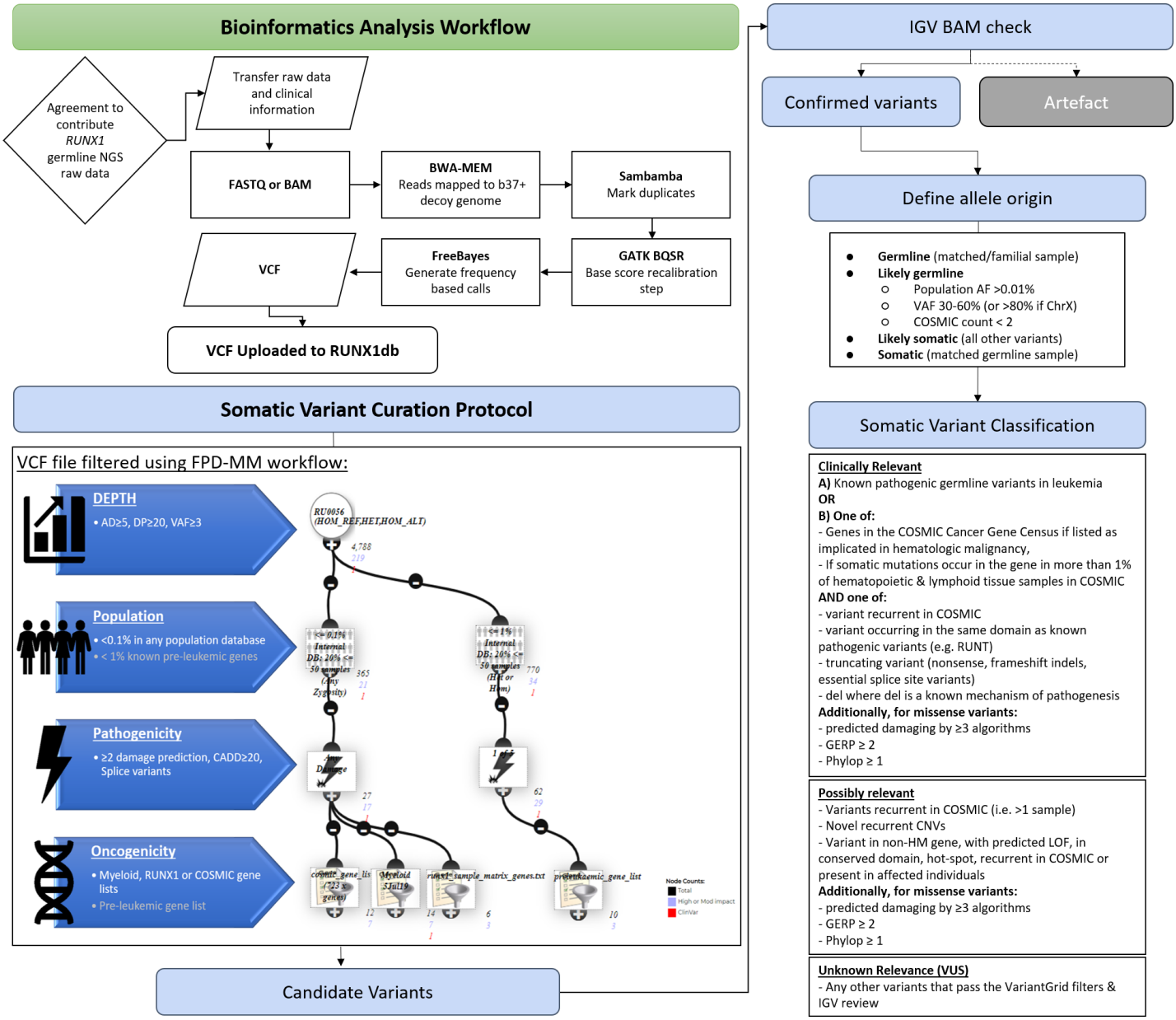
Claire C. Homan,[1,2] Sarah L. King-Smith,[1,2] David M. Lawrence,[1,2,3] Peer Arts,[1,2] Jinghua Feng,[2,3] James Andrews,[1,2] Mark Armstrong,[1,2] Thuong Ha,[1,2] Julia Dobbins,[1,2] Michael W. Drazer,[4] Kai Yu,[5] Csaba Bödör,[6] Alan Cantor,[7] Mario Cazzola,[8,9] Erin Degelman,[10] Courtney D. 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,[23] 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.

[1]Department of Genetics and Molecular Pathology, SA Pathology, Adelaide, SA, Australia; [2]Centre for Cancer Biology, SA Pathology and University of South Australia, Adelaide, SA, Australia; [3]Australian Cancer Research Foundation (ACRF) Cancer Genomics Facility, Centre for Cancer Biology, SA Pathology, Adelaide, SA, Australia; [4]Section 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; [5]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892; [6]HCEMM-SE Molecular Oncohematology Research Group, 1st Department of Pathology and Experimental Cancer Research, Semmelweis University, Budapest, Hungary; [7]Division of Hematology/Oncology, Boston Children's Hospital and Dana Farber Cancer Institute, Harvard Medical School, Boston, MA; [8]Department of Molecular Medicine, University of Pavia, Pavia, Italy; [9]Department of Hematology Oncology, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy; [10]Division of Hematology and Hematological Malignancies, Foothills Medical Centre, Calgary, AB, Canada; [11]Department of Leukemia, University of Texas MD Anderson Cancer Center, Houston, TX; [12]Laboratory of Hematology, Biology and Pathology Center, Centre Hospitalier Regional Universitaire de Lille, Lille, France; [13]Jean-Pierre Aubert Research Center, INSERM, Universitaire de Lille, Lille, France; [14]Assistance Publique-Hôpitaux de Paris, Armand Trousseau children's Hospital, Paris, France; [15]Department of Translational Medical Oncology, National Center for Tumor Diseases (NCT) and German Cancer Research Center (DKFZ), Heidelberg, Germany; [16]German Cancer Consortium (DKTK), Heidelberg, Germany; [17]Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, London, UK; [18]St Jude Children's Research Hospital, Memphis, Tennessee, United States; [19]Clinical Cooperation Unit Molecular Hematology/Oncology, German Cancer Research Center (DKFZ) and Dept. of Internal Medicine V, University of Heidelberg, Heidelberg, Germany; [20]Department of Hematology & Oncology, Graduate School of Medicine, The University of Tokyo, Japan; [21]Department of Haematology-Oncology, National University Cancer Institute, National University Health System, Singapore; [22]Institute of Cardiovascular Sciences, College of Medical and Dental Sciences, University of Birmingham, UK; [23]Imago Biosciences, Inc., San Francisco, CA, USA; [24]Institut Gustave Roussy, Université Paris Sud, Equipe Labellisée par la Ligue Nationale Contre le Cancer, Villejuif, France; [25]Department of Human Genetics, Hannover Medical School, Hannover, Germany; [26]Department of Pediatrics, Division of Pediatric Hematology and Oncology, Monroe Carell Jr. Children's Hospital, Vanderbilt University Medical Center, Nashville, TN, USA; [27]Department of Haematology, Addenbrooke's Hospital. Cambridge, CB2 0QQ; [28]Service 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; [29]Genetics Laboratory, Hospital Israelita Albert Einstein, Sao Paulo, Brazil. [30]School of Biological Sciences, University of Adelaide, Adelaide, SA, Australia and [31]School of Medicine, University of Adelaide, Adelaide, SA, Australia

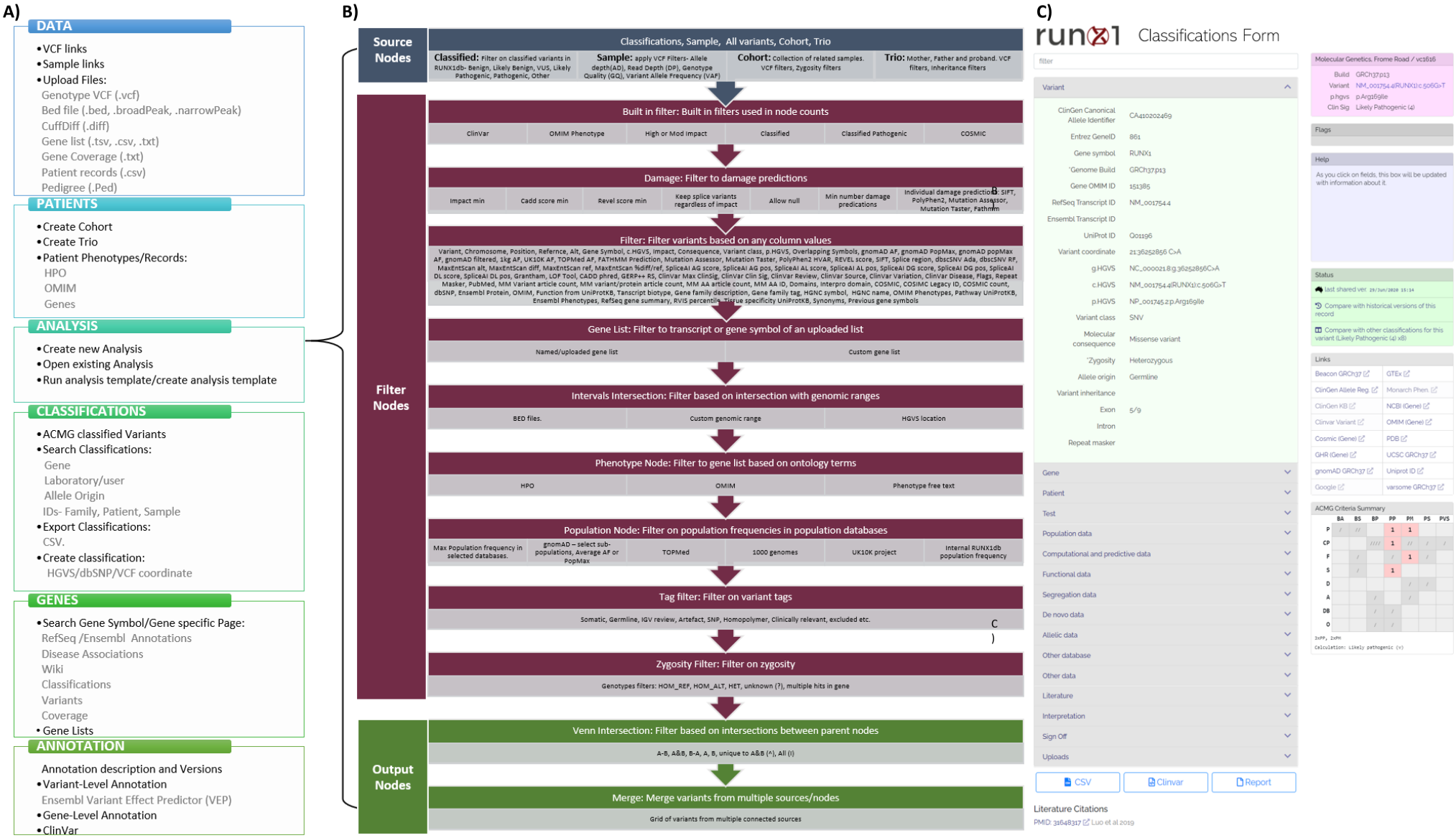
°Clinical Genome Resources Variation Myeloid Malignancy Expert Panel committee members

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

doi:10.3324/haematol.2021.278762



Supplementary Figure 1: Flow diagram outlining the RUNX1db genomics cohort bioinformatics analysis and somatic curation pipelines: Current NGS platforms represented within the database include: Whole exome Sequencing (Illumina), TruSight Myeloid Sequencing Panel (Illumina), Custom amplicon panels, custom capture panels and AmpliSeq panels (Ion torrent). Datasets for inclusion were preferentially obtained as raw data in the FASTQ format. Sequence reads were aligned to the GRCh37 (hs37d5) human reference genome with BWA-MEM (ver 0.7.12). Sambamba (ver 0.6.5) was used for marking PCR duplicates and GATK (ver 3.8-1) for recalibrating base-quality scores. Freebayes (ver 1.2) was used to call single nucleotide variants (SNVs) and insertions/deletions (INDELs). To increase sensitivity and permit the joint analysis of many samples, Freebayes was run in two passes, as previously described[1]. VCF output was uploaded onto the RUNX1 database. Variant, gene and protein level annotation were performed using an in-house pipeline (https://github.com/SACGF/variantgrid). VCFs were subsequently filtered (VariantGrid analysis software) and curated according to the outlined procedure to identify somatic variants of relevance. Grey writing in the FPD-MM filtering workflow indicates additional filtering applied to pre-leukemic samples only. Somatic variant filtering: Utilising the VariantGrid analysis software a somatic variant curation pipeline was developed. Sample Filter: AD≥5, DP≥20, VAF≥3%. Population Filter: Max population frequency of 0.1% in gnomAD (selected populations: African/African American, East Asian, Latino/Mixed Amerindian, non-Finnish European, South Asian), 1.0% for pre-leukemic samples. Damage Filter: Impact minimum=moderate, CADD score ≥20, Minimum 2 damage predictions, allow null (frameshift considered damaging) and keep splice variants. Oncogenicity Filters (https://runx1db.runx1-fpd.org/genes/gene_lists): Variants which passed all filtering criteria were subsequently manually curated.



Supplementary Figure 2: Capabilities and functionality of the RUNX1 database for genomics analysis: A) Highlighting the six top-level menus of the RUNX1db and the functions and/or data which can be utilised/accessed within each menu. B) Description of the nodes available and filters which can be applied for users to create their own analysis workflow utilising the sequencing data stored in the RUNX1db. Coloured boxes indicate the filtering node and description of filter use. Grey boxes indicate the sub-filters which can be applied within the node. C) Image depicts an example of the classification form which has been used to classify a *RUNX1* likely pathogenic germline variant. A number of fields in the classification form are auto populated from annotation data. Based on the input to the classification form the ACMG Criteria summary will output the variant pathogenicity prediction. This form is available for users to classify novel variants in the database.

Supplementary Table 1: *RUNX1* germline variant registry. All variants are classified according to MM-VCEP RUNX1-specific recommendations and links to the MM-VCEP variant interpretation page provided where available. Variants are annotated to RUNX1c; NM_001754.4; LRG_ 482.

Supplementary data reference list

1. Singhal D, Wee LYA, Kutyna MM, et al. The mutational burden of therapy-related myeloid neoplasms is similar to primary myelodysplastic syndrome but has a distinctive distribution. Leukemia. 2019;33(12):2842-2853.