LETTER TO THE EDITOR

Increased frequency of clonal hematopoiesis of indeterminate potential in Bloom syndrome probands and carriers

Authors

Isabella Lin,[1,2,3] Angela Wei,[1,2,3,4] Tsumugi A. Gebo,[5,6] Paul C. Boutros,[1,4,5,6,7,8] Maeve Flanagan,[9] Nicole Kucine,[9] Christopher Cunniff,[9] Valerie A. Arboleda[1,2,3,4,5,8#] and Vivian Y. Chang[5,6,10,11#]

[1]Department of Human Genetics, David Geffen School of Medicine, UCLA, Los Angeles, CA; [2]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, UCLA, Los Angeles, CA; [3]Department of Computational Medicine, David Geffen School of Medicine, UCLA, Los Angeles, CA; [4]Interdepartmental BioInformatics Program, UCLA, Los Angeles, CA; [5] Jonsson Comprehensive Cancer Center, UCLA, Los Angeles, CA; [6]Institute for Precision Health, University of California Los Angeles, Los Angeles, CA; [7]Department of Urology, University of California Los Angeles, Los Angeles, CA; [8]Molecular Biology Institute, University of California, Los Angeles, CA; [9]Department of Pediatrics, Weill Cornell Medicine, New York, NY; [10]Division of Pediatric Hematology/ Oncology, UCLA, Los Angeles, CA and [11]Children's Discovery and Innovation Institute, UCLA, Los Angeles, CA, USA

[#]VAA and VYC contributed equally as senior authors.

Correspondence:
V.Y. CHANG - vchang@mednet.ucla.edu

https://doi.org/10.3324/haematol.2024.285239

Received: March 20, 2024.
Accepted: July 18, 2024.
Early view: July 25, 2024.

Supplemental Figures and Legends for

Increased Frequency of Clonal Hematopoiesis of Indeterminate Potential in Bloom Syndrome Probands and Carriers

Isabella Lin[1,2,3], Angela Wei[1,2,3,4] ,Tsumugi A. Gebo[5,6], Paul C. Boutros[1,4,5,6,7,8], Maeve Flanagan[9], Nicole Kucine[9], Christopher Cunniff[9], Valerie A. Arboleda*[1,2,3,4,5,8], Vivian Y. Chang*[5,6,10,11].

[1] Department of Human Genetics, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA
[2] Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA
[3] Department of Computational Medicine, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA
[4] Interdepartmental BioInformatics Program, UCLA, Los Angeles, CA, USA
[5] Jonsson Comprehensive Cancer Center, UCLA, Los Angeles, CA, USA
[6] Institute for Precision Health, University of California Los Angeles, Los Angeles, CA, USA
[7] Department of Urology, University of California Los Angeles, Los Angeles, CA, USA
[8] Molecular Biology Institute, University of California, Los Angeles, CA, USA
[9] Department of Pediatrics, Weill Cornell Medicine, New York, NY, USA
[10] Division of Pediatric Hematology/ Oncology, UCLA, Los Angeles, CA, USA
[11] Children's Discovery and Innovation Institute, UCLA, Los Angeles, CA, USA

*These senior authors contributed equally

Correspondence:
Vivian Y. Chang, MD, MS
Associate Professor
Department of Pediatrics, Pediatric Hematology-Oncology
David Geffen School of Medicine at UCLA
Email vchang@mednet.ucla.edu
Phone 310-825-6708

Keywords: Cancer Predisposition, Bloom Syndrome, Somatic Mutation, Clonal Hematopoiesis

Supplemental Figures and Legends

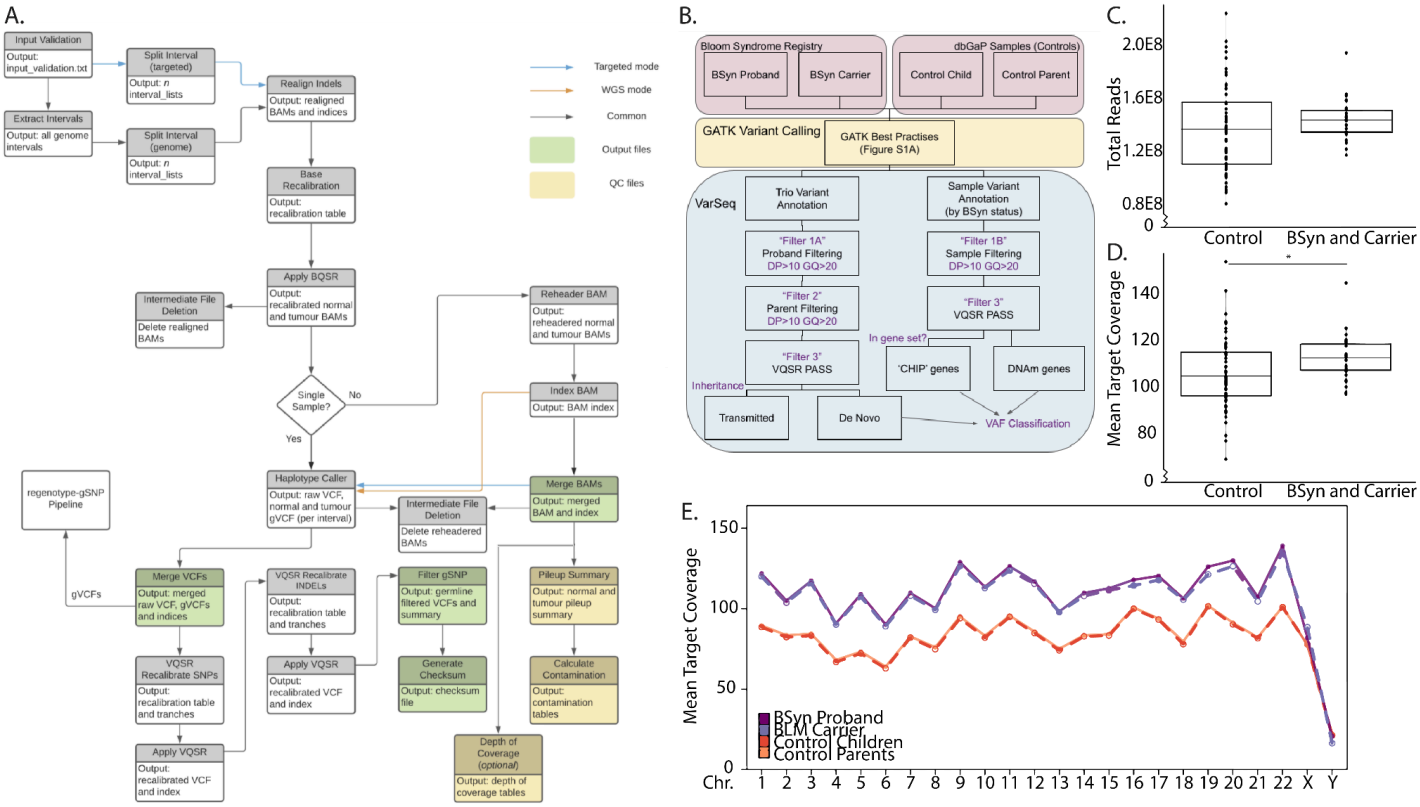

Figure S1: Analysis Pipelines and Sample Sequencing Quality Control.
(**A**) All exome files were processed using the UCLA CDS germline SNP pipeline v5.4.3 which incorporated GATK best practices. Base Quality Score Recalibration (BQSR) and Variant Quality Score Recalibration (VQSR) were performed using GATK v4.2.4.1 and local realignments were performed using GATK v3.7.0. (**B**) Merged raw variant call format (VCF) files generated from the UCLA CDS pipeline (yellow box) were then processed with the VarSeq v.2.3.0 Exome Trio Template (blue box). Annotated variants were then filtered for read depth (DP) and genotype quality (GQ) before classification (blue box). (**C**) We plotted total sequencing reads for each sample, and calculated mean total reads in the Bloom Syndrome (BSyn) cohort (1.44×10^8) and the control cohort (1.38×10^8) (t-test, *p*-value=0.268). (**D**) We plotted exome coverage for each sample, and calculated mean exome coverage in BSyn cohort (113.1x) and the control cohort (106.5x) (t-test, *p*-value=0.018). (**E**) Mean exome coverage between the cohorts - BSyn proband (n=10), and carrier (n=19), control children (n=19), and parents (n=38) - at each chromosome was plotted to examine any chromosome-specific enrichment in exome coverage.
p-values were determined by t-test, and ns denote *p*-value>.05, * denote *p*-value≤.05

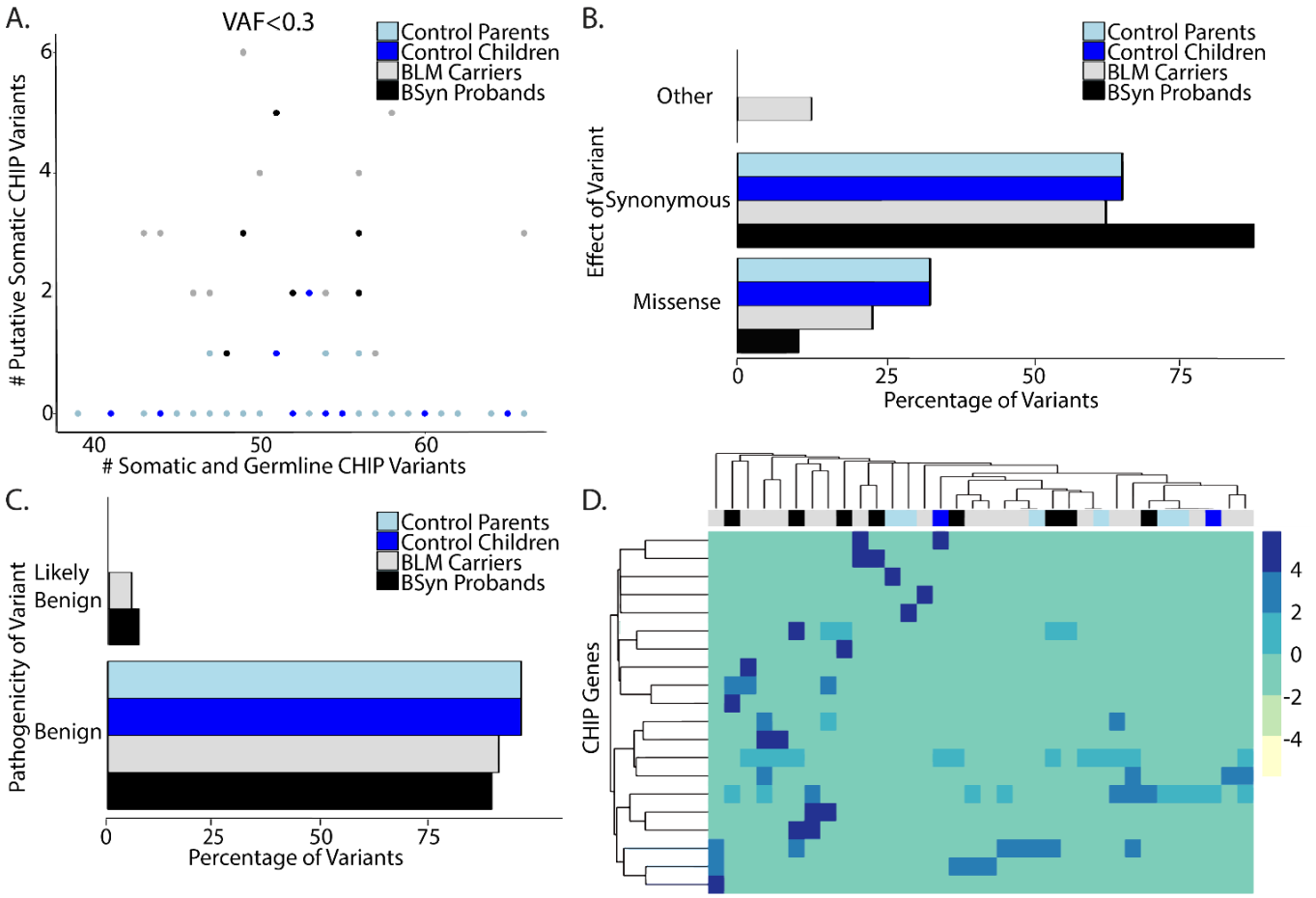

Figure S2: Clonal Haematopoiesis of Indeterminate Potential Variant Load and *De Novo* Variants in Bloom Syndrome Probands and Carriers

(**A**) Total number of clonal haematopoiesis of indeterminate potential (CHIP) gene variants identified in a sample compared to the number of somatic CHIP gene variants. (**B**) Consequences (RefSeq Genes 110, NCBI) of the putative somatic CHIP variants, specifically synonymous, missense, loss of function (LoF) or other (splice etc.) (**C**) Breakdown of somatic CHIP variants based on pathogenicity (ClinVar 2023-01- 05, NCBI). (**D**) Breakdown of putative somatic CHIP variants across all 56 CHIP genes (RefSeq Genes 110, NCBI) identified in literature, organized based on gene of variation (y axis) and by sample (x axis) using hierarchical clustering.

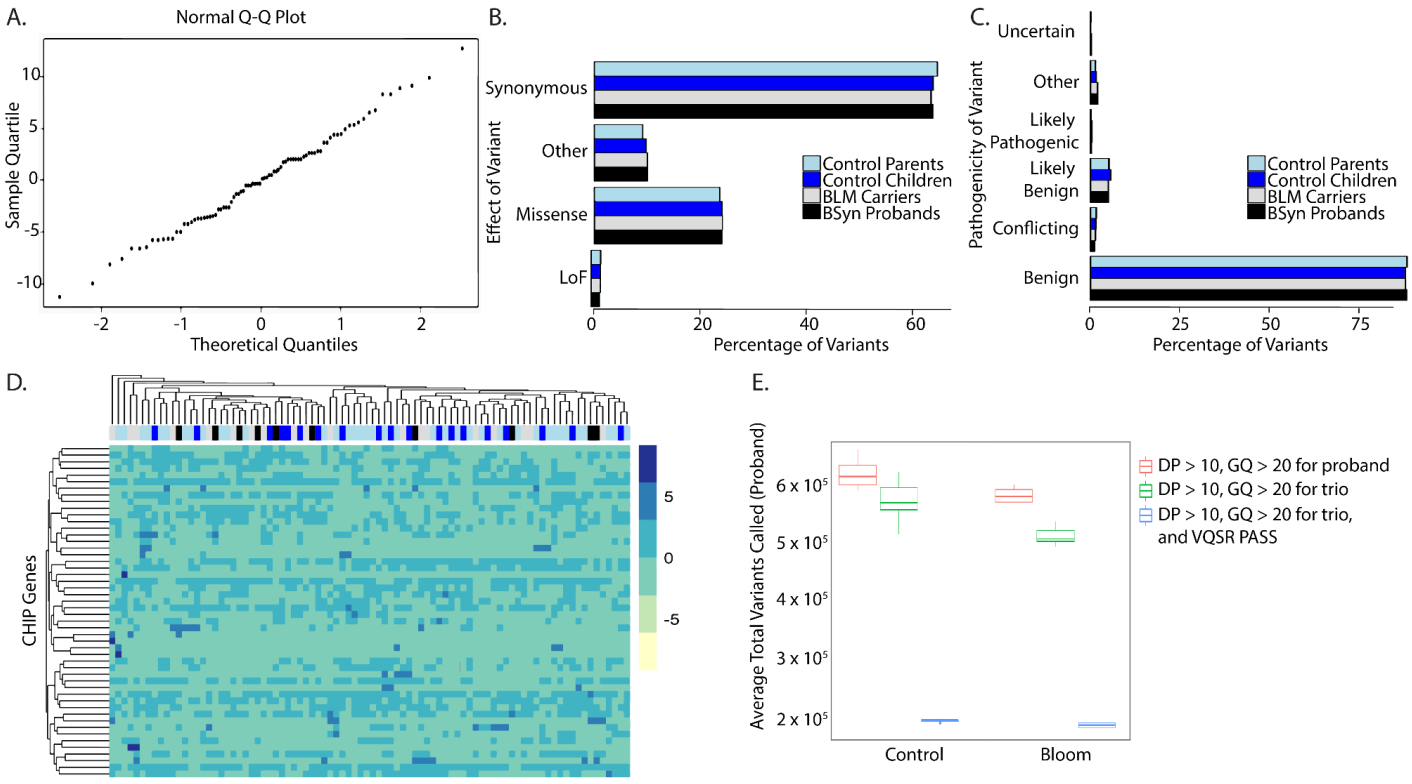

Figure S3: Total Variants in Clonal Haematopoiesis of Indeterminate Potential Genes in Bloom Syndrome and Control Cohorts

(**A**) A Quantile-Quantile (Q-Q) plot to assess the normal distribution of total clonal haematopoiesis of indeterminate potential (CHIP) gene variants for each sample. Each point represents a quantile of the observed data compared to the expected quantile of a normal distribution. (**B**) Consequences (RefSeq Genes 110, NCBI) of total CHIP gene variants, specifically synonymous, missense, loss of function (LoF) or other (splice etc.) (**C**) Pathogenicity of total CHIP gene variants (ClinVar 2023-01-05, NCBI). (**D**) Breakdown of CHIP gene variants across all 56 CHIP genes (RefSeq Genes 110, NCBI) identified in literature based on gene of variation (y axis) and by sample (x axis) using hierarchical clustering. The heatmap depicts the distribution of all CHIP variants regardless of variant allele frequency (VAF) identified in the samples, with each row representing a CHIP gene and each column representing a sample. (**E**) Trio exome samples for Bloom Syndrome (BSyn) probands (n=9) and control children (n=19) annotated and processed in VarSeq v.2.3.0 using read depth, genotype quality, and variant PASS filters.