haematologica
Journal of the Ferrata Storti Foundation

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

by Isabella Lin, Angela Wei, Tsumugi A. Gebo, Paul C. Boutros , Maeve Flanagan, Nicole Kucine, Christopher Cunniff, Valerie A. Arboleda, and Vivian Y. Chang

Received: March 20, 2024.
Accepted: July 18, 2024.

Citation: Isabella Lin, Angela Wei, Tsumugi A. Gebo, Paul C. Boutros , Maeve Flanagan, Nicole Kucine, Christopher Cunniff, Valerie A. Arboleda, and Vivian Y. Chang. Increased frequency of clonal hematopoiesis of indeterminate potential in Bloom syndrome probands and carriers. Haematologica. 2024 July 25. doi: 10.3324/haematol.2024.285239 [Epub ahead of print]

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

Data Sharing Statement. Data available on request to corresponding author.

Funding. This work was supported by the following funding sources awarded to VAA: NIH DP5OD024579 and the ASXL Research Related Endowment Pilot Grant (2020-2022), IL: NIH T32 GM008042 and VYC: NIH 1K08HL138305. CC and NK received support from The New York Community Trust.

Acknowledgements. We thank the UCLA Technology for Genomics and Bioinformatics for their sequencing expertise, and Dr. Jeff Gornbein from the UCLA Statistics Department for his expertise and aid in statistical analyses. We would like to acknowledge the families and patients who participated in this study and donated their time and efforts to make this possible.

Conflict of interest statement.

PCB sits on the Scientific Advisory Boards of Sage Bionetworks, Intersect Diagnostics Inc. and BioSymetrics Inc.

Contributorship Statement.

VC, VA, IL and AW designed and conceptualized the study, analyzed the generated data, and wrote the paper. PB and TAG performed mapping and variant calling of WES data and IL performed VarSeq trio exome analysis. MF, NK, CC coordinated sample collection and DNA extraction. All authors contributed to the final editing of the manuscript.

Bloom syndrome (BSyn, OMIM #210900) is a rare autosomal recessive disorder characterized by growth restriction, sun sensitivity, insulin resistance, mild immune deficiency, and increased risk of early-onset malignancy[1]. BSyn cases are caused by homozygous or compound heterozygous pathogenic variants (PVs) in *BLM*, with over 547 different PVs identified in ClinVar[2]. The Bloom Syndrome Registry (BSR) recently reported that 53% of participants had developed cancer, with hematologic malignancies being the most common cancer risk[3]. While several studies have shown no association of carriers having increased risk[4,5], recent studies have identified increased risk of cancers in *BLM* PV carriers such as colorectal cancer[6], breast cancer[7] and mesothelioma[8].

Clonal hematopoiesis of indeterminate potential (CHIP) is characterized by somatic mutations in leukemia-related genes detected in individuals without apparent hematologic malignancy[9]. CHIP is associated with an annual increased risk of leukemia ranging from 0.5% to 1.0%[9]. Increased age and presence of germline variants in DNA repair and telomere maintenance genes are associated with increased prevalence of CHIP[10].

The latter suggests that germline PVs can create a “permissive” environment for clonal evolution[11], leading to clonal selection in hematopoietic cells. In a longitudinal study spanning 14 years with 4,596 participants who developed blood malignancies, 18 genes were found to predispose individuals to clonal hematopoiesis. Notably, *BLM* was one of these genes [12].

Therefore, we hypothesized that 1 or 2 germline *BLM P*Vs may heighten CHIP risk in a dose dependent fashion with 1 germline PV associated with mildly elevated malignancy risk while 2 germline PVs are associated with increased malignancy risk at an early age. Using exome sequencing of BSyn patients and *BLM* carriers, we found that both BSyn probands and *BLM* carriers exhibited an increased frequency of CHIP compared to sex- and age-matched controls. This study sheds new light on the interplay between genetic predispositions and somatic variation and highlights the need for additional studies to further evaluate the mechanisms and potential clinical implications for patients with one or two *BLM* PVs.

All study participants provided informed consent under a protocol for the BSR approved by the Weill Cornell Medical College Institutional Review Board, and a material transfer agreement was obtained. We performed exome sequencing with the Nextera DNA Flex Pre-Enrichment Library Prep and the Roche NimbleGen exome capture kit following standard protocols. Libraries were indexed, multiplexed and sequenced on a 2x150 Illumina NovaSeq S1 flowcell at the UCLA Technology Center for Genomics and Bioinformatics.

Age- and sex-matched control trios were obtained from the publicly available dbGAP study phs000178.v11.p8.c1, submitted by the Center for Mendelian Genomics [CMG] - The Broad Institute Joint Center for Mendelian Genomics - The Broad Institute Joint Center for Mendelian Genomics. Control trios harbored undiagnosed disease without cancer phenotypes and samples were processed with the Illumina Nextera Exome Kit and sequenced on an Illumina HiSeq. The following public datasets were used: SRA ID SRS2136666, SRS2813808, SRS2136486, SRS2140039, SRS2140061, SRS2136721, SRS2203482, SRS2202906, SRS2202907, SRS2130875, SRS2136628, SRS2130876, SRS2197363, SRS2197826, SRS2197795, SRS2140305, SRS2137393, SRS2137389, SRS2200570, SRS2200550, SRS2200596, SRS2195820, SRS2195786, SRS2195798, SRS2200588, SRS2200627, SRS2200615, SRS2205811, SRS2195821, SRS2195834, SRS2136679, SRS2136619, SRS2136617, SRS2136629, SRS2136659, SRS2136613, SRS2203316, SRS2202950, SRS2202953, SRS2203490, SRS2203471, SRS2203336, SRS2200551, SRS2200573, SRS2200609, SRS2200624, SRS2200562, SRS2200610, SRS2200576, SRS2200561, SRS2288808, SRS2288810, SRS2288816, SRS2200626, SRS2200613, SRS2288805, SRS2200605

All FASTQ files underwent unified quality control, mapping and variant-calling based on GATK best-practices pipeline[13] (**Figure S1A and S1B**). Variant calls were filtered to maintain read depth (DP)>10 over the alternate allele. Variants were initially filtered for DP and analyzed

for coverage across regions of interest. Subsequent filtering for genotype quality (GQ) and a quality score of “PASS” were included (**Figure S1B**).

Variant allele frequencies (VAF), representing the percentage of sequencing reads matching a specific DNA variant, were used as a surrogate measure of allele proportion. A VAF<0.3 indicated acquired somatic variants, while VAF≥0.3 indicated likely germline or *de novo* variants[14]. CHIP is further defined as a somatic mutation in peripheral blood leukocytes with a VAF>0.02[15]**.**

We performed exome sequencing on 29 peripheral blood DNA samples obtained from the BSR. The cohort consisted of 10 BSyn probands and their biological parents who are obligate carriers of PVs in *BLM* (**Table 1**). Among the 10 BSyn probands, there were equal numbers of males and females, ranging from 10 months to 36 years of age at time of sample collection. Five of the BSyn probands had a history post-collection of at least one type of cancer (**Table 1**). None of the patients had been diagnosed with a hematologic malignancy at time of collection. The *BLM PVs* spanned amino acid 25 to 1243, with most variants clustered in the DEAH Helicase and RecQ Helicase C-terminal domains (**Figure 1A**). Visual inspection of *BLM* variants confirmed family structure (**Figure 1B**).

No significant difference in mean number of total reads was identified between BSyn and control trio samples (**Figure S1C**, t-test, *p*-value=0.268). After mapping to exome targets, BSyn trio samples had a mean coverage of 113.1x compared to controls with 106.5x coverage (**Figure S1D,** t-test, *p*-value=0.018). This coverage consistency extended across all chromosomes (**Figure S1E**).

Samples were categorized into four groups: BSyn proband samples (n=10) designated as “affected”, *BLM* variant carrier samples as “carrier” (n=19), and control proband (n=19) and control parents (n=38) as “unaffected.” In our somatic variant analysis, we separately

considered control parent and control children as age-matched controls to assess the frequency of CHIP in relation to age.

We explored multiple DP cut-offs for loci in CHIP genes (data not shown), and at DPs required to identify high quality, low frequency variants, we found that BSyn probands and *BLM* variant carriers had statistically significantly more low frequency, putative somatic variants (0.02 < VAF < 0.3) in CHIP genes with a median of 2, compared to control cohorts where no somatic CHIP gene variants were detected (median = 0) (Kruskal Wallis, *p*-value=1.50E-06 to 6.37E-03) (**Figure 2A**).

We further categorized variants in CHIP genes into putative somatic or germline based on VAF (**Figure 2B**). Consistently, significant differences were observed across all likely somatic variant comparisons between BSyn groups (model mean = 3.70 - 4.80%) and control groups (model mean = 0.30%) (*p*-value=1.41E-06 to 1.60E-03). No significant differences were found in the mean proportion of germline and somatic variants between BSyn probands and *BLM* carriers (*p*-value=0.447), nor between control probands and control parents (*p*-value=0.991) (**Figure 2B**).

Our analysis identified no significant correlations between mean somatic and germline variants in CHIP genes and the putative somatic subset (**Figure S2A**). Across the four sample groups, we identified no significant difference between mean somatic and germline number of variants in CHIP genes (**Figure 2C**). There were no significant differences identified using mean proportion comparison models in somatic (VAF < 0.3) CHIP variant analysis when comparing the type of variant (**Figure S2B**, Refseq Genes 110, NCBI), pathogenicity (**Figure S2C**, ClinVar 2023-01-05, NCBI), or CHIP genes to which these variants mapped (**Figure S2D**, Refseq Genes 110, NCBI).

One-way ANOVA with random family effect confirmed that all variants (no VAF cutoff) in CHIP genes followed a normal distribution (**Figure S3A**). No significant differences were observed in the type of variant (**Figure S3B**) or pathogenicity (**Figure S3C**). Breakdown of

these variants across all 56 established CHIP genes identified in literature was plotted using a heatmap based on gene of variation (y axis) and by sample (x axis) using hierarchical clustering (**Figure S3D**). The heatmap depicts the distribution of all CHIP variants regardless of VAF identified in the samples, with each row representing a CHIP gene and each column representing a sample. Despite the hierarchical clustering, there are no discernible relationships between the sample cohort and the genes where variants were identified, suggesting a heterogeneous pattern of CHIP gene variants across the samples.

We also explored the influence of age on the number of putative somatic variants in CHIP genes in each cohort (**Figure 2D**). Linear regression analysis identified very weak linear relationships between age and the frequency of putative somatic CHIP variants at VAF<0.3 in our cohorts (R^2 = 1.111E-05 to 0.009).

Lastly, we assessed whether *BLM* PVs affect germline *de novo* rates. Proband B380 was excluded from this analysis as sequencing from only one parent was available. High quality coding variants in probands of each trio (n=9, **Figure S3E**) that were not inherited from either parent were identified, representing *de novo* variants (DNVs, VAF≥0.3). No significant difference in total *de novo* germline variants was found between BSyn and control cohorts, regardless of cancer diagnosis (**Figure 2E**).

This study addresses the impact of *BLM* PVs on the incidence of *de novo* variants and somatic variants. Our findings reveal an increased frequency of low frequency, putatively somatic variants in CHIP genes in BSyn probands and *BLM* carriers, compared to sex- and age-matched controls.

An oral presentation at the American Society of Hematology 2023 meeting identified *BLM* as one of 18 clonal hematopoiesis genes associated with hematopoietic malignancy in the heterozygous state[12], consistent with our CHIP findings. These variants were predominantly synonymous variants in BSyn probands and splice variants in *BLM* carriers. In contrast to prior

studies on clonal hematopoiesis, which have identified somatic variants most frequently in *DNMT3A*, *ASXL1*, and *TET2*[9], we identified mainly synonymous and benign splice variants primarily in *NOTCH1* and *CUX1*.

The absence of significant differences in mean somatic and germline variants between BSyn probands and *BLM* carriers suggests other factors besides *BLM* mutations, such as environmental exposures or other genetic modifiers, may influence mutation patterns. Limitations of our study include small sample size, use of two different exome enrichment methods, and use of exome sequencing to detect ultra-low-frequency clones. Future studies using deep-amplicon sequencing of longitudinal samples could validate these findings.

Our study contributes to the growing literature on increased somatic mutation rate and cancer risk in carriers for genes important in maintaining genomic integrity. These findings may pave the way for early biomarkers in cancer detection and general health assessment in rare disease patients and carriers. Larger-scale studies with BSyn cohorts are imperative to unravel the mechanisms underpinning *BLM* PVs and their contribution to CHIP and cancer risk.

REFERENCES

1. Cunniff C, Bassetti JA, Ellis NA. Bloom's Syndrome: Clinical Spectrum, Molecular Pathogenesis, and Cancer Predisposition. Mol Syndromol. 2017;8(1):4-23.

2. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014;42(Database issue):D980-9855.

3. Sugrañes TA, Flanagan M, Thomas C, Chang VY, Walsh M, Cunniff C. Age of first cancer diagnosis and survival in Bloom syndrome. Genet Med. 2022;24(7):1476-1484.

4. Laitman Y, Boker-Keinan L, Berkenstadt M, et al. The risk for developing cancer in Israeli ATM, BLM, and FANCC heterozygous mutation carriers. Cancer Genet. 2016;209(3):70-74.

5. Bogdanova N, Togo AV, Ratajska M, et al. Prevalence of the BLM nonsense mutation, p.Q548X, in ovarian cancer patients from Central and Eastern Europe. Fam Cancer. 2015;14(1):145-149.

6. de Voer RM, Hahn M-M, Mensenkamp AR, et al. Deleterious Germline BLM Mutations and the Risk for Early-onset Colorectal Cancer. Sci Rep. 2015;5:14060.

7. Thompson ER, Doyle MA, Ryland GL, et al. Exome sequencing identifies rare deleterious mutations in DNA repair genes FANCC and BLM as potential breast cancer susceptibility alleles. PLoS Genet. 2012;8(9):e1002894.

8. Bononi A, Goto K, Ak G, et al. Heterozygous germline *BLM* mutations increase susceptibility to asbestos and mesothelioma. Proc Natl Acad Sci U S A. 2020;117(52):33466-33473.

9. 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.

10. Bick AG, Weinstock JS, Nandakumar SK, et al. Inherited causes of clonal haematopoiesis in 97,691 whole genomes. Nature. 2020;586(7831):763-768.

11. Schratz KE, Haley L, Danoff SK, et al. Cancer spectrum and outcomes in the Mendelian short telomere syndromes. Blood. 2020;135(22):1946-1956.

12. Liu J, Tran D, Chan ICC, et al. Genetic Determinants of Clonal Hematopoiesis and Progression to Hematologic Malignancies in 479,117 Individuals. Blood. 2023;142(Suppl.1):811.

13. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-1303.

14. Coffee B, Cox HC, Bernhisel R, et al. A substantial proportion of apparently heterozygous TP53 pathogenic variants detected with a next-generation sequencing hereditary pan-cancer panel are acquired somatically. Hum Mutat. 2020;41(1):203-211.

15. Steensma DP, Bejar R, Jaiswal S, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. Blood. 2015;126(1):9-16.

TABLES:

Sample ID	Status	Sex	Age Range	Trio Grouping	BSyn Status	HGVS Nomencalture of BLM Mutation (NM_000057.4)	Cancer History (years pc*)
B179	Proband	F	5-9	BLOOM_179	Affected	NM_000057.4:c.[3727_3728insA];[3727_3728insA]	N
C185	Father	M	35-39		Carrier	NM_000057.4:c.3727_3728insA	N
C184	Mother	F	30-34		Carrier	NM_000057.4:c.3727_3728insA	N
B286	Proband	F	1-4	BLOOM_286	Affected	NM_000057.4:c.[1933C>T];[c.3261del]	N
C285	Father	M	30-34		Carrier	NM_000057.4:c.1933C>T	N
C287	Mother	F	25-29		Carrier	NM_000057.4:c.3261del	N
B360	Proband	M	<1	BLOOM_360	Affected	NM_000057.4:c.[2207_2212delinsTAGATTC];[2207_2212delinsTAGATTC]	Y (1x ALL[1] 7y pc)
C353	Father	M	30-34		Carrier	NM_000057.4:c.2207_2212delinsTAGATTC	N
C354	Mother	F	20-24		Carrier	NM_000057.4:c.2207_2212delinsTAGATTC	N
B364	Proband	M	35-39	BLOOM_364	Affected	NM_000057.4:c.[2506_2507del];[2506_2507del]	Y (3x cSCC[2] 1y pc, 1x ALL[1] 4y pc)
C556	Father	M	65-69		Carrier	NM_000057.4:c.2506_2507del	N
C557	Mother	F	60-64		Carrier	NM_000057.4:c.2506_2507del	N
B380	Patient	M	1-4	BLOOM_380	Affected	NM_000057.4:c.[2207_2212delinsTAGATTC];[2207_2212delinsTAGATTC]	Y (1x GIST[3] 17y pc, 1x DLBCL[4] 20y pc)
C381	Mother	F	30-34		Carrier	NM_000057.4:c.2207_2212delinsTAGATTC	N
B409	Proband	M	1-4	BLOOM_409	Affected	NM_000057.4:c.[2207_2212delinsTAGATTC];[2207_2212delinsTAGATTC]	Y (AML[5] 20y pc)
C396	Father	M	25-29		Carrier	NM_000057.4:c.2207_2212delinsTAGATTC	N
C397	Mother	F	20-24		Carrier	NM_000057.4:c.2207_2212delinsTAGATTC	N
B488	Proband	F	1-4	BLOOM_488	Affected	NM_000057.4:c.[275del];[275del]	N
C489	Father	M	25-29		Carrier	NM_000057.4:c.275del	N
C490	Mother	F	30-34		Carrier	NM_000057.4:c.275del	N
B498	Proband	M	5-9	BLOOM_498	Affected	NM_000057.4:c.[2695C>T];[2695C>T]	N
C499	Father	M	30-34		Carrier	NM_000057.4:c.2695C>T	N
C500	Mother	F	35-39		Carrier	NM_000057.4:c.2695C>T	N
B502	Proband	F	25-29	BLOOM_502	Affected	NM_000057.4:c.[1933C>T];[1933C>T]	Y (BLCA[6] 6y pc)
C503	Father	M	45-49		Carrier	NM_000057.4:c.1933C>T	N
C504	Mother	F	45-49		Carrier	NM_000057.4:c.1933C>T	N
B615	Proband	F	1-4	BLOOM_615	Affected	NM_000057.4:c.[2695C>T];[3171_3172insT]	N
C613	Father	M	25-29		Carrier	NM_000057.4:c.3171_3172insT	N
C614	Mother	F	20-24		Carrier	NM_000057.4:c.2695C>T	N

* pc = post collection

[1] ALL = acute lymphoblastic leukemia

[2] cSCC = cutaneous squamous cell carcinoma

[3] GIST =gastrointestinal stromal tumour

[4] DLBCL = diffuse large B cell lymphoma

[5] AML = acute myeloid leukemia

[6] BLCA = bladder carcinoma

Table 1: Bloom Syndrome Patient and *BLM* Pathogenic Variant Carrier Demographics

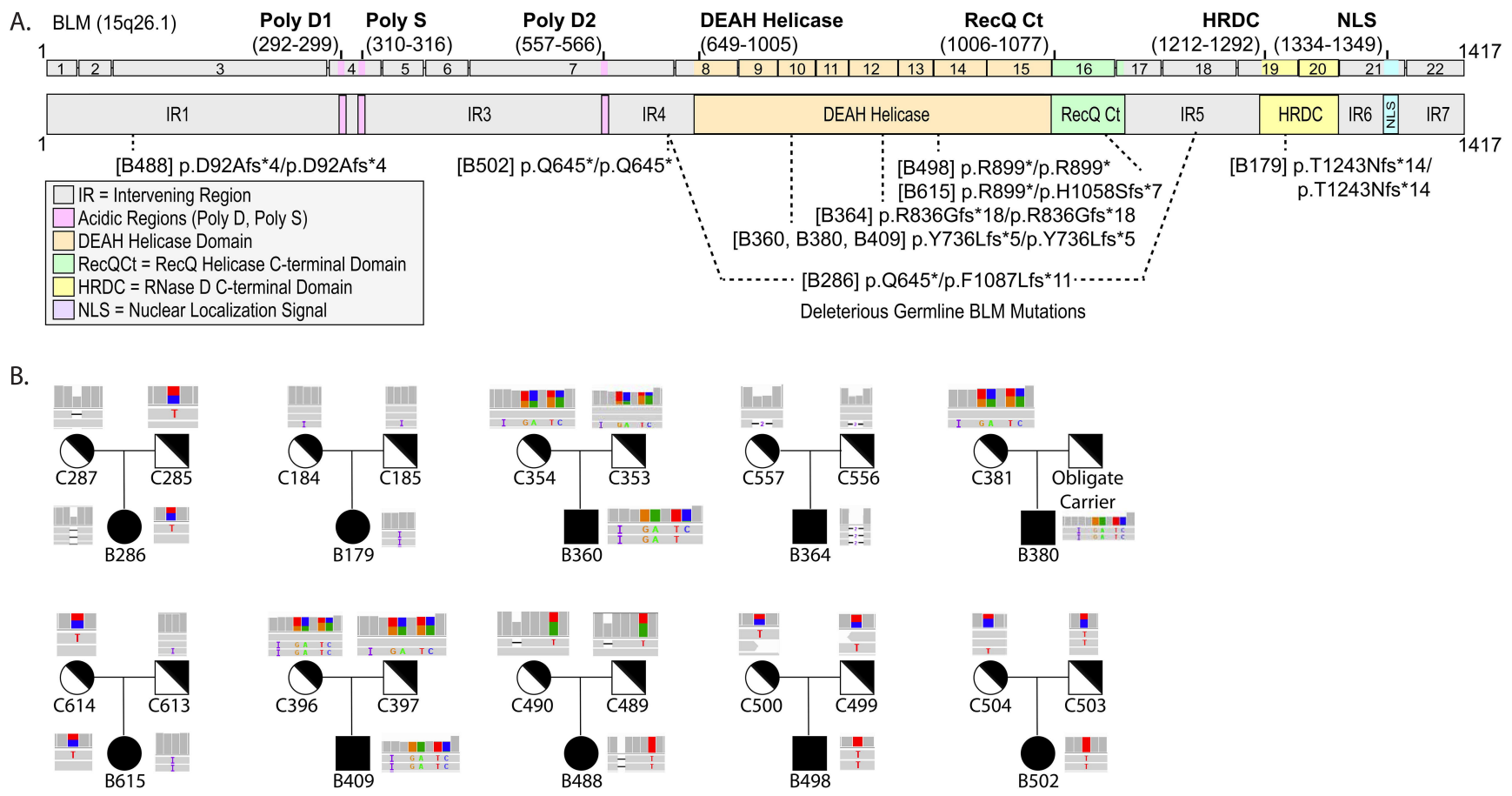
Bloom Syndrome patient (B#, n=10) and *BLM* pathogenic variant carrier (C#, n=19) parents included in our studies are identified with their patient IDs. Individual sex, age range, family grouping, *BLM* pathogenic variant and history of cancer are also shown.

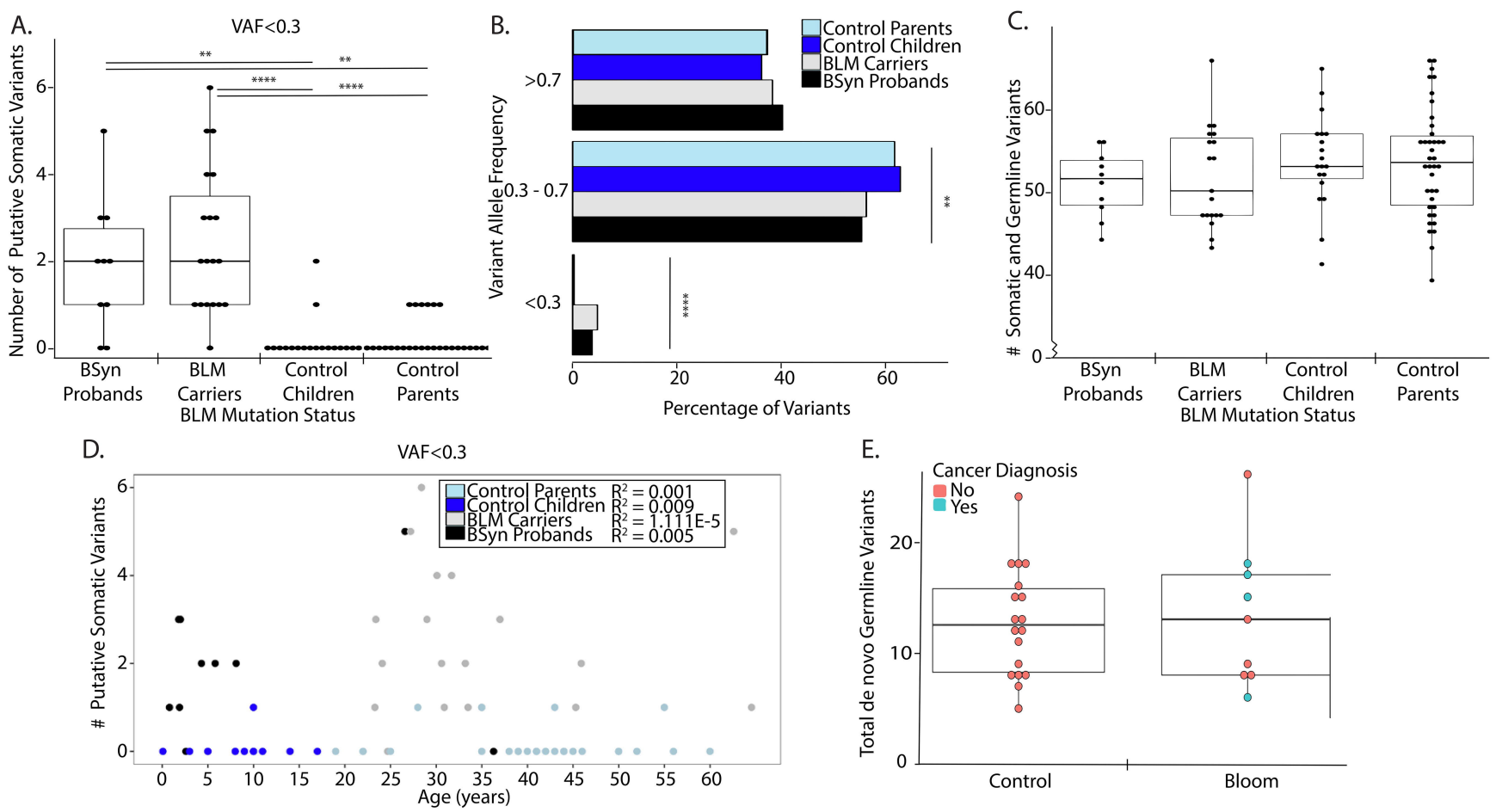
FIGURE LEGENDS

Figure 1: Genomic Analysis of *BLM* Pathogenic Variants in Bloom Syndrome Patients and Carriers. (**A**) Schematic representation of the *BLM* transcript (ENST00000355112.8) and protein (GenBank: BLM; NM000057.4; GRCh38), its functional domains (solid lines above transcript), and pathogenic variants (dotted lines below transcript) causing Bloom Syndrome (BSyn). Variants listed correspond to BSyn probands and are tagged with patient identifiers (B#, see Table 1). Deleterious biallelic pathogenic variants are shown with one dotted line, while compound heterozygous pathogenic variants are shown with two dotted lines. (**B**) Variants in each BSyn proband (B#) and BSyn carrier (C#) verified in *Integrative Genomic Viewer v.2.9.4*. Each *IGV* screenshot shows coverage at the *BLM* variant at the top and the first two to three sequencing reads below with reference bases in grey and genetic variants in color. Histograms represent the coverage around the Bsyn variants at that site. Each trio relationship is depicted.

Figure 2: Somatic CHIP Gene Variants and De Novo Variants in Bloom Syndrome Probands and Carriers. Bloom Syndrome (BSyn) probands (n=10, black), BSyn carrier parents (n=19, grey), control children (n=19, light blue) and control parents (n=38, dark blue). (**A**) Number of putative somatic CHIP variants using a variant allele frequency (VAF) cutoff < 0.3. (**B**) Mean proportions shown for CHIP gene variants subsetted based on VAF grouping. (**C**) Total number of germline and somatic CHIP gene variants. (**D**) Effect of age on number of putative somatic variants in CHIP genes. Linear regression analysis performed separately for each cohort and R^2 values indicate goodness of fit for each model: R^2 = 0.005 (BSyn proband), R^2 = 1.111E-05 (carrier), R^2 = 0.009 (control child), R^2 = 0.001 (control parent). (**E**) Number of germline *de novo* variants (VAF≥0.3) for each sample.

t-test, ns or no stars denote *p*-value>.05, * denote *p*-value≤.05, ** denote *p*-value<.01, *** denote *p*-value<.001, **** denote *p*- value<.0001


A.
BLM (15q26.1)
Poly D1
(292-299)
Poly S
(310-316)
Poly D2
(557-566)
DEAH Helicase
(649-1005)
RecQ Ct
(1006-1077)
HRDC
(1212-1292)
NLS
(1334-1349)
1
1417
IR1
IR3
IR4
DEAH Helicase
RecQ Ct
IR5
HRDC
IR6
NLS
IR7
[B488] p.D92Afs*4/p.D92Afs*4
[B502] p.Q645*/p.Q645*
[B498] p.R899*/p.R899*
[B615] p.R899*/p.H1058Sfs*7
[B364] p.R836Gfs*18/p.R836Gfs*18
[B360, B380, B409] p.Y736Lfs*5/p.Y736Lfs*5
[B286] p.Q645*/p.F1087Lfs*11
[B179] p.T1243Nfs*14/
p.T1243Nfs*14
IR = Intervening Region
Acidic Regions (Poly D, Poly S)
DEAH Helicase Domain
RecQCt = RecQ Helicase C-terminal Domain
HRDC = RNase D C-terminal Domain
NLS = Nuclear Localization Signal
Deleterious Germline BLM Mutations
B.
C287
C285
B286
C184
C185
B179
C354
C353
B360
C557
C556
B364
C381
Obligate
Carrier
B380
C614
C613
B615
C396
C397
B409
C490
C489
B488
C500
C499
B498
C504
C503
B502


A.
VAF<0.3
Number of Putative Somatic Variants
BSyn Probands
BLM Carriers
Control Children
Control Parents
BLM Mutation Status
B.
Control Parents
Control Children
BLM Carriers
BSyn Probands
Variant Allele Frequency
>0.7
0.3 - 0.7
<0.3
Percentage of Variants
C.
Somatic and Germline Variants
BSyn Probands
BLM Carriers
Control Children
Control Parents
BLM Mutation Status
D.
VAF<0.3
Control Parents R² = 0.001
Control Children R² = 0.009
BLM Carriers R² = 1.111E-5
BSyn Probands R² = 0.005
Putative Somatic Variants
Age (years)
E.
Cancer Diagnosis
No
Yes
Total de novo Germline Variants
Control
Bloom

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

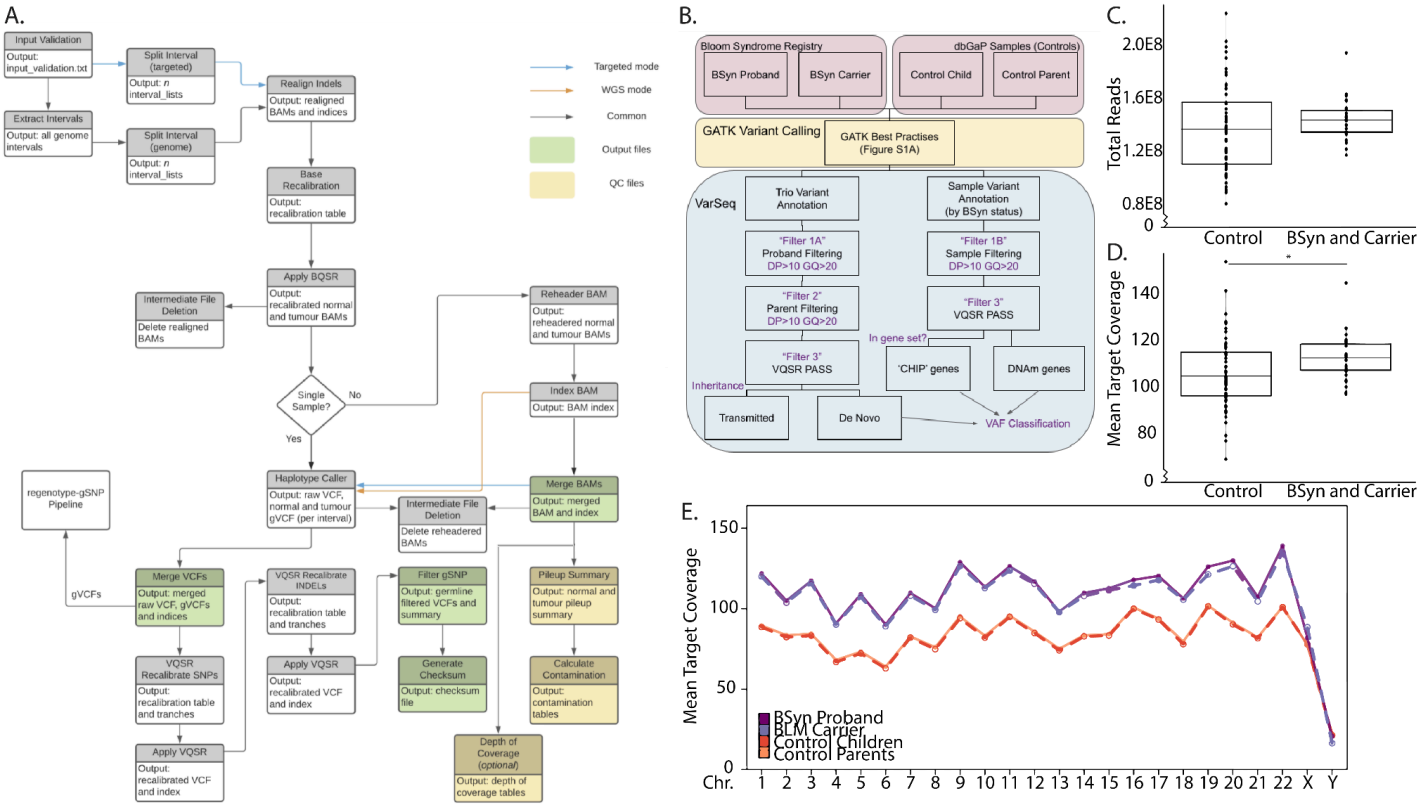


Figure S1: Analysis Pipelines and Sample Sequencing Quality Control.
(**A**) All exome files were processed using the UCLA CDS germline SNP pipeline v5.4.3 which incorporated GATK best practices. Base Quality Score Recalibration (BQSR) and Variant Quality Score Recalibration (VQSR) were performed using 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

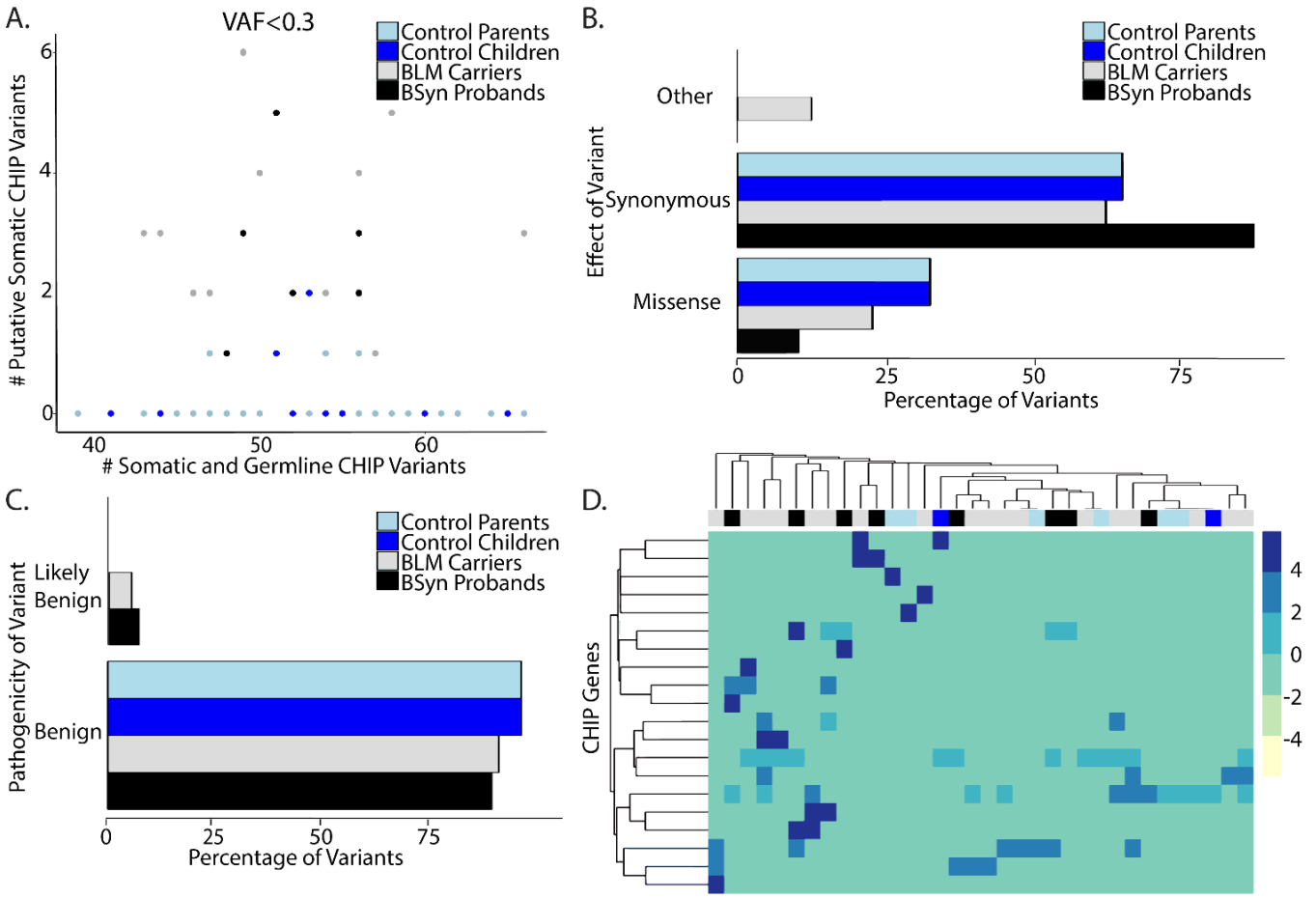


Figure S2: Clonal Haematopoiesis of Indeterminate Potential Variant Load and *De Novo* Variants in Bloom Syndrome Probands and Carriers

(**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.

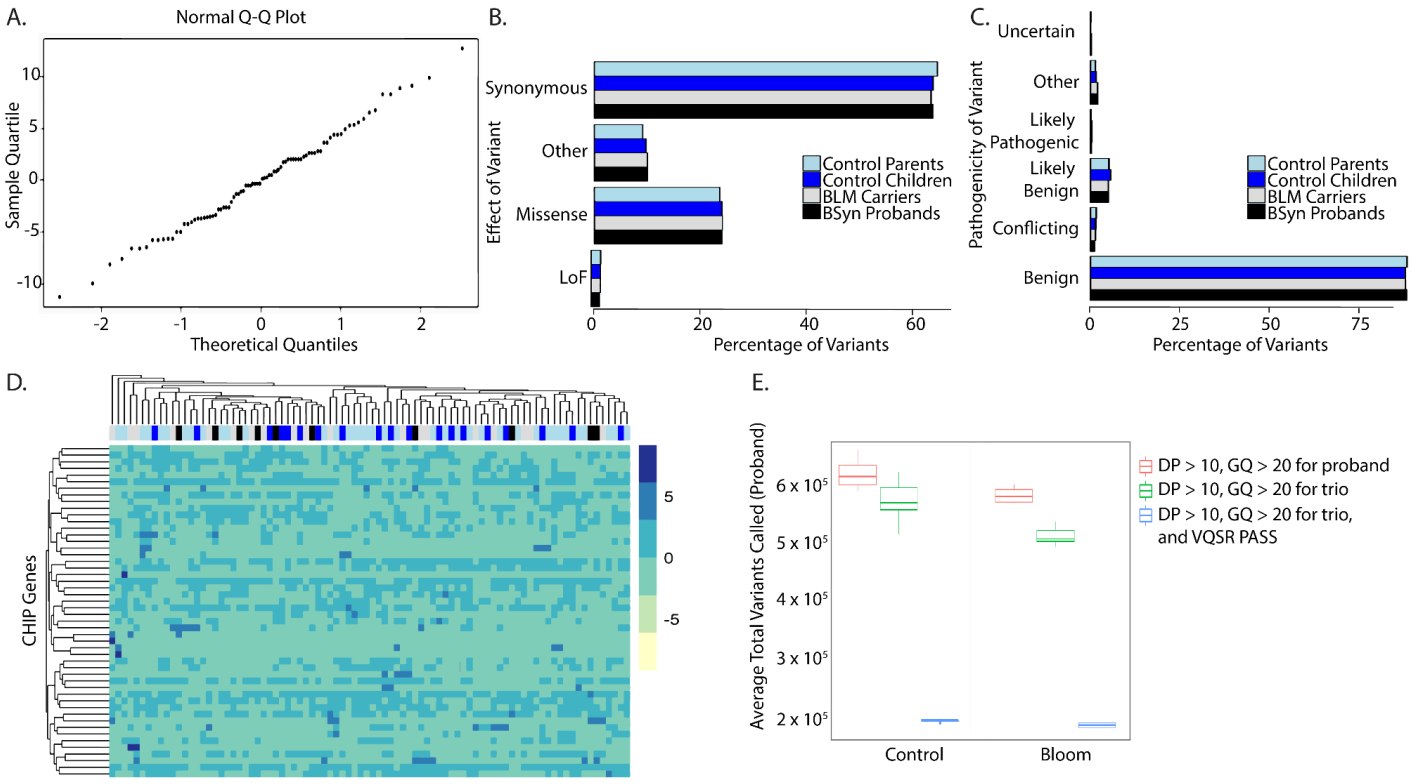


Figure S3: Total Variants in Clonal Haematopoiesis of Indeterminate Potential Genes in Bloom Syndrome and Control Cohorts

(**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.