

Supplementary Appendix

Supplement to: DeBoy EA, Tassia MG, Schratz KE, et al. Familial clonal hematopoiesis in a long telomere syndrome. *N Engl J Med* 2023;388:2422-33. DOI: 10.1056/NEJMoa2300503

This appendix has been provided by the authors to give readers additional information about the work.

TABLE OF CONTENTS

Supplementary Methods	2-7
Author contributions	8
Figure S1. POT1 levels and radiosensitivity in cells derived from <i>POT1</i> mutation carriers	9
Figure S2. T cell receptor repertoire (TCR) V β clonality assessment by sequencing and flow cytometry	10
Figure S3. Age-dependent clonal hematopoiesis rates among <i>POT1</i> mutation carriers	11
Figure S4. Ultra-deep and leukocyte fraction analyses of <i>JAK2</i> and <i>DNMT3A</i> somatic mutations	12
Figure S5. Estimated age of onset of driver clonal hematopoiesis mutation in three individuals	13-14
Figure S6. Time-scaled phylogenies of hematopoietic clones annotated for coding mutations in 87 CH-driver genes	15-16
Figure S7. Somatic single base substitution COSMIC signature per clone in two <i>POT1</i> mutation carriers and a related control who does not carry the mutation	17
Figure S8. Simulated dynamic of mean variant allele frequency of single heterozygous driver mutation acquired at birth and over a lifetime	18
Figure S9. Mean variant allele frequency of a clonal hematopoiesis driver mutation acquired at birth, across 10,000 simulations for each of four groups graphed separately	19
Figure S10. Proportion of simulated individuals with clonal hematopoiesis of indeterminate potential (CHIP).....	20
Figure S11. Read alignment and unfiltered variant call statistics for whole genome sequence data	21
Figure S12. Summary of variant filtering data for Individuals A, B and C, respectively	22-25
Figure S13. Distribution of somatic variants	26
Table S1. <i>POT1</i> mutations and their prevalence in germline and somatic databases	27
Table S2. Onset of hair graying in adults with <i>POT1</i> mutations	28
Table S3. Somatic variants identified in 12 <i>POT1</i> mutation carriers and 21 <i>POT1</i> non-carrier relatives	29
Table S4. Somatic variants in 30 controls over age 70	30
Table S5. Primer sequences	31
Table S6. List of genes analyzed for somatic mutation studies and their accession identifiers ...	32
Supplementary References	33-34

Supplementary Methods

Germline genotyping. *POT1* variants were genotyped from DNA isolated from peripheral blood by PCR amplification and Sanger sequencing. For deceased individuals, sequencing was performed on archived formalin fixed paraffin embedded tissue. The *JAK2* ‘GGCC’ haplotype block was inferred by sequencing rs12343867, which is in linkage disequilibrium with three risk polymorphisms^{1,2}. Primers are listed in Table S5. Variants were deposited in the ClinVar Annotation Database.

Radiosensitivity. Lymphoblastoid cell lines (LCLs) were derived from peripheral blood mononuclear cells by Epstein Barr virus immortalization³. Sensitivity to ionizing radiation was assayed by irradiating 3000 LCLs/well in four replicates using a Gammacell 40 Extractor as described^{4,5}, and cells were counted at 7 days (Cell Counting Kit-8, Dojindo Molecular Technologies) to assess viability.

***POT1* immunoblot and quantitative real time PCR.** LCLs were lysed in RIPA lysis buffer (Cell Signaling Technology) supplemented with protease inhibitor (Roche) and benzonase nuclease (Millipore). Protein was resolved on Bolt 8% Bis-Tris gels with MOPS-SDS buffer and transferred to Polyvinylidene difluoride membrane using the iBlot 2 Dry Blotting System (Thermo Fisher Scientific). Membranes were pre-treated with SuperSignal Western Blot Enhancer then blocked with Intercept blocking buffer (LI-COR). In vitro transcribed *POT1* (NM_015450) was generated as described⁶. Primary antibodies details are as follows: *POT1* (rabbit, NB500-176, 1:500, Novus), Myc (mouse, clone 4A6, 1:1000, Millipore), and β -tubulin (rabbit, ab6046, 1:100,000, Abcam). *POT1* secondary antibody was conjugated to horseradish peroxidase (HRP) (Goat, 7074, 1:10,000; Cell Signaling Technology), and blots were visualized by chemiluminescence (SuperSignal West Pico PLUS, Thermo Scientific). Other secondary antibodies were conjugated to IR680 or IR800 (donkey, 1:10,000, LI-COR) and blots were visualized by LICOR’s Odyssey scanner and quantified as previously described⁷. *POT1* cDNA levels were measured by quantitative reverse transcription PCR using the SYBR green method as described⁸ (primers listed in Table S5).

***POT1* functional studies.** We tested the impact of *POT1* missense mutations on telomere DNA binding in a gel shift assay as described⁶. To assess the effect of splice junction variants, LCLs were incubated with cycloheximide diluted in dimethyl sulfoxide-containing medium (50 μ g/mL, Cell Signaling) as described⁹. cDNA was amplified using exon-spanning primers and products were gel isolated for Sanger sequencing (primers listed in Table S5).

***T* cell receptor (TCR) diversity.** TCR-V β Complementarity Determining Region 3 (CDR3) repertoire diversity was examined using the immunoSEQ Survey Assay (TCRv4b, Adaptive Biotechnologies)¹⁰, and clonality metrics were obtained using immunoSEQ Analyzer 3.0. TCR clonality was plotted relative to data for healthy bone marrow donors with known cytomegalovirus serostatus (data were derived from Cohort 1 in Emerson et al.¹¹). Percentile lines were derived as described previously¹².

Immunophenotyping by flow cytometry for B and T cell clonality. Immunophenotyping was performed on ficoll-separated peripheral blood mononuclear cells at the Johns Hopkins Flow

Cytometry Laboratory. Cell suspensions were incubated with monoclonal antibody combinations (BD Biosciences) that were used at concentrations titrated for optimal staining. The panel included antibodies for CD3, CD19, CD200, CD56, CD45, CD10, CD20, surface kappa and lambda light chains, and CD38. Some specimens were subjected to extended antibody panels, which included polyvalent kappa and lambda light chains, CD2, CD7, CD4, CD34, CD5, CD8, CD27, and CD28. An antibody directed at the TCR- β chain constant region 1 was added to identify isotypes (C β TCR; clone JOVI-1, BD Biosciences). Specimens were analyzed on the BD FACSCanto 10-color system (BD Biosciences). List mode data files were acquired and analyzed for each specimen using FACSDiva (BD Biosciences) and Infinicyt (Cytognos), respectively. An antigen was considered positive if the cell population of interest showed a homogeneous distribution with the median intensity at least 2 log channels above that seen in the control, or if there was a heterogeneous distribution of antigen expression, such that a subpopulation of cells was above that seen in the control.

Clonal hematopoiesis next-generation sequencing from bulk DNA. Somatic variant sequencing was performed using a customized clinical pipeline at the Johns Hopkins Pathology Laboratories. The panel included 87 genes implicated in hematologic malignancies (Table S6). Libraries were prepared using xGen Dual Index UMI adaptors and hybridized to an IDT DNA probe set (Integrated DNA Technologies). Final libraries were generated using IDT xGen Hybridization blockers and Wash Kit and treated with blocking reagent (Illumina Free) to reduce sequencing aberrancies. Sequencing was performed on NovaSeq 6000 (S1 Reagent Kit v1, 200 cycles, Illumina). Mean deduplicated consensus coverage was 612x (range 326-1151) with 96% on target at 150x or greater depth. Reads were aligned to GRCh37 using BWA v0.7.17. Variants were called and annotated using a custom pipeline, cross referenced with HaplotypeCaller (GATK 3.3), and annotated with ANNOVAR (v10122020)¹³. Data from septuagenarian controls were from the Women's Health and Aging Study (WHAS), and the sequencing and analysis of variants were performed using the same clinical pipeline as for the remaining samples, except the depth of coverage was two-fold higher (mean 1336x).

Somatic mutation analysis. Non-synonymous and canonical splicing variants with variant allele frequency (VAF) greater than or equal to 0.01 and minimum 50x coverage were retained. Variants with VAF of either 0.35-0.65 (heterozygous calls) or >0.95 (homozygous calls) were excluded as likely germline polymorphisms. Recurrent technical artifacts were excluded if they were present in clinical validation samples that were sequenced on the same platform. Filtered variants were manually inspected in Integrated Genome Viewer (IGV), and included if they were present on at least 5 unique molecular bar codes (MBC) with all reads for each MBC containing the variant. Variants that had minor allele frequency >0.001 in gnomAD v2.2.1 were also excluded as potentially germline. Two of the authors generated and verified each variant manually in IGV.

DNMT3A and JAK2^{V617F} targeted genotyping and droplet digital PCR (ddPCR). To assess for low frequency somatic JAK2^{V617F} mutations, we performed targeted deep sequencing (mean coverage 5100x, range 2800-10115) using the Ion AmpliSeq Cancer HotSpot Panel v2 (Life Technologies). Amplified products were sequenced on Ion S5 XL (Thermo Fisher). To test the lineage of origin, we performed droplet digital PCR (ddPCR) on DNA isolated from leukocyte fractions (EasySep Human Myeloid Positive Selection and Human T cell Enrichment kits, Stem

Cell Technologies) as previously described⁶. Primers are listed in Table S5. Somatic *DNMT3A* (NM_022552) strand phasing was performed by cloning and sequencing products of cDNA amplification. cDNA was derived from LCLs and the amplicon encompassed the two *DNMT3A* somatic mutations (550 base pairs apart).

Whole genome sequencing of hematopoietic colonies. Ficoll-separated peripheral blood mononuclear cells were plated at a concentration of 2×10^5 cells per 3 cm tissue culture dish in Methocult media following manufacturer's instructions (Stem Cell Technologies, Vancouver) and incubated at 37°C. Single erythroid colonies were plucked on day 14 after seeding; these erythroid were chosen for their larger size and higher DNA content. DNA was extracted using MagMAX DNA Multi-Sample Ultra kit (Applied Biosystems). Libraries were prepared from 10 ng of DNA derived from each single colony (sample extracted from whole blood was used as a 'bulk' sample) using the Qiagen DNA FX library preparation kit with unique dual indexes (UDI). Fragmentation was run for 16 minutes with no enhancer. Adapters were diluted 1:10 and samples were PCR amplified for 10 cycles. Paired-end whole genome sequencing was performed (150 bp) at 15x targeted depth of coverage on an Illumina NovaSeq v1.5 kits (S4 and S2, 300 cycle). We sequenced 50 colonies in addition to the bulk DNA sample for 3 individuals. A single colony sample failed to sequence and was removed from subsequent analyses. Raw sequencing reads are under controlled access and available via the NCBI's dbGaP accession phs003207.v1.p1.

Read alignment. Raw reads for the whole genome sequencing data were aligned to the GRCh38 human reference genome assembly (GRCh38.d1.vd1; obtained from the NIH National Cancer Institute Genomic Data Commons) using bwa-mem¹⁴. File conversion, indexing, sorting, and removal of PCR duplicates were performed with the SAMtools suite¹⁵. Read alignment statistics are included in Fig. S11A.

Variant calling. As each sequenced clone was experimentally designed to represent a homogenous population of cells derived from a single hematopoietic progenitor; as such, genotyping was performed using the GATK HaplotypeCaller algorithm¹⁶ to produce gVCF files for each clone. Variant calls were merged across all clonal samples using the GATK CombineGVCFs command, and joint genotyping was performed using the GenotypeGVCFs command to produce variant call files composed of both germline and somatic mutations for every clone from a given donor. Bulk peripheral blood samples (henceforth, "germline matched reference") were also produced and analyzed independently of clone samples. Aggregated variant calls for all the whole genome sequence data are shown in Fig. S11B.

Variant call filtering. We developed a stringent variant filtering strategy with the goal of isolating and retaining high-confidence somatic mutations while removing germline variation and spurious technical artifacts. This somatic mutation filtering heuristic is modified from previous work^{17,18} to leverage the clonal sequencing protocol presented above. Each filter is described below:

- *VQSR*: We applied the GATK VariantQualityScoreRecalibrator (VQSR) to filter sites for technical artifacts which could be inferred from base quality and read-mapping statistics, according to GATK best practices (<https://github.com/gatk-workflows/broad-prod-wgs->

[germline-snp-indels.git](https://github.com/germline-snp-indels)) with the following modifications tailored to our somatic clone sequencing scenario: 1) Excess heterozygosity was excluded from the metrics under consideration, as this measurement relies on assumptions of Hardy-Weinberg Equilibrium, which are violated by our study design; and 2) `BaseQRankSum` (z-score from Wilcoxon rank sum test of alternate vs. reference base qualities) was added as a metric under consideration.

- *GermlineHet*: Heterozygous sites present in each individual's matched germline reference (which met the AlleleBalance filter of 0.25, detailed below) were removed.
- *CommonPopVariant*: Common germline polymorphisms observed in the population (allele frequency ≥ 0.01) were removed using the gnomAD v3.1.2¹⁹ variant database as reference.
- *MissingData*: Removed sites where over 20% of samples lacked genotypes.
- *UnreliableSite*: The 1000 Genomes Project genome accessibility mask²⁰ was used to remove variants in regions of low technical accessibility.
- *MultiAllelic*: Multiallelic variants were removed.
- *HomAlt*: Variant sites with an allele frequency of 1.00 were removed.
- *HighHet*: Variants where over 80% of genotyped samples were called as heterozygous were removed.
- *ClusteredEvents*: Variants within 10 bp of any other variant were removed.
- Two depth filters were applied:
 - *DepthFilter*: Removed sites where $>80\%$ of heterozygous samples possessed sequence depths below the 10th percentile of coverage, given the mean sequencing depth of clones from that donor and assuming a Poisson distribution.
 - *AlleleBalance*: An allele balance filter to remove sites where fewer than half of heterozygous samples (which met the depth filter) had an allele balance of at least 1/3.

Fig. S12A-C show the respective variant filtering statistics for each individual studied. Post-filtering summaries of somatic variant distributions are included in Fig. S13. All somatic variants per clone were annotated using GATK's Funcotator tool (Fig. S7).

Phylogenetics. SNP sites which passed all filtering steps were used to infer the phylogenetic relationships of the clones with maximum-likelihood using IQTree v.2.2.0²¹. The maximum-likelihood topology was inferred with the best-fit substitution model using ModelFinder²², and node support values were estimated using 10,000 ultrafast bootstrap replicates²³ with nearest-neighbor interchange to avoid overestimating branch lengths due to severe model violations. All nodes under 95% ultrafast bootstrap support were collapsed.

Using the topologies generated above, branch lengths were transformed to reflect the acquisition and accumulation of somatic SNPs (Figure 4) using the maximum-likelihood algorithm implemented in 'treemut' (<https://github.com/NickWilliamsSanger/treemut>; Williams et al. 2022¹⁸). Ultrametric branch length adjustments to visualize time-dependent somatic SNPs acquisition (Fig. S5) were estimated using the 'iterated reweighted means approach' detailed in Mitchell et al. 2022¹⁷. Unedited tree files with node supports presented as '[SH-like approximate likelihood ratio test value] / [ultrafast bootstrap support percentage]' and branch length estimations measured in substitutions per site are available on GitHub

(https://github.com/mccoyle-lab/pot1_ch_2022). Visualization of phylogenetic results was performed using ape (version 5.6-2)²⁴, and TreeTools (version 1.8.0) libraries for R (version 4.2.0).

Mutation signature assignment. SigProfilerAssignment

(<https://github.com/AlexandrovLab/SigProfilerAssignment.git>) was used to assign single-base substitution (SBS) signatures from The Catalog of Somatic Mutation in Cancer (COSMIC; Alexandrov et al. 2020²⁵) to the collection of somatic mutations per clone, independently. Each inferred SBS profile was then reconciled with all other samples from the respective individual and visualized as SBS signature count distributions (Fig. S7). Mutation signature assignments were calculated directly from the somatic SNV calls and do not account for shared ancestry among clones. As such, mutation burden reflected in the SBS signature (Fig. S7) is elevated relative to those calculated using phylogenetic trees in Fig. 5 (which used root-to-tip distance following assignment of SNVs to branches using ‘treemut’).

In silico simulation. We used simulations to test the role of telomere length (TL) on the longevity of a single heterozygous gain-of-function somatic mutation arising in one of 100,000 hematopoietic stem cells at birth. The simulation was run for 32,850 days, reflecting a 90-year lifespan, and for each condition, 10,000 simulations were run to capture variability. We simulated four TL backgrounds, each representing the value at a certain percentile as defined by flowFISH in the human population. These TL backgrounds were at the 1st (8.6 kb), 50th (11.0 kb), and 99th (13.4 kb) population percentiles¹², as well as a *POT1* mutant background with 13.4 kb TL. To reflect heterogeneity within the stem cell population’s TL at birth, we applied a TL standard deviation of 1.25 kb, as estimated by single colony TL measurements by whole genome sequencing in Mitchell *et al.*¹⁷. A 100 bp/division telomere shortening rate in hematopoietic cells was utilized based on¹⁷, and 50 bp/division gain in the *POT1* mutant group was estimated based on the data in Figure 5D (other permutations were also considered in the simulation in Fig. S8B).

We constructed our simulation using the software SLiM²⁶. Within each day of the simulation, each cell underwent one of two fate decisions: division or non-division, where division could be asymmetric (producing a differentiated daughter cell), symmetric (creating an additional progenitor), or differentiation (dropout from the progenitor population). Cells also acquired a TL change after each division. This introduces an element of stochastic drift, whereby cells carrying a driver mutation may drop out or never divide regardless of their fitness advantage. All cells also acquired a TL change after each division, drawn from a normal distribution to further incorporate variability in TL distributions. Finally, every cell with TL <4 kb experienced cell death with a probability dependent entirely on TL, with likelihood of death increasing linearly from 0 (certain survival) to 1 (certain death) between 4 kb and 2 kb (based on data in¹²).

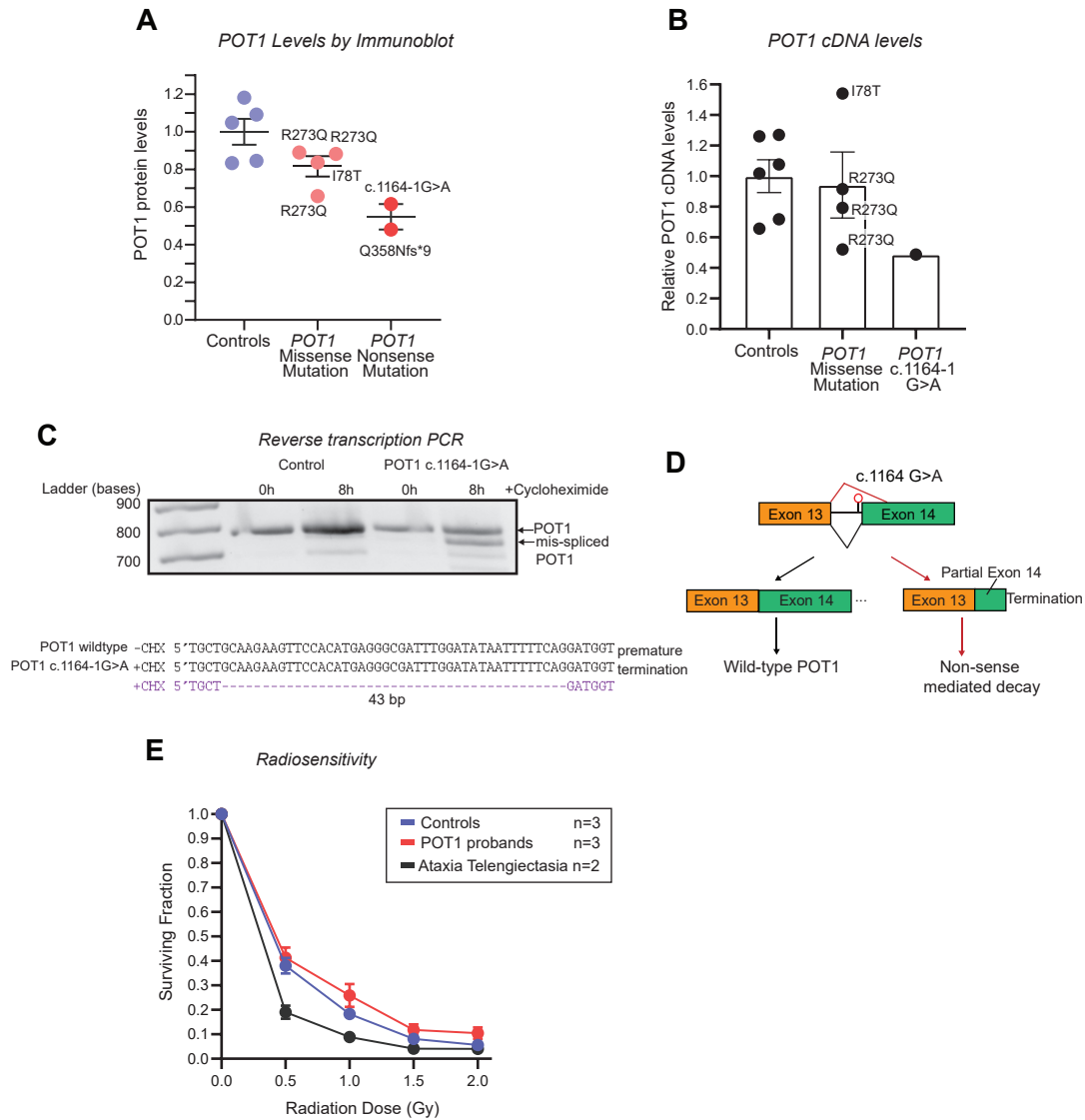
Premises for the model were as follows: The number of hematopoietic progenitors at birth was based on¹⁷. We set the progenitor division rate at once every 280 days (daily division probability = 1/280), as described in¹⁷. Progenitor cells with the driver clonal mutation had a higher probability of entering the cell cycle (1.1x daily division probability), a conservative rate based on literature estimating the replicative advantage of cells with *JAK2* and *DNMT3A* hotspot mutations (^{27,28}and references therein). Probabilities of fate decisions were derived from²⁹ as follows: 0.9 asymmetric division, 0.0525 symmetric division, and 0.0475 differentiation. In cells

with driver mutations, the probabilities of asymmetric, symmetric division and differentiation were 0.8, 0.14, and 0.06, respectively, as per²⁹. The simulation was limited to a single mutation acquired at birth and did not account for secondary mutations that may have been gained over a lifetime. Code for the simulation is available at https://github.com/mccoyle-lab/pot1_ch_2022.

Author contributions

EAD performed *POT1* variant genotyping, molecular studies (Western blot, qPCR, ddPCR) and analyzed bulk CH and TCR-seq data. MGT analyzed whole genome sequence data and performed phylogenetic analyses. KES analyzed CH data and procured and analyzed clinical data. SMY performed the simulations. ZLC performed germline genotyping and gel shift/splicing experiments. EJM analyzed TCR-seq data. DLG analyzed germline sequence variants and provided input on molecular studies. ZX performed telomere length and ddPCR experiments. DBL and ESA evaluated and analyzed clinical data. CDG analyzed bulk CH and flow cytometry data. RCM analyzed and oversaw the computational studies for the phylogenetic and simulation analyses. MA designed the study, analyzed the data, oversaw the project and wrote the manuscript. All the authors reviewed the manuscript, provided input and approved its final form.

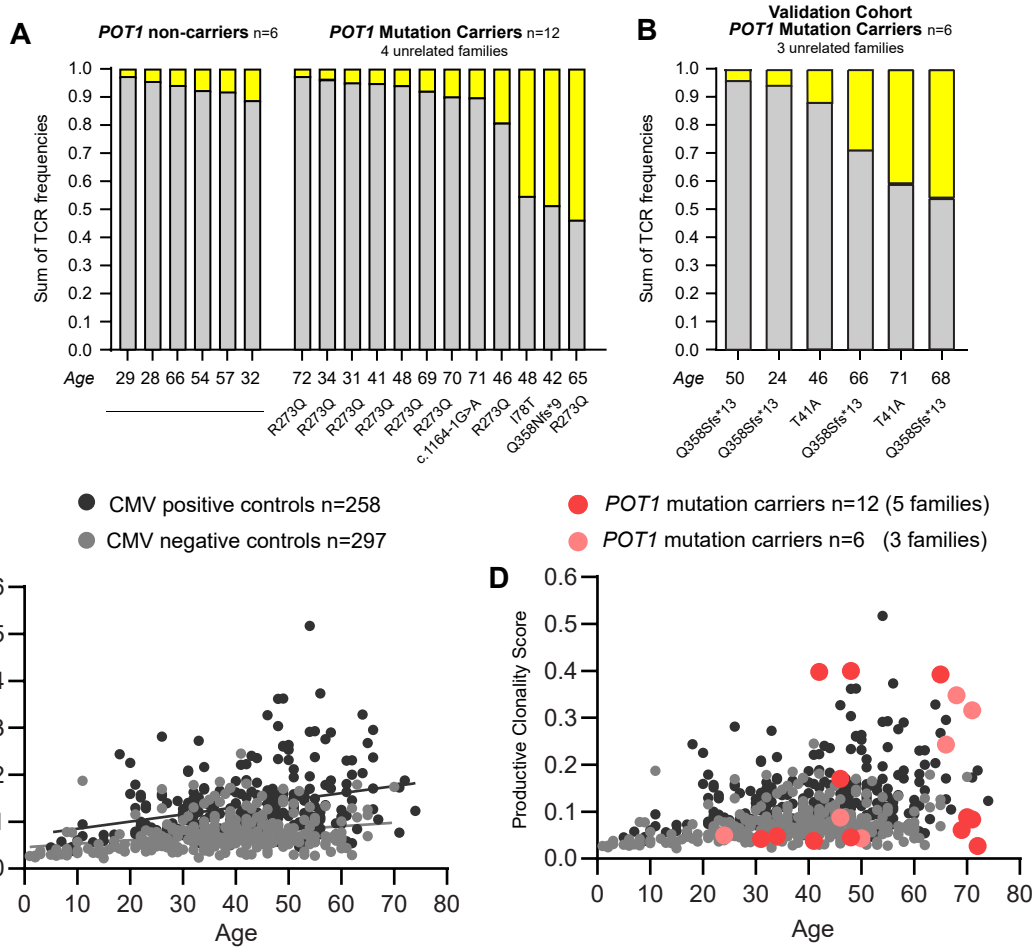
Figure S1



POT1 levels and radiosensitivity in cells derived from *POT1* mutation carriers. **A.** Endogenous POT1 protein levels in lymphoblastoid cell lines (LCLs) derived from *POT1* mutation carriers relative to unrelated controls measured by immunoblotting. Data are normalized relative to β -tubulin based on 2-6 blots for each individual. **B.** Quantitative reverse transcription polymerase chain reaction (qRT-PCR) for *POT1* cDNA. Data were replicated in a second experiment. **C.** Reverse transcription products from LCLs from a control and *POT1* mutation carrier with a splice junction variant prior to and after treatment with cycloheximide, an inhibitor of nonsense mediated decay. The scheme in **D** shows the effect of mis-splicing on cDNA based on sequence analysis of gel-extracted products. **E.** Surviving fraction of LCLs derived from controls and *POT1* mutation carriers after increasing doses of gamma irradiation. LCLs from individuals with ataxia telangiectasia show the typical radiosensitivity.

Figure S2

■ Frequency of top 10 TCR clones ■ Remaining TCR repertoire



● CMV positive controls n=258
● CMV negative controls n=297

● *POT1* mutation carriers n=12 (5 families)
● *POT1* mutation carriers n=6 (3 families)

E

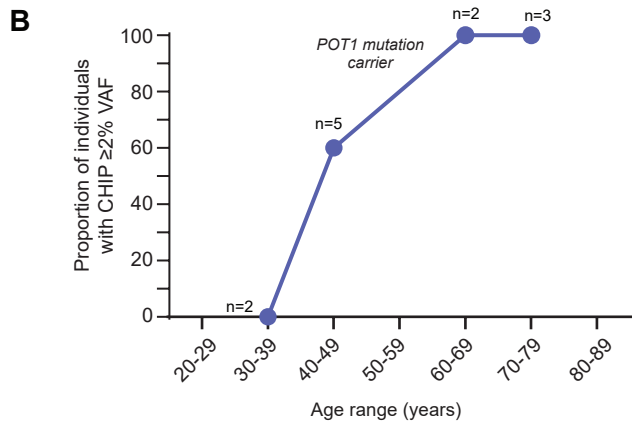
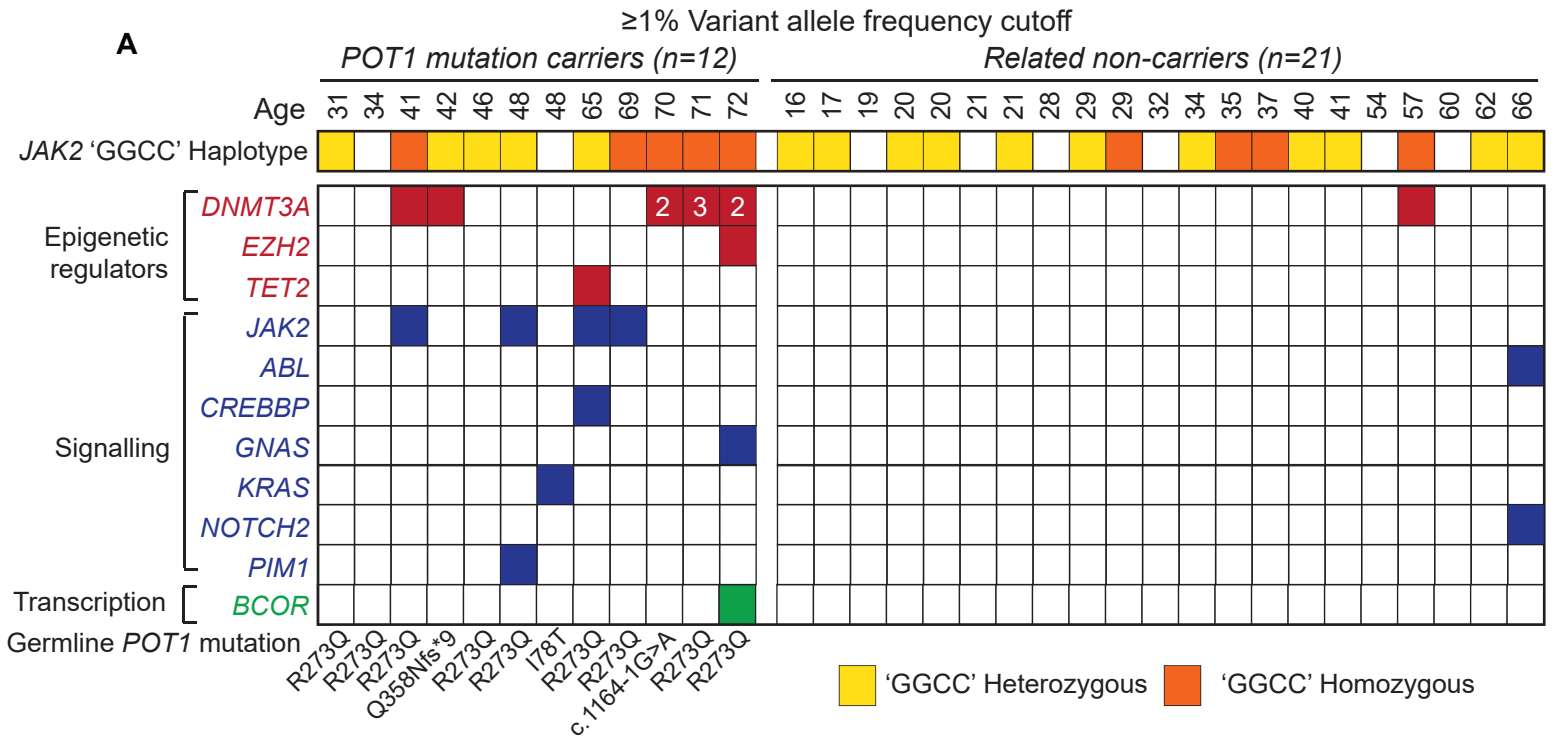
Age	M/F	<i>POT1</i> mutation	Telomere length	Lymphoid Neoplasm diagnosis?	TCR-seq clonality score relative to age	T cell/ B cell Clone?	T cell clone Immunophenotype	% of T cells	B cell clone immunophenotype	% of B cells
42	F	p.Q358Nfs*9	90-99%ile	No	>99%ile	Abnormal T cells Abnormal B cells	CD3+CD5+CD8+ (No <i>TCRB1</i> staining, cannot rule out gamma/delta T cells)	1%	Lambda predominant population but obscured by high background staining	Cannot assess definitively
52	M	p.I78T	99%ile	CLL CTCL	>99%ile	Clonal T cells Clonal B cells	CD4+CD3dimCD5dim CD7+CD26bright (>95% <i>TCRB1</i> skewed)	14%	k+CD19+CD20dim, CD5+CD200+ (CLL)	>99%
65	M	p.R273Q	99%ile	No	>99%ile	Normal T cells Abnormal B cell	-----	NA	kappa predominant (k:l ratio 5:1) CD19+CD20+CD5-CD10-CD38-	80%
69	F	p.Q358Sfs*13	>99%ile	No	>99%ile	Clonal T cells (2) Clonal B cells	CD4+CD7dimCD26- (>95% <i>TCRB1</i> skewed) CD8+ suspicious population (89% <i>TCRB1</i> staining)	29% 2%	k+CD19+CD20+CD10-CD5- CD38-	14%
71	M	p.T41A	>99%ile	No	90-99%ile	Abnormal T cells Clonal B cells	CD8+CD7+CD5+CD3+ (<i>TCRB1</i> 89% skewed) CD3-CD5+CD4+ (Note)	7% 1.4%	kdimCD20+CD19+ CD5-CD10-	33%

Note: Small but distinct CD3 negative TRBC1 positive population

T cell receptor repertoire (TCR) Vβ clonality assessment by sequencing and flow cytometry. A and B.

Each bar represents the entire TCR repertoire from a single individual with the yellow indicating the proportion occupied by the top most frequent clones (yellow). Age at the time of analysis and each *POT1* mutation are shown below for each bar. Data are shown for 18 *POT1* mutation carriers from 7 families and relatives in panel A. C. Productive clonality scores for cytomegalovirus (CMV) sero-positive and sero-negative controls with linear regression line shows higher clonality among CMV positive controls (data generated from Emerson et al. *Nature Genetics* 2017). D. Independent of whether *POT1* mutation carriers are plotted relative to CMV positive or negative controls, a subset of them shows higher clonality. Red and pink dots refer to original and validation groups, respectively as in the key. E. Flow cytometry analysis of peripheral blood among 5 *POT1* mutation carriers with high TCR-seq clonality scores performed by a clinician unaware of individual status. Atypical T cell and B cell populations even among individuals with no known hematologic malignancy diagnoses are identified. TRBC1 refers to an antibody against the TCR-β chain constant region 1.

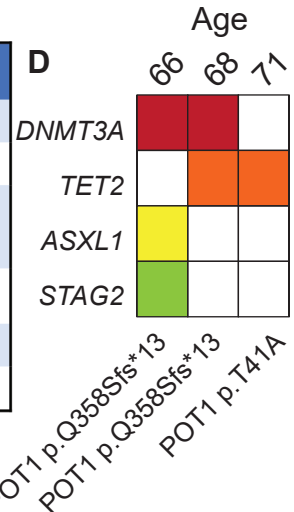
Figure S3



Age-dependent clonal hematopoiesis rates among *POT1* mutation carriers. **A.** Co-mutation table showing data for clonal hematopoiesis mutations with ≥1% variant allele frequency (VAF) among *POT1* mutation carriers and their relatives. Mutations are annotated in Table S3. **B.** Proportion of 12 *POT1* mutation carriers from five unrelated families with clonal hematopoiesis mutation graphed by age group (≥2% VAF threshold). **C.** CHIP mutations among three additional *POT1* mutation carriers from two unrelated families with germline mutations listed below in **panel D**. Data for ≥2% VAF threshold are shown in **panels C and D**.

C

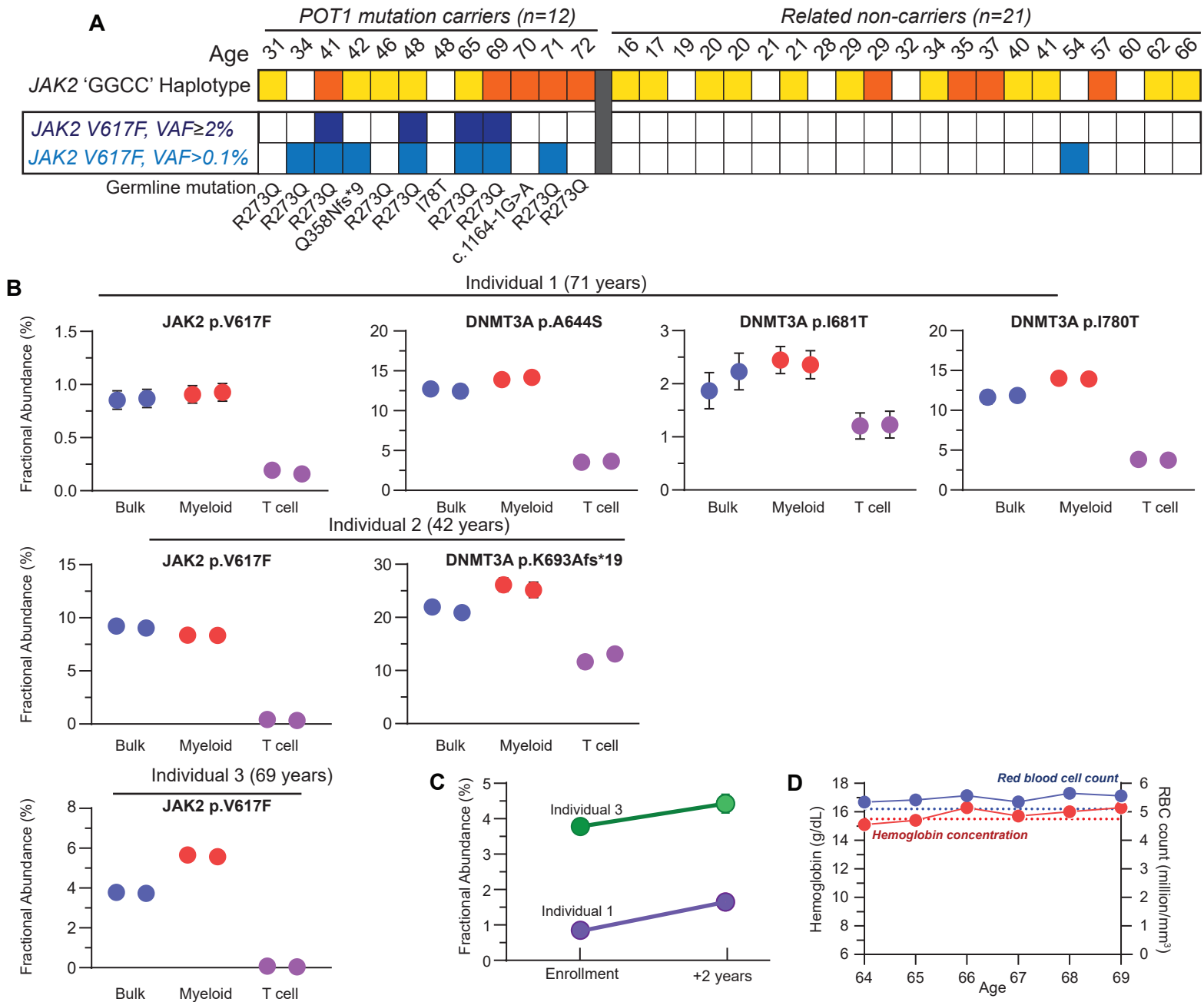
Gene	Protein change	Coding change	Chromosome Position	Reference Allele	Alternate Allele	VAF (%)	COSMIC v90
<i>ASXL1</i>	p.W1411X	c.4233G>A	chr20:31024748	G	A	4.95	AML
<i>DNMT3A</i>	p.F608Sfs*43	c.1823delT	chr2:25467051	GA	G	5.61	---
	p.R771Q	c.2312G>A	chr2:25463181	C	T	24.4	AML, Mast cell neoplasm, various solid
<i>STAG2</i>	p.R1205X	c.3613C>T	chrX:123229240	C	T	10.03	AML, Breast, Prostate
<i>TET2</i>	p.G1282Wfs*18	c.3843dupT	chr4:106180814	G	GT	5.49	---
	p.W1291R	c.3871T>A	chr4:106180843	T	A	13.67	---



*Coordinates shown are from UCSC version hg19 (NCBI build GRCh37).

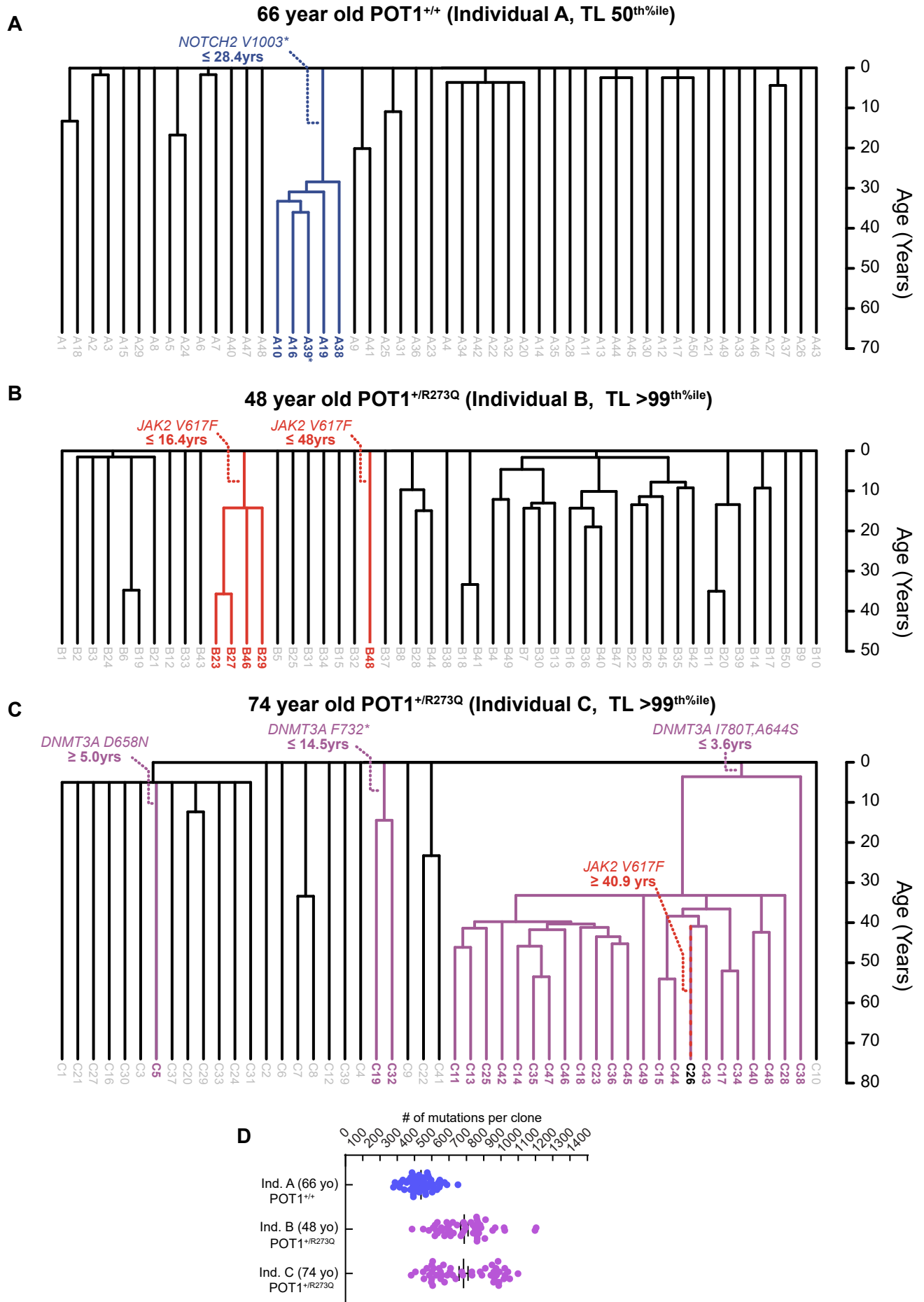
**Previously identified mutations in COSMIC are listed from top to bottom in order of prevalence. AML, acute myelogenous leukemia

Figure S4



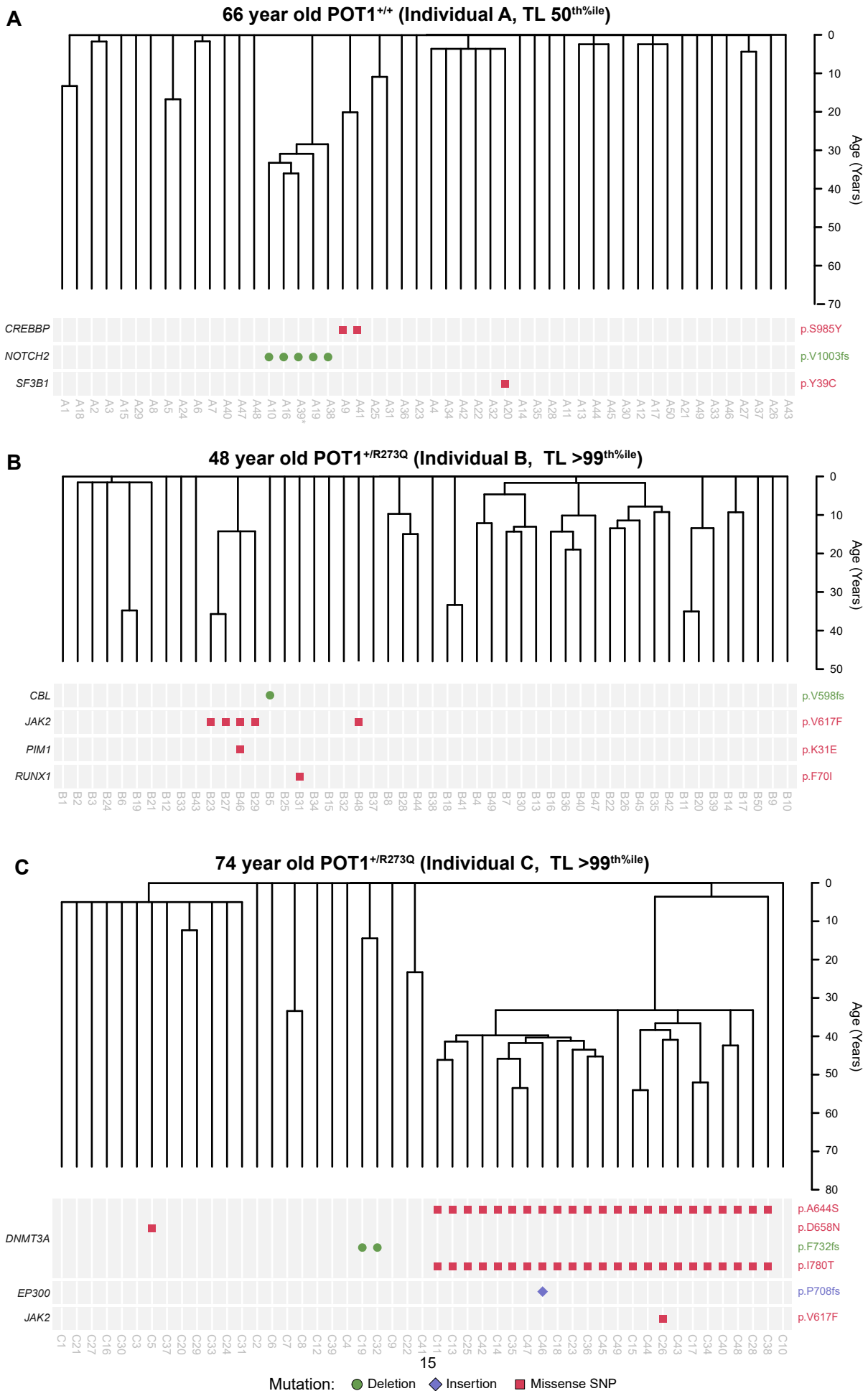
Ultra-deep and leukocyte fraction analyses of *JAK2* and *DNMT3A* somatic mutations. A. Co-mutation table shows *JAK2*^{V617F} variant allele frequencies by targeted panel testing and ultradeep sequencing relative to the 'GGCC' haplotype. The left annotation reflects the threshold of sensitivity of detection. **B.** *JAK2*^{V617F} and *DNMT3A* somatic mutation burden quantified by droplet digital PCR (ddPCR) in bulk, myeloid, and T cell DNA fractions among three individuals. *JAK2*^{V617F} mutations are exclusive to myeloid cells, but *DNMT3A* mutations show variable myeloid predominance and myeloid-lymphoid enrichment. For each datapoint, ddPCR replicates are shown with error bars (where visible) representing the minimal and maximum 95% confidence interval derived from the Poisson distribution. **C.** Increasing *JAK2*^{V617F} mutation burden (measured by ddPCR) over a 2-year period in bulk DNA for two individuals studied in **panel B**. For Individual 3, the increase in *JAK2*^{V617F} burden is associated with elevated hemoglobin concentration and red blood cell (RBC) counts (**panel D**). **D.** Red blood cell indices (in Individual 3 in the last panel, a 69-year old person with a *POT1* mutation) are plotted over a 5-year period. The upper limit of the normal range is shown in the dashed lines, respectively.

Figure S5



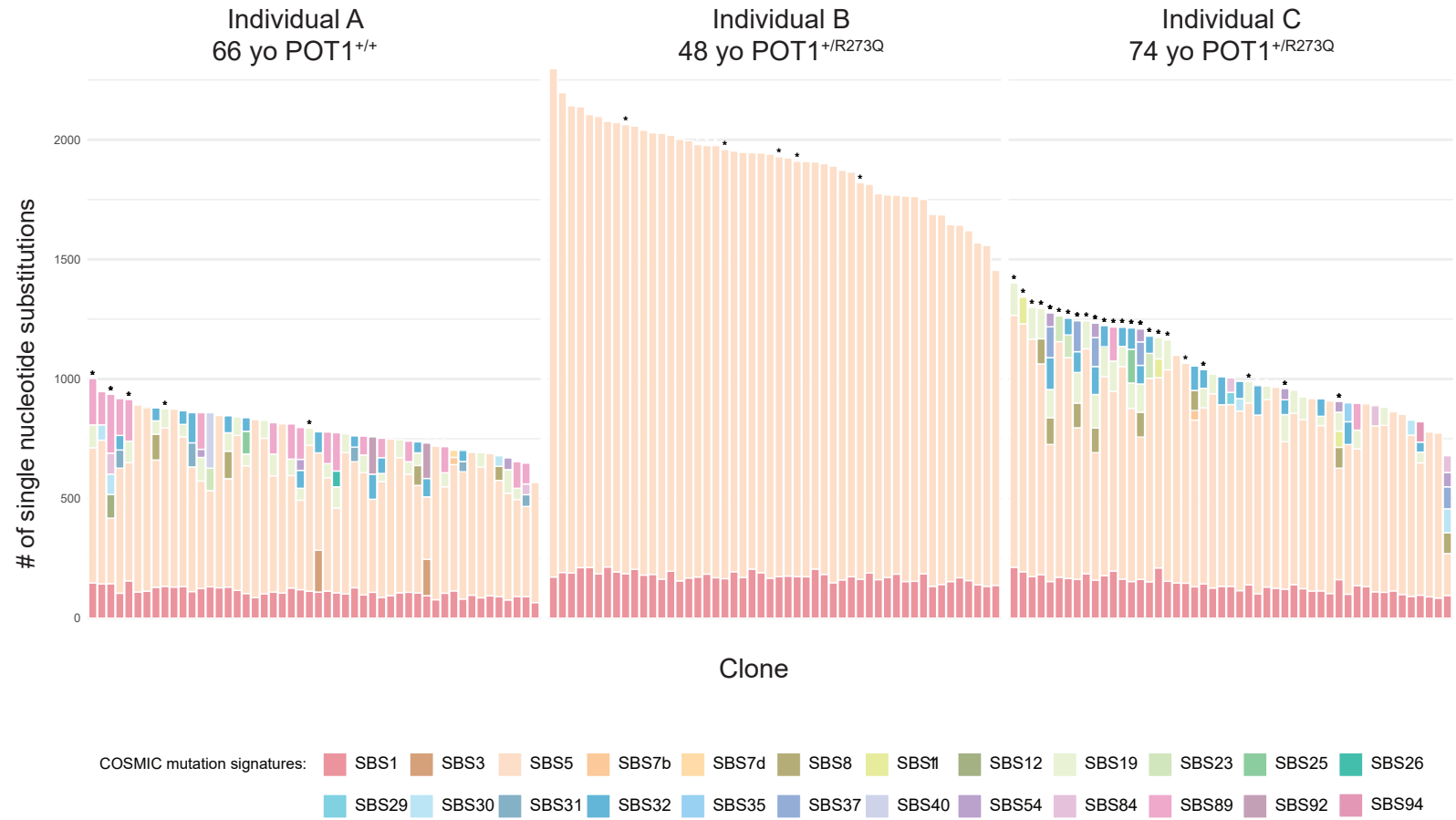
Estimated age of onset of driver clonal hematopoiesis mutation in three individuals. Panels A-C show time-calibrated phylogenetic trees with *POT1* mutation status and telomere length noted above and age on the right. The estimated onset of each somatic driver mutation is annotated above each tree, respectively. Somatic mutations in the two *POT1* mutation carriers in **panels B and C** arose earlier than in the related *POT1* wildtype individual in **panel A**. The tree topology in the two *POT1* mutation carriers (**panels B and C**) also shows oligoclonality with fewer branches arising in the first decade of life, in contrast to the tree in the related individual in **panel A**. Two *JAK2*^{V617F} mutations arose independently in **panel B**, and the two *DNMT3A* mutations in **panel C** were found to be biallelic in phasing studies. In **panel A**, * refers to a *NOTCH2* genotype that was inferred by PCR amplification and Sanger sequencing due to low sequencing coverage at that single nucleotide variant in that colony. **Panel D** is a plot of the mean and standard error of the number of somatic mutations for the sequenced single cell-derived colonies. The data show a higher mean and wider distribution in the two *POT1* mutation carriers compared to the related *POT1* wildtype individual.

Figure S6



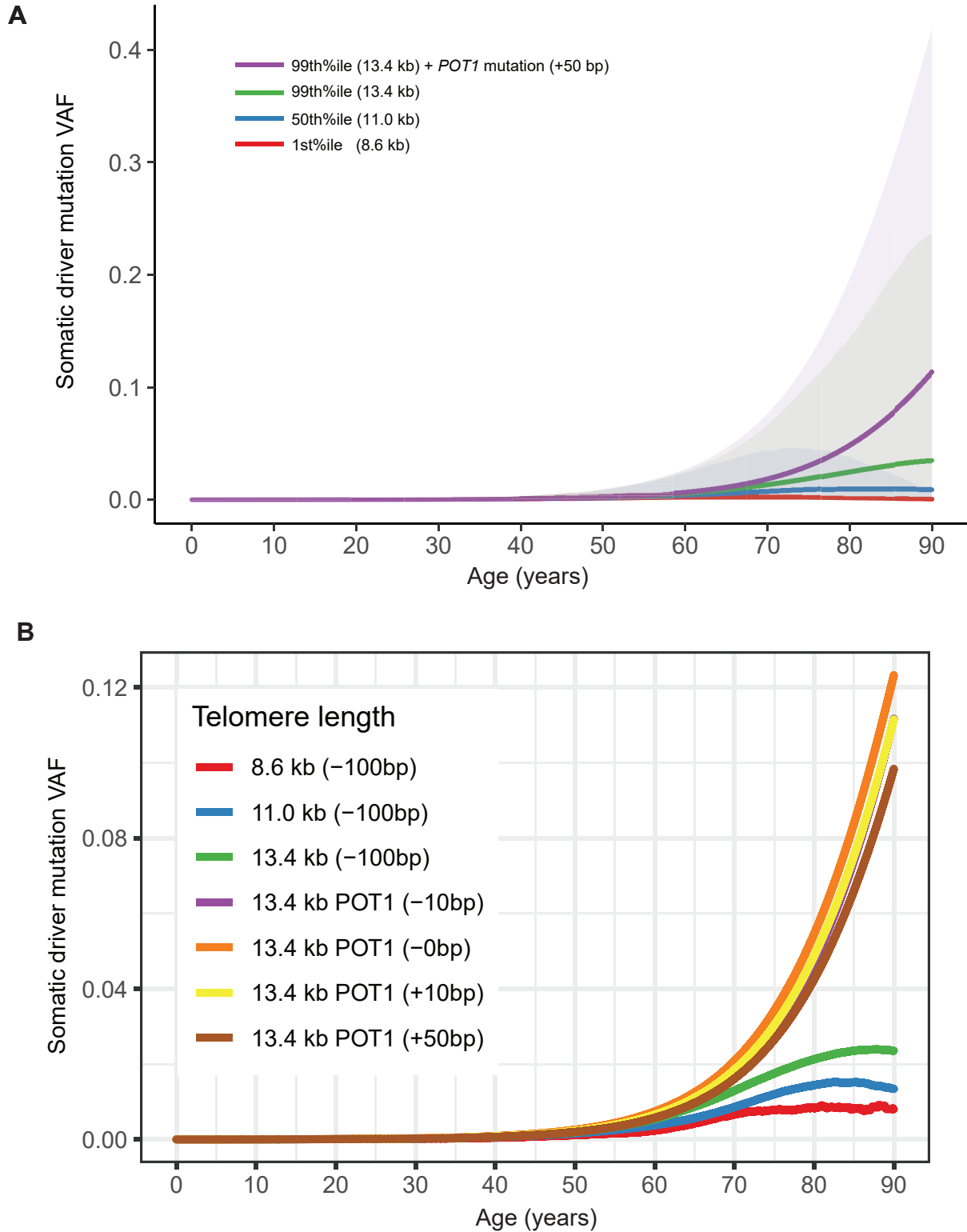
Time-scaled phylogenies of hematopoietic clones annotated for coding mutations in 87 CH-driver genes. Panels A-C show time-calibrated phylogenetic trees with *POT1* mutation status and telomere length noted above and age on the right. Below each tree are the protein-coding mutations, with genes labeled to the left, protein coding effect on the right, and colony identifiers below each matrix. **Panel A** reflects an individual without a *POT1* mutation who carries a non-canonical *NOTCH2* clonal mutation (p.V1003*) in the ancestral lineage to five contemporary samples (variant in A39 was confirmed with PCR and Sanger sequencing), as well as two additional and independent mutations outside the clade. **Panels B** and **C** show trees for two *POT1* mutation carriers. **Panel B** shows convergent *JAK2* p.V617F single nucleotide variants (SNVs), one in an ancestral lineage to four contemporary samples, and the second mutation occurring in an independently evolving lineage; three other protein-coding mutation events are shown, one overlapping with the shared *JAK2* mutation. **Panel C** shows four *DNMT3A* mutations, two SNVs co-occurring in the ancestral lineage to 23 contemporary samples, and two additional mutations identified outside the clade. An EP300 p.P708* frame-shift insertion and *JAK2* p.V617F also emerged within the large *DNMT3A* mutant clade. The list of clonal hematopoiesis genes annotated is in Table S3.

Figure S7



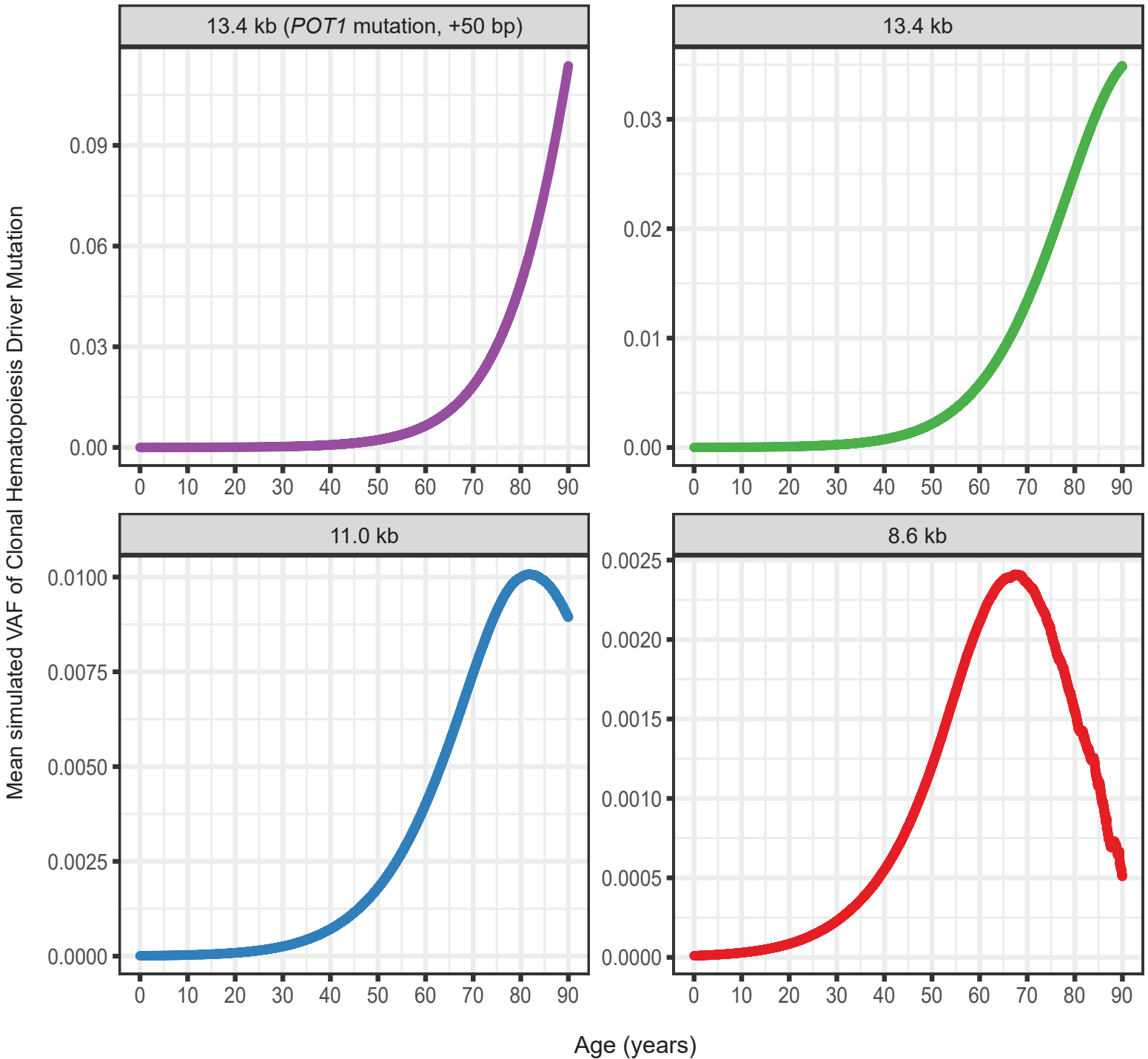
Somatic single base substitution (SBS) COSMIC signature per clone in two *POT1* mutation carriers and a related control who does not carry the mutation. Data are derived from single cell colony whole genome sequence data. The number of mutations per clone is higher in *POT1* mutation carriers and the predominant COSMIC signature is that associated with clock-like DNA replication errors as is seen with aging (C>T transitions, SBS1 and SBS5 as defined by Alexandrov et al. *Nature* 2020). Clones with driver mutations are noted by * above, and some of these clones (with somatic *DNMT3A* mutations especially) have additional secondary signatures. In this analysis, the number of somatic mutations is not adjusted to reflect phylogenetic inference of mutation onset, thus the somatic burden per clone is inflated relative to that presented in Figure 4 and Fig. S5D.

Figure S8



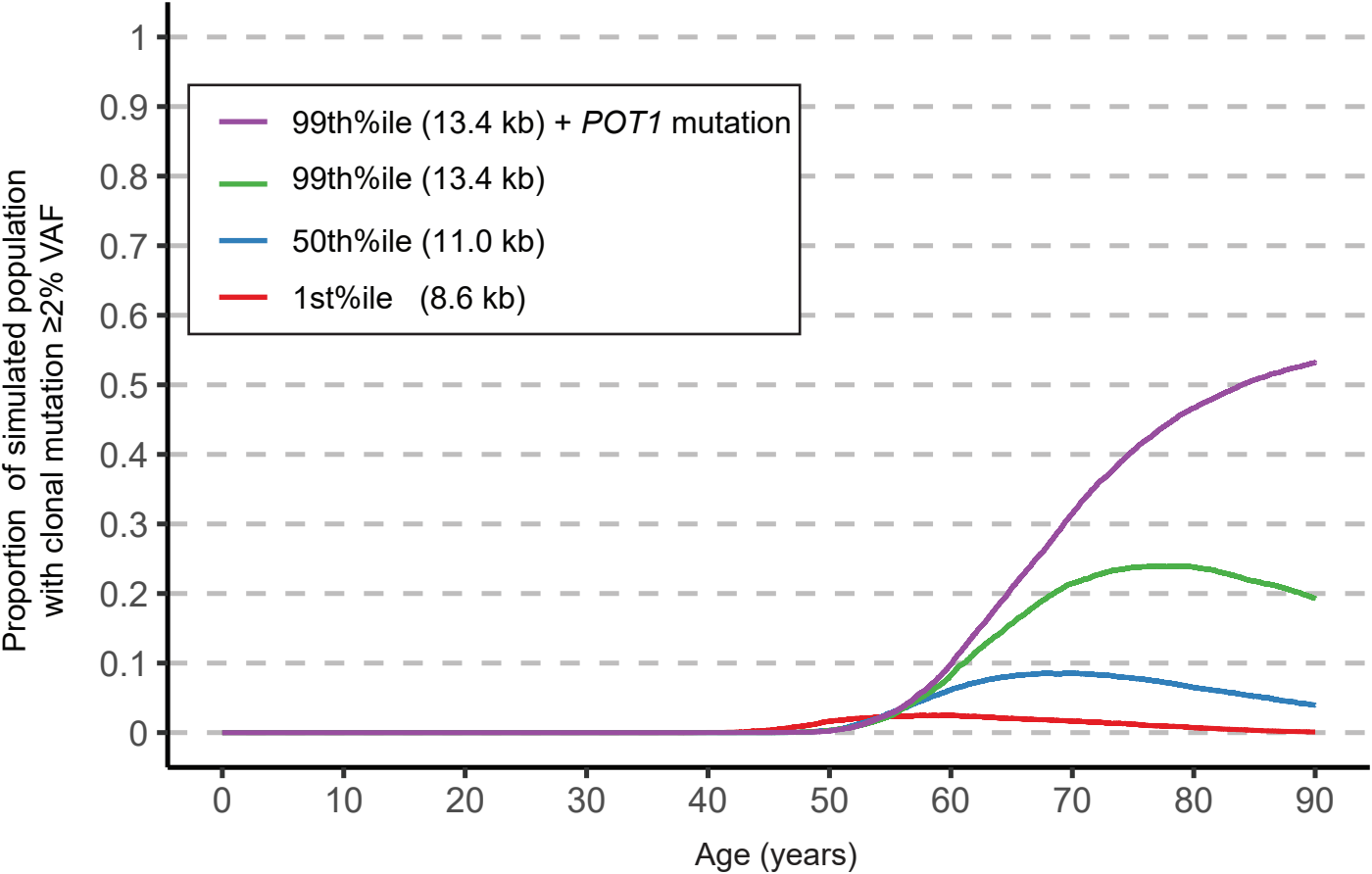
Simulated dynamic of mean variant allele frequent (VAF) of single heterozygous driver mutation acquired at birth and over a lifetime. A. Mean lines with 95% confidence intervals shown in the shaded areas respectively for each group shown in the key. Data reflect 10,000 simulations for each group. Telomere loss in all the *POT1* wild-type groups in panel A was simulated at -100 bp/cell division. **B.** Analysis of varied telomere length gain in the simulation for *POT1* mutation carriers. Key shows initial telomere length at birth with the net telomere length gain/loss per cell division simulated for each group. Data in **panel B** are from 100 simulations for each parameter graphed.

Figure S9



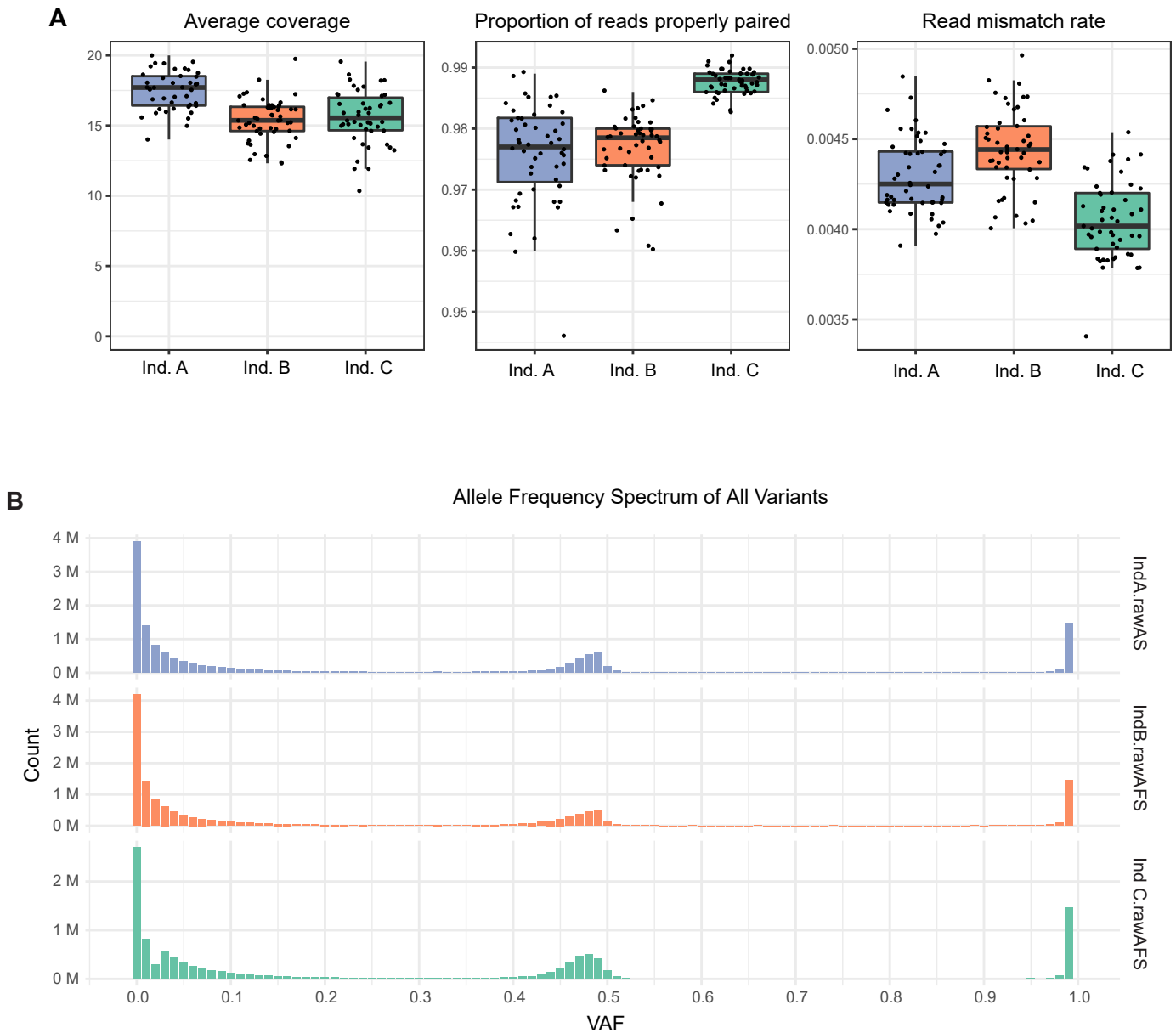
Mean variant allele frequency (VAF) of a clonal hematopoiesis driver mutation acquired at birth, across 10,000 simulations for each of four groups graphed separately. The panel in the top left corner shows the mean VAF for 10,000 simulations for the *POT1* mutation group allowing for a +50 base pair gain with each cell division, while *POT1* wildtype simulations allowed for -100 base pair loss with each cell division. The data shown here are plotted in aggregate in **Figure 4E**.

Figure S10



Proportion of simulated individuals with clonal hematopoiesis of indeterminate potential (CHIP). CHIP was defined using standard $\geq 2\%$ variant allele frequency (VAF) criteria, and proportions were calculated from 10,000 simulations for each of the four groups. The data supplement those shown in **Figure 5E** where the mean clone size across 10,000 simulations is plotted.

Figure S11



Read alignment and unfiltered variant call statistics for whole genome sequence data. A.

Aggregated chromosomal read mapping statistics for colony and bulk sequence data for each individual. Data are shown for 51 whole genomes sequenced (1 bulk and 50 colonies) except for Individual C for whom there was sequence data for 1 bulk and 49 colonies (1 sample failed sequencing). **B.** Allele frequency spectra for all unfiltered variants aggregated across clones per individual.

Figure S12

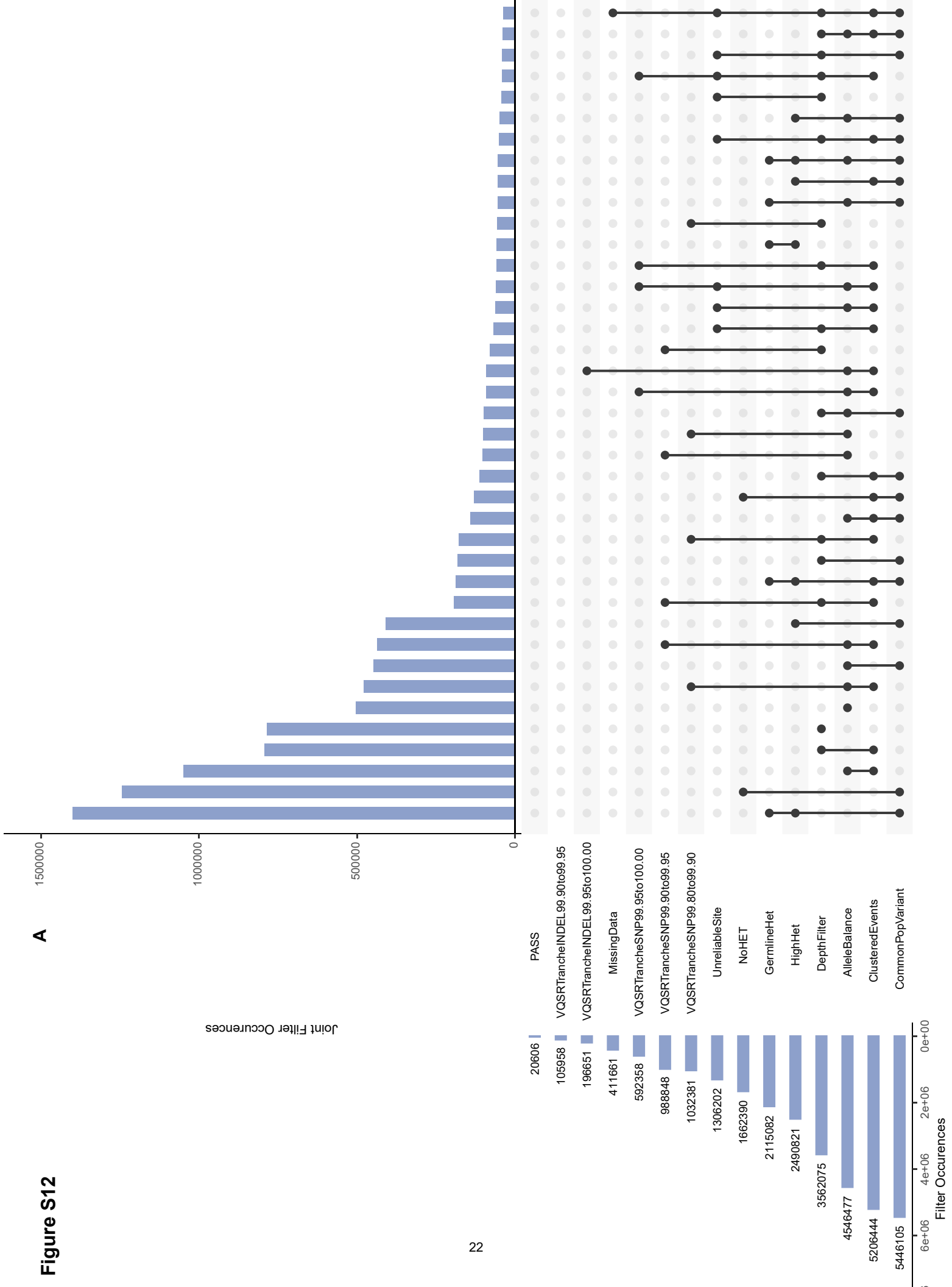


Figure S12 (cont'd)

B

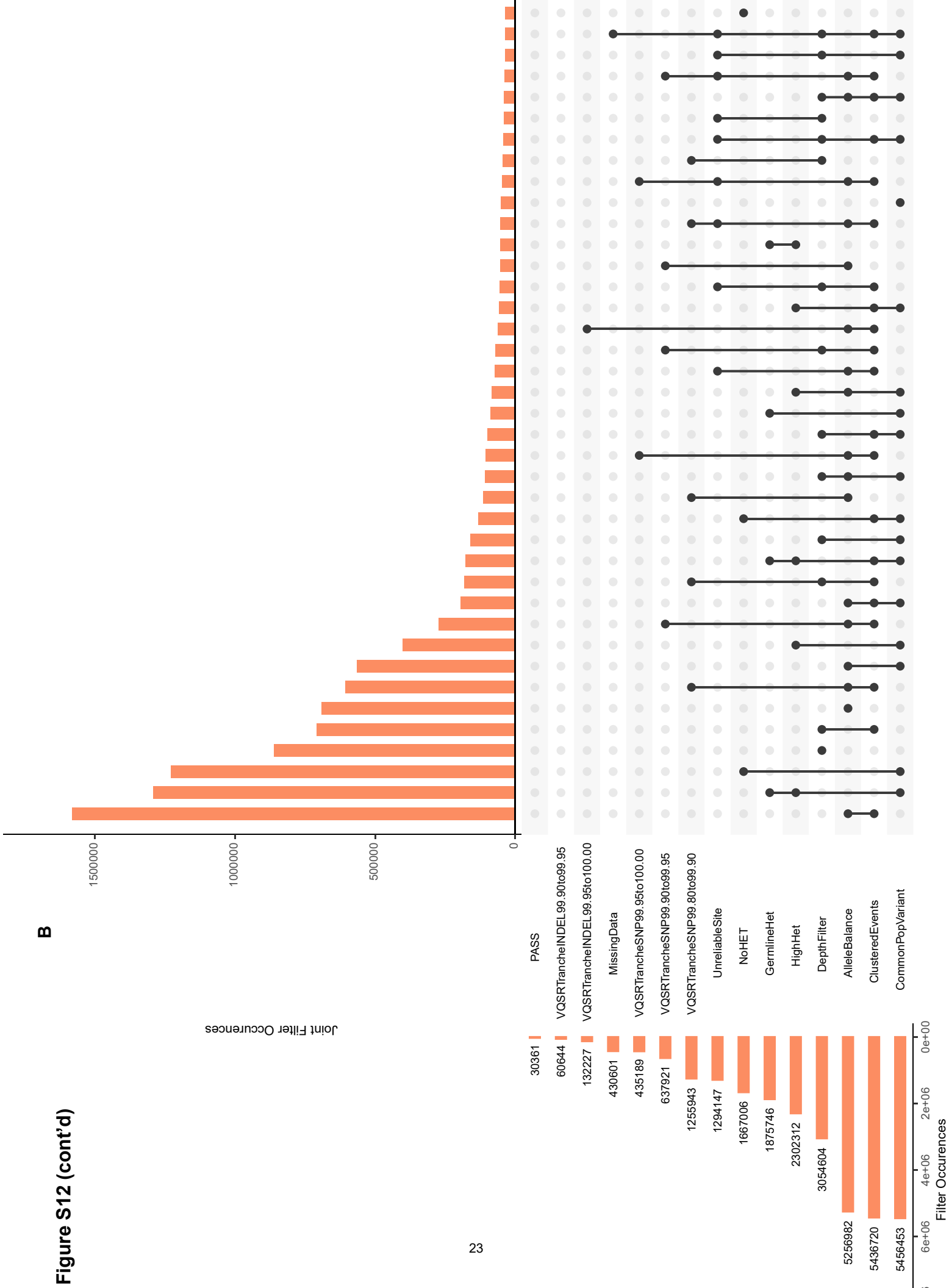
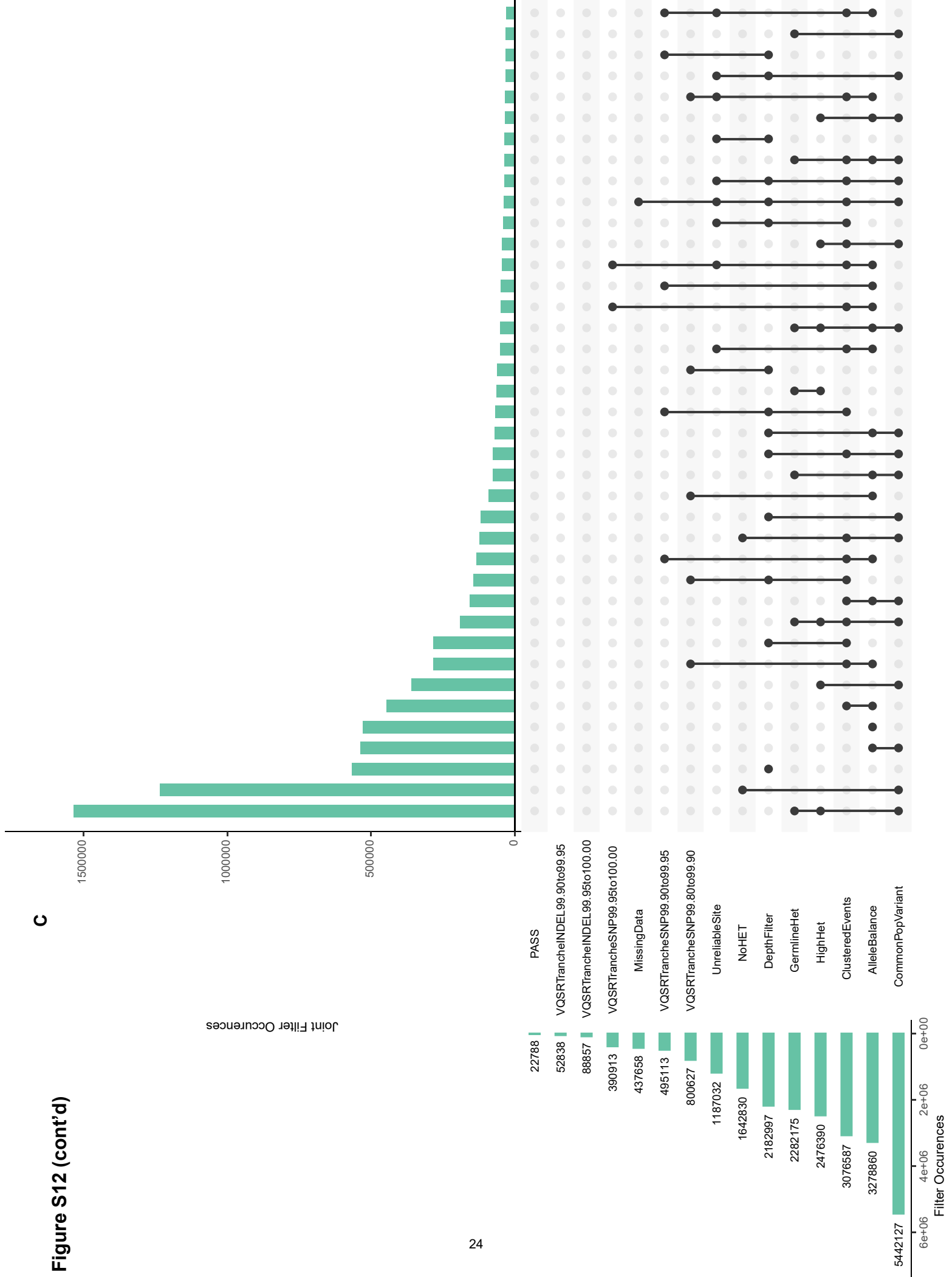
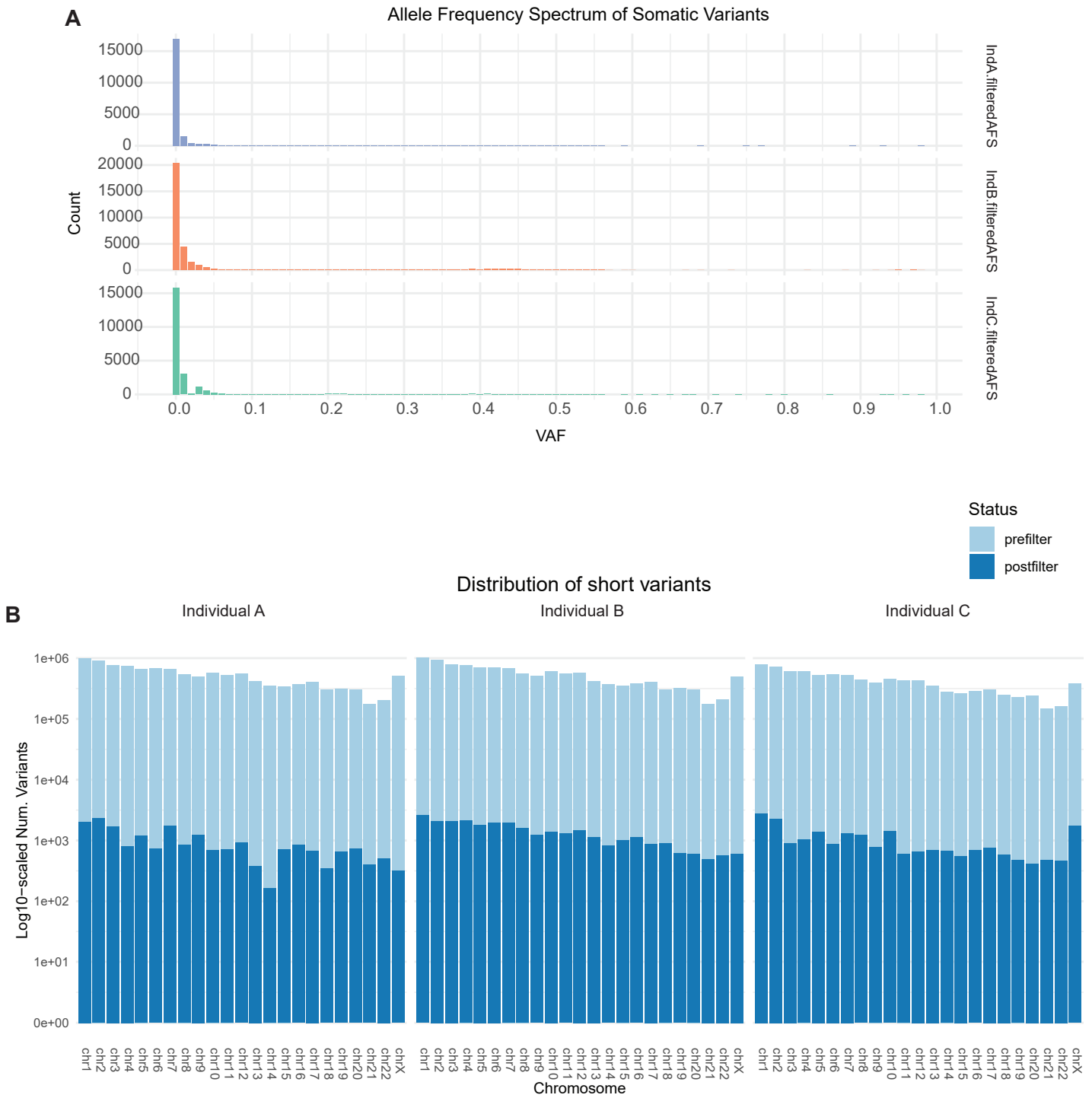


Figure S12 (cont'd)



Summary of variant filtering data for Individuals A, B and C, respectively. Each panel shows each filter's individual application (horizontal bar plot) and joint application (vertical bar plot) to variant calls. Filter combination for each category of the joint filter plot are shown as dot plot matrix belying their respective column and are sorted in descending order of occurrence. Only the top 40 most frequent filter combinations are shown. The "PASS" filter category denotes variants which passed all filters and were subsequently annotated as somatic variants.

Figure S13



Distribution of somatic variants. **A.** Allele frequency spectra for all somatic variants aggregated across clones per individual (ind.). **B.** Count of all variants (light blue) and somatic variants (dark blue) identified or each chromosome

Table S1. *POT1* germline mutations and their prevalence in germline and somatic databases

Coding change (NM_015450)	Protein change (NP_056265)	Genomic coordinate (GRCh37/hg19)	gnomAD v2.1.1* (MAF)	Somatic Prior Report (COSMIC v90)*	CADD Score ^{*,**}	Prior report in germline
c.147delA	p.I49Mfs*7	Chr7:124511073delA	<i>Not reported</i>	<i>Not reported</i>	Nonsense mutation	Reference ³³
c.233T>C	p.I78T	Chr7:124510987A>G	4 of 246,852 (0.000016) 3 of 9,890 (0.000303)	Hairy cell leukemia Serous ovarian cancer	25	Reference ³⁴
c.818G>A	p.R273Q	Chr7:124493077C>T	<i>Not reported</i>	CLL Colon adenocarcinoma Stomach adenocarcinoma Soft tissue sarcoma	30	-
c.1071delT	p.Q358Nfs*9	Chr7:124482953delA	<i>Not reported</i>	<i>Not reported</i>	Nonsense mutation	-
c.1164-1G>A	splice	Chr7:124481233C>T	1 of 31,346 (0.000032)	<i>Not reported</i>	33	Reference ³⁵

Abbreviations: CADD, Combined Annotation-Dependent Depletion; CLL, chronic lymphocytic leukemia, MAF, mean allele frequency

* Databases were accessed last on December 20, 2021

**CADD scores were obtained from the CADD v1.6 tool. Higher CADD scores are associated with a higher probability of deleterious effect. The single nucleotide variant scores are ranked relative to all possible substitutions of the human genome such that scores ≥ 20 indicate the top 1% of most deleterious possible substitutions and ≥ 30 indicate the 0.1% most deleterious.

Notes: All five the variants in this table meet the 2015 American College of Medical Genetics and Genomics (ACMG) criteria of pathogenicity.

Table S2. Onset of hair graying in adults with *POT1* mutations

Age at assessment*	M/F	<i>POT1</i> mutation	Age when hair graying >10%**
(d.83)	F	c.1164-1G>A	70
(d.76)	M	p.I78T	60
72	F	p.R273Q	62
71	F	p.R273Q	Not reached
70	F	c.1164-1G>A	Not reached
70	F	p.R273Q	Not reached
65	M	p.R273Q	Not reached
52	M	p.I78T	Not reached
(d.50)	M	p.R273Q	<i>Not assessable</i>
48	M	p.R273Q	<i>Not assessable/Bald</i>
46	M	p.R273Q	Not reached
42	F	p.Q358Nfs*9	Not reached
41	F	p.R273Q	Not reached
39	F	p.I49Mfs*7	Not reached
34	M	p.R273Q	Not reached

*Age refers to age at last assessed, d. refers to age at the time of death

**Estimates were obtained based on review of photographs and participant's self-report corroborated by two of the authors involved in clinical care.

Table S3A. Somatic variants identified in 12 *POT1* mutation carriers (n=17 variants)

Gene	Protein change	Coding change	Chromosome Position*	Reference Allele	Alternate Allele	VAF (%)	COSMIC v90
<i>BCOR</i>	p.P1344Hfs*2	c.4031delC	chrX:39922038	TG	T	2.7	Not reported
<i>CREBBP</i>	p.R672C	c.2014C>T	chr16:3828111	G	A	2.5	Stomach Endometrium Large intestine
<i>DNMT3A</i>	p.K693Afs*19	c.2076_2077del	chr2:25464435	TTC	T	19.5	Not reported
	p.A644S	c.1930G>T	chr2:25466773	C	A	15.2	Not reported
	p.L295P	c.884T>C	chr2:25470590	A	G	10.7	MDS
	p.I780T	c.2339T>C	chr2:25462068	A	G	10.4	Lymphoid, MDS/AML
	p.Y735C	c.2204A>G	chr2:25463289	T	C	6.0	MDS/AML Lymphoid Rare solid
	p.A575T	c.1723G>A	chr2:25467152	C	T	4.8	Not reported
	p.I681T	c.2042T>C	chr2:25464471	A	G	1.4	Not reported
	Splicing	c.2478+2T>G	chr2:25459803	A	C	1.4	AML
	p.R882H	c.2645G>A	chr2:25457242	C	T	1.3	MDS/AML MPN Lymphoid
<i>EZH2</i>	p.Q526*	c.1576C>T	chr7:148512102	G	A	10.3	Not reported
<i>GNAS</i>	p.R201C	c.601C>T	chr20:57484420	C	T	1.5	Pancreas Pituitary Large intestine Bone Small intestine
<i>JAK2</i>	p.V617F	c.1849G>T	chr9:5073770	G	T	10.2	MPN
						5.9	MDS/AML
						3.1	CML
						2.6	Lymphoid
<i>KRAS</i>	p.G13D	c.38G>A	chr12:25398281	C	T	1.8	Large intestine MDS/AML MPN CML Lymphoid Lung Pancreas Stomach
<i>PIM1</i>	p.K122E	c.364A>G	chr6:37138557	A	G	2.0	Not reported
<i>TET2</i>	p.C1298Lfs*65	c.3893delG	chr4:106180864	TG	T	7.7	AML CML

Table S3B. Somatic variants identified in 21 *POT1* non-carrier relatives (n=3)

Gene	Protein change	Coding change	Chromosome Position*	Reference Allele	Alternate Allele	VAF (%)	COSMIC v90
<i>ABL1</i>	p.E758*	c.2272G>T	chr9:133759949	G	T	1.2	Not reported
<i>DNMT3A</i>	p.E505*	c.1513G>T	chr2:25468163	C	A	5.3	AML
<i>NOTCH2</i>	p.V1003*	c.3007_3008del	chr1:120483352	AAC	A	2.1	Not reported

*Coordinates shown are from UCSC version hg19 (NCBI build GRCh37).

**COSMIC v90 was accessed on December 30, 2021. Previously identified mutations in COSMIC are listed from top to bottom in order of prevalence. Cancer subtypes are listed for hematologic malignancies.

Abbreviations: AML, acute myelogenous leukemia; CML, chronic myelogenous leukemia; MDS, myelodysplastic syndromes; MPN, myeloproliferative neoplasms.

Table S4. Somatic variants in 30 controls over age 70 (n=22 variants)[†]

Gene	Protein change	Coding change	Chromosome Position	Reference Allele	Alternate Allele	VAF (%)	COSMIC v97
ASXL1	p.I1502M	c.4506C>G	chr20:31025021	C	G	6.51	
DNMT3A	p.R366S	c.1096C>A	chr2:25469946	G	T	3.48	AML
	p.T395Qfs*12	c.1183_1187del	chr2:25469580	GGCAGT	G	13.33	
	p.L411del	c.1231_1233del	chr2:25469534	CCAG	C	4.9	
	NA	c.1851+1G>T	chr2:25467023	C	A	3.32	
	p.V657M	c.1969G>A	chr2:25464544	C	T	5.94	Astrocytoma Blastic plasmacytoid dendritic cell neoplasm MPN Stomach
	p.Q692*	c.2074C>T	chr2:25464439	G	A	2.94	
	p.D702Y	c.2104G>T	chr2:25463578	C	A	6.11	
	p.F731Lfs*49	c.2192_2193ins AA	chr2:25463300	G	GTT	20.29	
	p.R736L	c.2207G>T	chr2:25463286	C	A	2.88	MPN Primary CNS lymphoma
	p.L815Q	c.2444T>A	chr2:25459839	A	T	8.62	
p.R882C	c.2644C>T	chr2:25457243	G	A	2.4	MDS/AML MPN Lymphoid	
JAK2	p.V617F	c.1849G>T	chr9:5073770	G	T	2.22	MPN MDS/AML CML Lymphoid
KRAS	p.V14I	c.40G>A	chr12:25398279	C	T	24.76	Large intestine AML ALL Mast cell neoplasm Mantle cell lymphoma Pancreas Skin Lung
NF1	p.I70Nfs*15	c.209delT	chr17:29486031	AT	A	7.09	
NLRP1	p.A436V	c.1307C>T	chr17:5462709	G	A	3.23	Large intestine
TET2	p.Q108Rfs*5	c.323delA	chr4:106155421	CA	C	3.08	
	p.N598Ifs*3	c.1792delA	chr4:106156890	CA	C	2.12	CML
	p.A1159Vfs*67	c.3476delC	chr4:106162561	GC	G	2.74	
TNFAIP3	p.E132Sfs*2 ***	c.393_394insTC CT	chr6:138196079	G	GTCCT	2.17	
	p.E132* ***	c.394G>T	chr6:138196080	G	T	2.23	
TP53	p.L201M	c.601T>A	chr17:7578248	A	T	5.42	

*Coordinates shown are from UCSC version hg19 (NCBI build GRCh37).

**COSMIC v97 was accessed on January 9, 2023. Previously identified mutations in COSMIC are listed from top to bottom in order of prevalence. Cancer subtypes are listed for hematologic malignancies.

***multinucleotide variant

Abbreviations: AML, acute myelogenous leukemia; CNS, central nervous system; MDS, myelodysplastic syndrome; MPN, myeloproliferative neoplasm; ALL, acute lymphoblastic leukemia; CML chronic myeloid leukemia

[†]Note: This group of controls was sequenced and analyzed on the same platform and pipeline as the *POT1* mutation carriers (data in Table S3) but at a two-fold higher depth of coverage (1300x vs. 600x) as detailed in the Supplementary Methods.

Table S5. Primer sequences

Name	Primer Sequence
ARF3 qRT-PCR F	5' TCACCACCATCCCTACCATT 3'
ARF3 qRT-PCR R	5' AGGTGGCCTGAATGTACCAG 3'
DNMT3A c1930G>T pA644S	BioRad dHsaMDS316402850*
DNMT3A c2042T>C pI681T	BioRad dHsaMDS196828418*
DNMT3A c2339T>C pI780T	BioRad dHsaMDS415068371*
DNMT3A c2076 2077del pK693Afs*19	BioRad dHsaMDS992328224*
JAK2 c1849G>T V617F	BioRad dHsaMDV27944642*
JAK2 rs12343867 F	5' GAGATTATGGCAGGTTCAAC 3'
JAK2 rs12343867 R	5' TAGTAGTTTCTGTGAACACC 3'
POT1 qRT-PCR F	5' GATCTCTGAAGGTTGGAAGC 3'
POT1 qRT-PCR R	5' GTTACTTTCTGGCAAGACCC 3'
POT1 R273Q F Genotyping	5' GGTTGGAAGCTTTCTTAGAATC 3'
POT1 R273Q R Genotyping	5' GAAATCGGCTTAATCGATACC 3'
POT1 ex11-12 F Splicing	5' CTTTCCAAGCTCTGGATCAGTAT 3'
POT1 ex17-18 R Splicing	5' GATGCTGGAATCTGGAAGAATTTG 3'
POT1 Q358Nfs F Genotyping	5' GGTTTCAGGAGGATGCATGTC 3'
POT1 Q358Nfs R Genotyping	5' ACTGAAATAGTCTTCTGGGC 3'
POT1 c1164-1G>A F Genotyping	5' TCACGCTTACACCAAATCG 3'
POT1 c1164-1G>A R Genotyping	5' ATATCCAAATCGCCCTCATG 3'
DNMT3Aphasing F	5' TCCAAAGGTTTACCCACCTG 3'
DNMT3Aphasing R	5' GTTCATACCGGGAAGGTTAC 3'

*These identifiers refer to probe sets that were designed using the BioRad Assay Design Engine and are searchable at Bio-Rad.com/digital-assays.

Table S6. List of genes analyzed for somatic mutation studies and their accession identifiers (n=87)

ABL1	NM_005157	FBXW7	NM_033632	PAX5	NM_016734
ASXL1	NM_015338	FLT3	NM_004119	PDGFRA	NM_006206
ATM	NM_000051	FOXO1	NM_002015	PHF6	NM_032458
ATRX	NM_000489	GATA1	NM_002049	PIGA	NM_002641
BCL2	NM_000633	GATA2	NM_032638	PIM1	NM_001243186
BCL6	NM_001706	GNA13	NM_006572	PLCG2	NM_002661
BCOR	NM_017745	GNAS	NM_000516	PPM1D	NM_003620
BCORL1	NM_021946	IDH1	NM_005896	PRDM1	NM_001198
BRAF	NM_004333	IDH2	NM_002168	PTEN	NM_000314
BTK	NM_000061	IKZF1	NM_006060	PTPN11	NM_002834
CALR	NM_004343	JAK2	NM_004972	RAD50	NM_005732
CARD11	NM_032415	KDM2B	NM_032590	RECQL4	NM_004260
CBL	NM_005188	KDM6A	NM_021140	RHOH	NM_004310
CBLB	NM_170662	KIT	NM_000222	RUNX1	NM_001754
CD79A	NM_001783	KMT2A	NM_001197104	SETBP1	NM_015559
CD79B	NM_000626	KMT2D/MLL2	NM_003482	SF3B1	NM_012433
CDKN2A	NM_000077	KRAS	NM_004985	SGK1	NM_005627
CEBPA	NM_004364	MEF2B	NM_001145785	SRSF2	NM_003016
CHEK2	NM_007194	MPL	NM_005373	STAG2	NM_006603
CREBBP	NM_004380	MYC	NM_002467	STAT3	NM_139276
CSF3R	NM_156039	MYD88	NM_002468	STAT5B	NM_012448
CXCR4	NM_001008540	NF1	NM_000267	STAT6	NM_003153
DDX41	NM_016222	NLRP1	NM_033004	TET2	NM_001127208
DNMT3A	NM_022552	NOTCH1	NM_017617	TNFAIP3	NM_001270507
EP300	NM_001429	NOTCH2	NM_024408	TNFRSF14	NM_003820
PERBB2	NM_004448	NPM1	NM_002520	TP53	NM_000546
ETV6	NM_001987	NRAS	NM_002524	U2AF1	NM_006758
EZH2	NM_004456	NSD1	NM_022455	WT1	NM_024426
FAS	NM_000043	NUP98	NM_016320	ZRSR2	NM_005089

This panel of genes was captured and sequenced on a clinical pipeline at Johns Hopkins Hospital Pathology Laboratories.

References for Supplementary Appendix

1. Olcaydu D, Harutyunyan A, Jäger R, et al. A common JAK2 haplotype confers susceptibility to myeloproliferative neoplasms. *Nat Genet* 2009;41:450-4.
2. Jones AV, Chase A, Silver RT, et al. JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nat Genet* 2009;41:446-9.
3. Penno MB, Pedrotti-Krueger M, Ray T. Cryopreservation of whole blood and isolated lymphocytes for B-cell immortalization. *J Tiss Cult Meth* 1993;15:43-8.
4. Sun X, Becker-Catania SG, Chun HH, et al. Early diagnosis of ataxia-telangiectasia using radiosensitivity testing. *J Pediatr* 2002;140:724-31.
5. Stanley SE, Rao AD, Gable DL, McGrath-Morrow S, Armanios M. Radiation Sensitivity and Radiation Necrosis in the Short Telomere Syndromes. *Int J Radiat Oncol Biol Phys* 2015;93:1115-7.
6. Schratz KE, Gaysinskaya V, Cosner ZL, et al. Somatic reversion impacts myelodysplastic syndromes and acute myeloid leukemia evolution in the short telomere disorders. *J Clin Invest* 2021;131.
7. Gable DL, Gaysinskaya V, Atik CC, et al. ZCCHC8, the nuclear exosome targeting component, is mutated in familial pulmonary fibrosis and is required for telomerase RNA maturation. *Genes Dev* 2019;33:1381-96.
8. Stanley SE, Gable DL, Wagner CL, et al. Loss-of-function mutations in the RNA biogenesis factor NAF1 predispose to pulmonary fibrosis-emphysema. *Science translational medicine* 2016;8:351ra107.
9. Gaysinskaya V, Stanley SE, Adam S, Armanios M. Synonymous Mutation in DKC1 Causes Telomerase RNA Insufficiency Manifesting as Familial Pulmonary Fibrosis. *Chest* 2020;Dec:2449-57.
10. Carlson CS, Emerson RO, Sherwood AM, et al. Using synthetic templates to design an unbiased multiplex PCR assay. *Nature communications* 2013;4:2680-.
11. Emerson RO, DeWitt WS, Vignali M, et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat Genet* 2017;49:659-65.
12. Alder JK, Hanumanthu VS, Strong MA, et al. Diagnostic utility of telomere length testing in a hospital-based setting. *Proc Natl Acad Sci U S A* 2018;115:E2358-e65.
13. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164.
14. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv: Genomics* 2013.
15. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. *GigaScience* 2021;10.
16. Van der Auwera GA, O'Connor BD. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*: O'Reilly Media; 2020.
17. Mitchell E, Spencer Chapman M, Williams N, et al. Clonal dynamics of haematopoiesis across the human lifespan. *Nature* 2022;606:343-50.
18. Williams N, Lee J, Mitchell E, et al. Life histories of myeloproliferative neoplasms inferred from phylogenies. *Nature* 2022;602:162-8.
19. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;581:434-43.

20. Zheng-Bradley X, Streeter I, Fairley S, Richardson D, Clarke L, Flicek P. Alignment of 1000 Genomes Project reads to reference assembly GRCh38. *Gigascience* 2017;6:1-8.
21. Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* 2020;37:1530-4.
22. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 2017;14:587-9.
23. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* 2018;35:518-22.
24. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 2019;35:526-8.
25. Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. *Nature* 2020;578:94-101.
26. Haller BC, Messer PW. SLiM 3: Forward Genetic Simulations Beyond the Wright-Fisher Model. *Mol Biol Evol* 2019;36:632-7.
27. Yang L, Rau R, Goodell MA. DNMT3A in haematological malignancies. *Nat Rev Cancer* 2015;15:152-65.
28. Nangalia J, Griffin J, Green AR. Pathogenesis of Myeloproliferative Disorders. *Annu Rev Pathol* 2016;11:101-26.
29. Lahouel K, Younes L, Danilova L, et al. Revisiting the tumorigenesis timeline with a data-driven generative model. *Proc Natl Acad Sci U S A* 2020;117:857-64.
30. Alder JK, Chen JJ, Lancaster L, et al. Short telomeres are a risk factor for idiopathic pulmonary fibrosis. *Proc Natl Acad Sci U S A* 2008;105:13051-6.
31. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna, Austria. 2022.
32. Robles-Espinoza CD, Harland M, Ramsay AJ, et al. POT1 loss-of-function variants predispose to familial melanoma. *Nat Genet* 2014;46:478-81.
33. Jones S, Anagnostou V, Lytle K, et al. Personalized genomic analyses for cancer mutation discovery and interpretation. *Science translational medicine* 2015;7:283ra53.
34. Wong K, Robles-Espinoza CD, Rodriguez D, et al. Association of the POT1 Germline Missense Variant p.I78T With Familial Melanoma. *JAMA Dermatol* 2018.
35. Speedy HE, Kinnersley B, Chubb D, et al. Germ line mutations in shelterin complex genes are associated with familial chronic lymphocytic leukemia. *Blood* 2016;128:2319-26.