Supplementary Methods, Tables, Figures

Gliogene recruitment descriptive statistics

The Gliogene Consortium recruited 435 glioma families from 14 centers in the US, Sweden, Denmark, UK and Israel between 2007-2011. The mean and age range of the participants in the cohort was 49.1 and range 1-91 years, respectively (Supplementary Table 1). We identified families through a screening questionnaire at each of the centers in addition to recruitment through the internet. Eligible families were consented, interviewed, and blood or saliva samples were obtained from cases and unaffected family members. We verified all glioma cases with medical record or pathology report.

Supplementary Table 1: Description of Gliogene Families with ≥ 2 Gliomas Verified by Country

	Total	**US**	**Sweden**	**Denmark**	**Israel**
# families	435	309	44	52	30
# of gliomas per family (n, %)					
2	347 (79.8)	242 (78.3)	40 (90.1)	39 (75.0)	26 (86.7)
3	62 (14.3)	47 (15.2)	3 (6.8)	8 (15.4)	4 (13.3)
≥ 4	26 (6.0)	20 (6.5)	1 (2.2)	5 (9.6)	0 (0)
# of generations (n, %)					
1	117 (27.0)	79 (25.6)	7 (15.8)	15 (29.6)	16 (53.3)
2	283 (64.8)	203 (65.7)	37(84.2)	29 (55.6)	14(46.7)
3	33 (7.7)	25 (8.1)	0 (0)	8 (14.8)	0 (0)
4	2 (.4)	2 (.6)	0 (0)	0 (0)	0 (0)
Age at diagnosis					
Mean ± SD	49.1±19.1	49.4±19.3	52.4±12.8	50.0±16.0	42.0±21.8
Range	1 -91	1-91	22-79	17-82	1-77

Gliogene Exome Sequencing Selection Methods

The selection criteria for inclusion of exome sequencing was: 1) three or more gliomas in the family with DNA from at least one affected, or 2) two gliomas in the family with DNA available from both affected. For this paper, we included 90 affected samples from 55 US families of which 22 (40%) had two gliomas, 23 (42%) had three gliomas, and 10 (18%) had 4 or more gliomas reported in the

family. Average coverage of exome sequenced samples was 95x with an average of 86% of the targeted region at 20x coverage or higher.

The glioma subtypes among the 90 included samples were the following: 32 glioblastomas (36%), 17 anaplastic astrocytomas (19%), 12 diffuse astrocytomas(13%), 4 juvenile pilocytic astrocytomas (4%), 4 unclassified gliomas (4%), 1 ganglioglioma (1%), 10 oligodendrogliomas (11%), 5 anaplastic oligodendrogliomas (6%), 3 oligoastrocytomas (3%), and 2 anaplastic oligoastroycytomas (2%). Out of the 20 cases with oligodendroglial features (oligodendrogliomas, anaplastic oligodendrogliomas, oligoastrocytomas, and anaplastic oligoastroycytomas), 3 (15%) had a *POT1* mutation.

Supplementary Table 2: Mutations in Glioma Families

Family	Gene	Mutation	Affected Individuals Available for Sequencing	Sequencing strategy
(A)	*POT1*	NM_015450:p.G95C, HG19:chr7:g.124503667C>A	Proband, Uncle	Exome
(B)	*POT1*	NM_015450:p.E450X, HG19:chr7:g.124481048C>A	Proband, Uncle	Exome
(C)	*POT1*	p.D617Efs*8, HG19:chr7:g.124464068TTA>T	Proband	Amplicon

Telomere Content Analysis

Methods for qPCR and Statistical Analysis

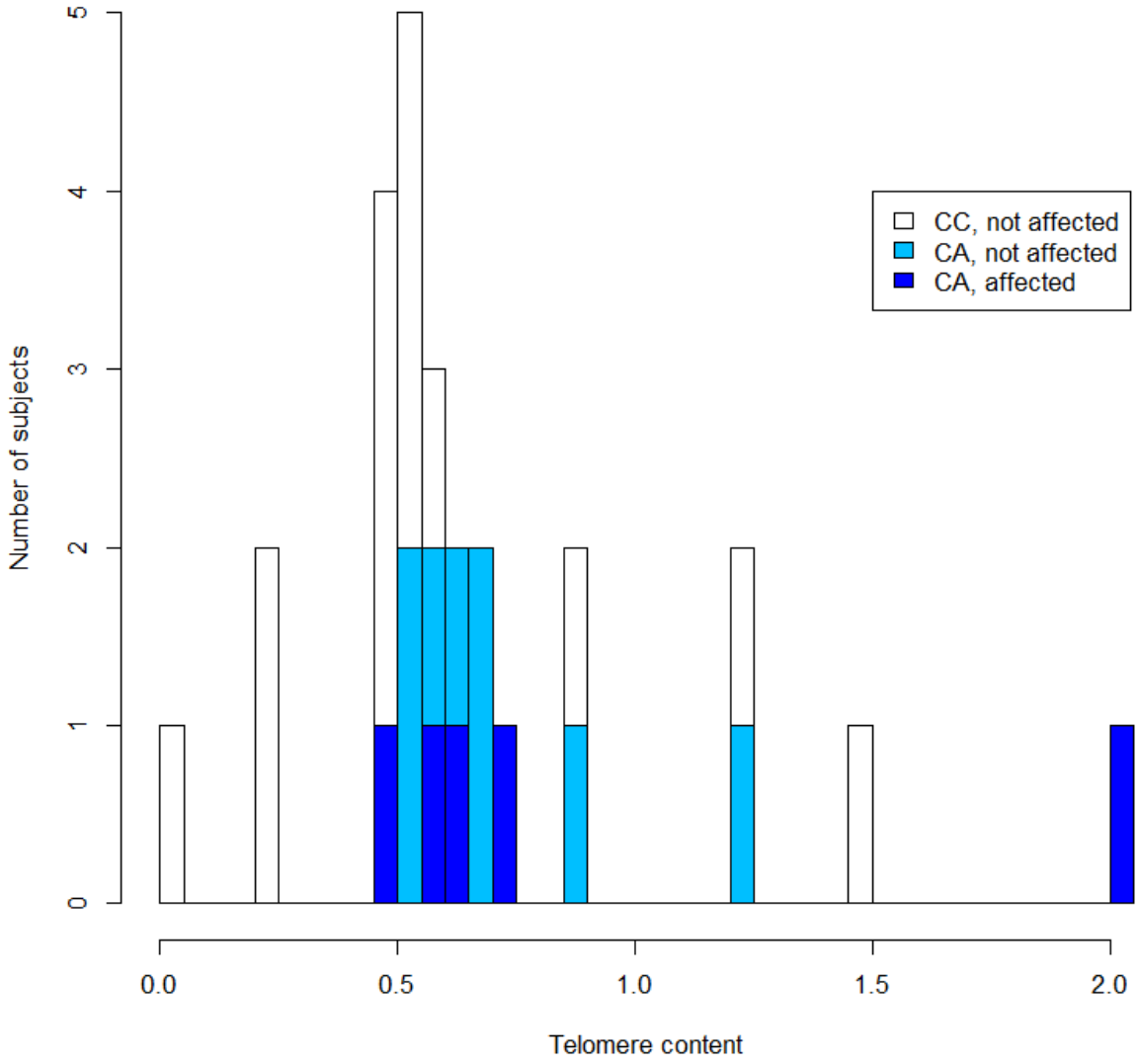
We estimated telomere content (TC) using qPCR as previously described[1]. We randomly intermixed the samples so that the investigators were blinded to the case/control status at the time of TC measurement and calculation. The DNA for the TC was derived from either blood or saliva, both of which contain predominantly lymphocyte DNA[1]. Cases (those carrying the mutation) and controls (those not carrying the mutation) were matched by gender and race. The mean TC's in cases and controls were first compared using a two-tailed t-test. We then generated a receiver operating characteristic (ROC)

curve by plotting sensitivity vs. specificity for TC in predicting whether a subject is/is not affected with the disease and does/does not carry the mutation, using bootstrapping to compute confidence intervals for the area under the curve (AUC).

Results: qPCR for Telomere Content

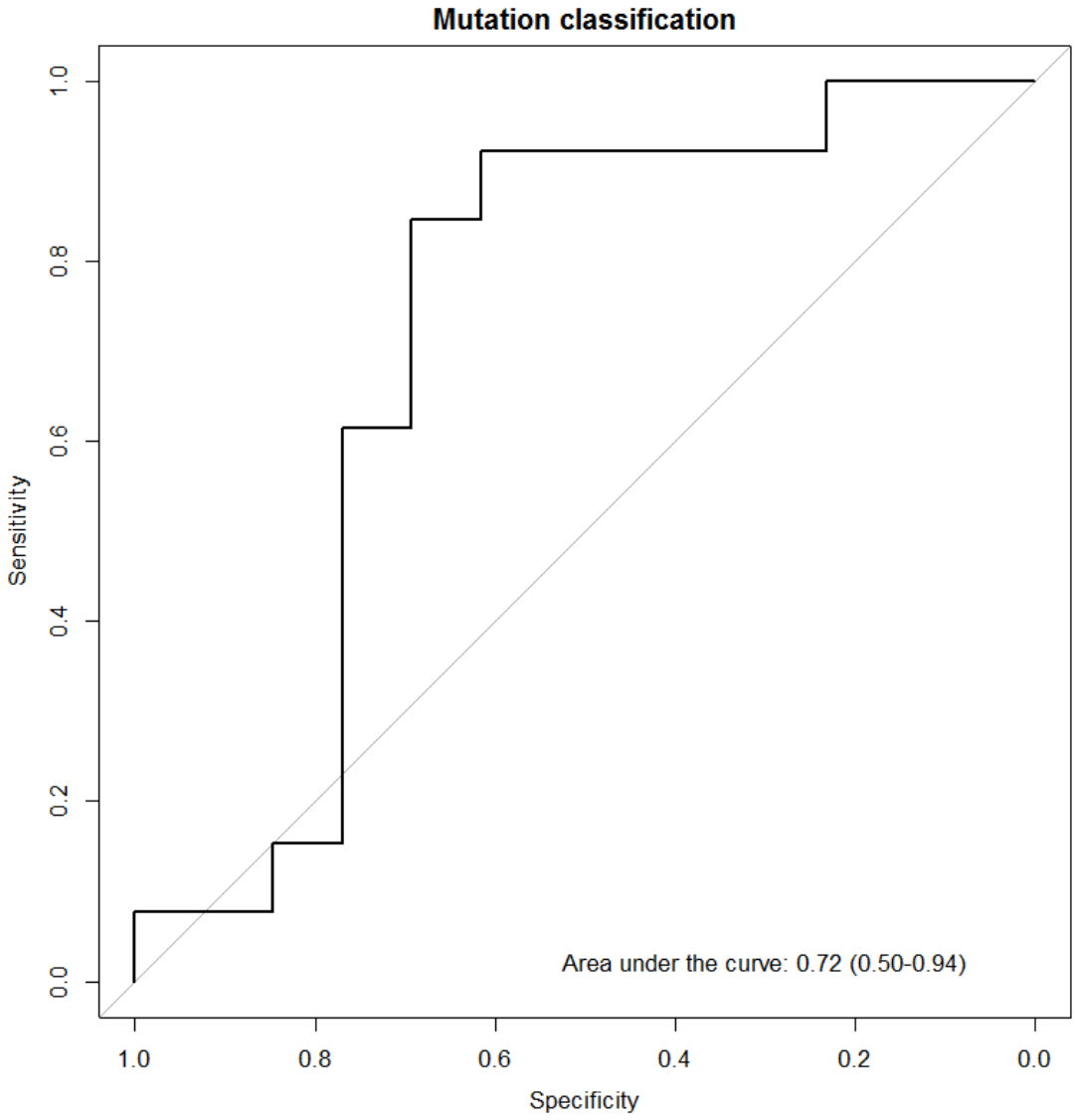
Given evidence for an association between *POT1* mutations and alterations in telomere length, we used qPCR to measure telomere DNA content (TC), a proxy for telomere length, in our family cohort samples from Families A, B, and C as well as in healthy controls. A total of 26 samples were analyzed by qPCR, including 13 samples from subjects with the wild-type CC (3 of which were familial controls), and 13 samples from subjects with the CA mutation, 5 of which were affected by the disease. The mean TC for both cases and controls was 0.68, SD 0.4. T-test comparison between those classified as CC or CA was not significant (CC mean TC 0.59, SD 0.4; CA mean TC 0.78, SD 0.4; p=0.25). T-test comparison between those classified as unaffected or affected was also not significant (CC or CA, unaffected mean TC 0.64, SD 0.34; CA, affected mean TC 0.89, SD 0.6; p=0.25). The distribution of TC in this cohort is shown in Supplementary Figure 1. Though the resulting t-test comparisons were limited by small sample size, the trend observed was as expected, suggesting longer telomeres in those with the CA genotype. We then conducted a multivariable analysis with TC as a continuous explanatory variable. Multivariate odds ratios (OR) for either glioma cases or mutation carriers were calculated by logistic regression with corresponding 95% confidence limits and Wald $\chi 2$ tests for significance. However, for both glioma cases and mutation carriers, the results were not significant, (glioma cases: OR 3.57, CI 0.4-32.0, p=0.25, mutation carriers: adjusted OR 3.67, CI 0.4-35.3, p=0.26). An ROC curve was then generated testing the ability of TC to predict mutational and affected status within the study cohort. The results for unaffected vs. affected were not significant. However, the results were statistically significant for mutational status (CC vs. CA), with an AUC of 0.72, CI 0.5-0.94 (Supplementary Figure 2).

Supplementary Figure 1: The Distribution of Telomere Content in Relation to Number of Subjects Tested, Including Affected Carriers, Unaffected Carrier, and Unaffected Wild-Type Controls from Families A, B, and C



Telomere content estimations are shown relative to population density, with non-carriers who are not affected shown in white (n=13), carriers who are not affected shown in light blue (n=8), and carriers who are affected shown in dark blue (n=5).

Supplementary Figure 2: ROC for Mutational Status (CC vs. CA)

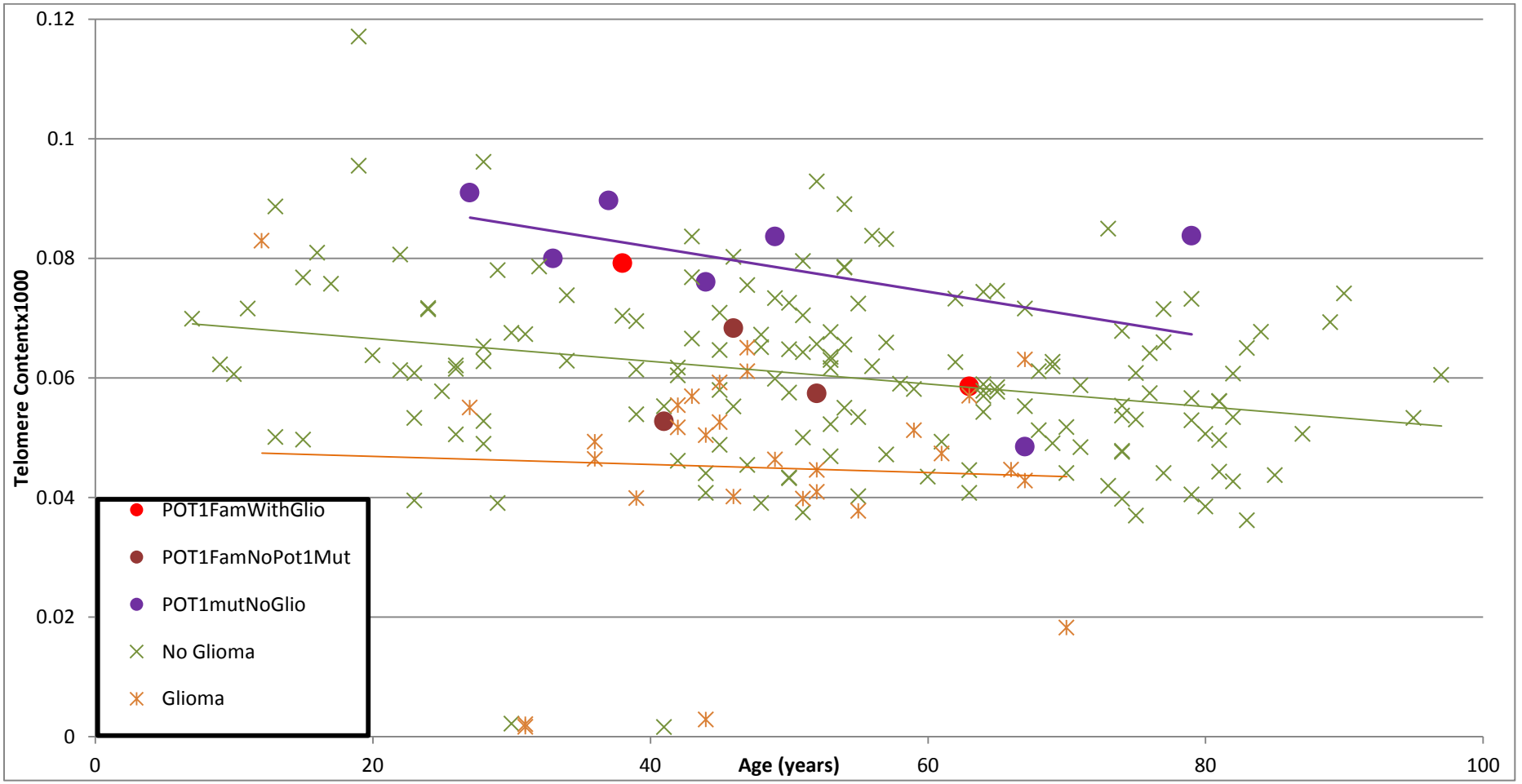


The true positive rate, or sensitivity, for TC as predictive of mutational status is plotted as a function of the false positive rate (100-specificity), with the area under the curve (AUC) representing the ability of TC to distinguish between the CC and CA genotypes. The AUC is just significant, with 95% confidence intervals ranging from 0.501 to 0.943.

Capture-Sequencing Telomere Content (TC) Methods

Data were generated using a regional capture reagent, which targets the coding exons, transcription factor binding sites and conserved 3'UTR elements from the p-arm of chromosome 17 from 31 familial glioma cases and 173 unaffected family members. Additional DNA was not available from 3 members of family A.

We calculated TC using a previously reported method[6]. Briefly, we counted telomere aligning ("off-target") reads, and reads that contained at least 4-telomere repeats and normalized this to the total number of reads produced for each capture experiment. We conducted covariant analysis using the R-statistical package.



Supplementary Figure 3: Telomere content vs. age to show relationship between *POT1* mutation status (purple and red circles), glioma status (red circles and orange asterisks) and TC. Also shown are *POT1* family members without *POT1* mutation (dark red circles). Best fit linear regression lines are shown for POT1 mutated samples (both with and without glioma) (purple line), glioma affected individuals (orange line) and those without glioma (green line).

Sequencing and Analysis

Whole Exome and Regional Capture Methods

Library Preparation:

DNA samples were constructed into Illumina paired-end pre-capture libraries according to the manufacturer's protocol (*Illumina Multiplexing_SamplePrep_Guide_1005361_D*) with modifications as described in the BCM-HGSC protocol (https://hgsc.bcm.edu/sites/default/files/documents/Illumina_Barcoded_Paired-End_Capture_Library_Preparation.pdf). Libraries were prepared using Beckman robotic workstations (Biomek NXp and FXp models). Briefly, 1 ug of DNA was sheared into fragments of approximately 300-400 base pairs with the Covaris E210 system followed end-repair, A-tailing, and ligation of either 6bp Illumina multiplexing adaptors or 9bp barcoded adaptors. Pre-capture ligation-mediated PCR (LM-PCR) was performed for 6-8 cycles of amplification using the 2X SOLiD Library High Fidelity Amplification Mix (a custom product manufactured by Invitrogen). Purification was performed with Agencourt AMPure XP beads after enzymatic reactions, and following the final purification, quantification and size distribution of the pre-capture LM-PCR product was determined using the LabChip GX electrophoresis system (PerkinElmer).

Capture Enrichment:

In preparation for co-capture, the uniquely barcoded pre-capture libraries were pooled in equimolar amounts (totaling 1ug/pool). For exome capture, each library pool (6 samples/pool) was then hybridized in solution to the HGSC-designed VCRome 2.1 capture reagent (42Mb, NimbleGen) according to the manufacturer's protocol (*NimbleGen SeqCap EZ Exome Library SR User's Guide*) with minor revisions, while for regional capture, 46-plex library pools used the custom Gliogene capture reagent (1.6Mb, NimbleGen). Human COT1 DNA was added into the hybridization to block repetitive genomic sequences, followed by post-capture LM-PCR amplification using the 2X SOLiD Library High

Fidelity Amplification Mix with 14 cycles of amplification. After the final SPRI bead purification, quantity and size of the capture library was analyzed using the Agilent Bioanalyzer 2100 DNA Chip 7500. The efficiency of the capture was evaluated by performing a qPCR-based quality assay on the four standard NimbleGen internal controls. Successful enrichment of the capture libraries was estimated to range from a 6 to 9 of ΔCt value over the non-enriched samples.

Sequencing:

Library templates were prepared for sequencing using Illumina's cBot cluster generation system with TruSeq PE Cluster Generation Kits. Briefly, these libraries were denatured and diluted in hybridization buffer in order to achieve a load density of ~800K clusters/mm^2. Each library pool was loaded in a single lane of a HiSeq flow cell, and each lane was spiked with 2% phiX control library for run quality control. The sample libraries then underwent bridge amplification to form clonal clusters, followed by hybridization with the sequencing primer. Sequencing runs were performed in paired-end mode using the Illumina HiSeq 2000 platform. Using the TruSeq SBS Kits, sequencing-by-synthesis reactions were extended for 101 cycles from each end, with an additional 7 or 10 cycles for the index read. Sequencing runs generated approximately 300-400 million successful reads on each lane of a flow cell, yielding an average of ~6 Gb per exome sample and ~800 Mb per regional sample. With these sequencing yields, exome samples achieved an average of 81% of the targeted bases covered to a depth of 20X or greater and regional samples achieved 86%.

Sequence alignment, variant calling, annotation, and verification

Illumina data were aligned by use of Burrows-Wheeler Aligner (BWA) software. Variants were called using ATLAS-SNP v2.0 and the SAMtools program Pileup. De novo variants were found by in silico subtraction of the variants discovered in either parent. Variants were subsequently annotated for effect on the protein, known minor allele frequencies, and gene function using AnnoVar and custom in-house developed software. Candidate variants were verified and segregation examined using Sanger

capillary sequencing. Deleteriousness was determined either by mutation type (truncating, splicing, frameshifting) or with dbNSFP[2], a collection of 8 different in silico prediction methods. MAFs were calculated using Thousand Genomes[3], ESP[4] and ARIC data[5]. Which contain samples that are ethnically matched (non-Jewish, European origin Americans) to our POT1 families. Variants were prioritized based on minor allele frequency and deleteriousness (truncating and splicing).

We prioritized variants based on deleteriousness of the mutation (truncating and splice-affecting), quality of the mutation (>30% allele fraction in reads), gene function or previous association to cancer, and repeated observation in our dataset. POT1 was the only gene in which we found both a very rare (MAF ~0%), high-quality truncating mutation and any-other mutation in an unrelated family. It was further considered a good candidate based on our observation that mutations, of any kind, in this gene are very rare in a control, ethnically matched population. Finally, POT1 was also considered a good candidate because of the important role of telomeres in cancer biology.

High-throughput amplicon sequencing of POT1 gene

A total of 20 amplicons (124-270 bp) were designed to amplify the *POT1* gene in 354 samples. For each of the 20 primer pairs, the design included a 6-bp barcode sequence to serve as a sample identifier, with 12 unique barcodes and 240 primer pairs designed in total. PCR reactions were prepared using 10 ng of DNA with a Multiplex PCR Hot start kit (QIAGEN, Cat. no. 206105). Touchdown PCR was performed with the following amplification parameters: initial denaturation at 95°C for 15 min., followed by seven cycles of denaturation, annealing, and extension at 94°C for 30s, 57°C for 1.5 min., and 72°C for 1.5 min. The reaction then continued with 31 cycles of denaturation, annealing, and extension at 94°C for 30s, 55°C for 15 min., and 72°C for 15 min., followed by a final extension of 72°C for 10 min.

All 20 PCR products from a single sample were pooled together and then combined further for 9-12 samples/pool. Using 500ng of such a pool, libraries were prepared using the Ion Xpress Plus

Fragment Library Kit (Cat. no. 4471269) per manufacturer protocol. Briefly, amplicon DNA was processed through DNA End-Repair using the provided End-Repair Enzyme Mix and were then ligated with Ion Xpress Barcode Adapters (1-32; Cat. no. 4471250 and 4474009). Agencourt XP beads (Beckman Coulter Genomics, Inc., Cat. no. A63882) were used to purify DNA after each reaction step. Libraries were analyzed for concentration and size distribution using the Caliper GX 1K/12K/High Sensitivity Assay Labchip (Cat. no. 760517), and then pooled for sequencing (16 libraries/pool).

Library pools were prepared for sequencing on the Ion PGM platform using the manufacturer's protocols and reagents. Briefly, emulsion PCR reactions were performed with the Ion OneTouch 2 instrument and used a combination of the Ion PGM Template OT2 200 Kit and Ion PGM Sequencing 200 Kit v2 (Life Technologies, Cat. no. 4480974 and 4482006). Using the Ion OneTouch ES module, enrichment was then completed by selectively binding the Ion Sphere Particles (ISPs) containing amplified library fragments to streptavidin coated magnetic beads, removing empty ISPs through washing steps, and denaturing the library strands to allow for collection of the template-positive ISPs. After quantification using the Guava EasyCyte 5 (Millipore Technologies), approximately 25-35 million template-positive ISPs per run were deposited onto the Ion PGM 318C chip (Life Technologies, Cat. no. 4469497), and sequencing was performed according to the 500 flow run format.

Sanger validation

Mutation validation was performed with bidirectional sequencing of the selected sample sites. PCR reactions were prepared using 5 ng of genomic DNA, 0.4 µM oligonucleotide primers, and 0.7X Qiagen Multiplex Master Mix (Cat. no. 206145) containing HotStar Taq, buffer, and polymerase. Reactions were performed with and without QSOL PCR additive to enhance PCR and final sequence performance. Touchdown PCR was performed with the following parameters: 98^{o}C for 5 min. and 10 cycles of 98^{o}C for 30 sec., 72^{o}C for 30 sec. (decreasing by 1^{o}C per cycle), and 72^{o}C for 1 min. The reaction then continued with 30 cycles of 98^{o}C for 30 sec., 63^{o}C for 30 sec., and 72^{o}C for 1 min, followed

by a final extension at 72^{o}C for 5 min. The PCR products were purified with a 1:15 dilution of Exo-SAP, diluted by 0.6X, and then cycle-sequenced for 25 cycles using a 1/64th dilution of BigDye Terminator v3.1 reaction mix (Applied Biosystems, Cat. No. 4337456). Finally, reactions were precipitated with ethanol, resuspended in 0.1 mM EDTA, and analyzed on AB 3730xl sequencing instruments using the Rapid36 run module and 3xx base-caller. SNPs were identified using SNP Detector software and then visually validated with Consed.

References:

1. Gramatges, M. M. *et al.* Telomere content and risk of second malignant neoplasm in survivors of childhood cancer: a report from the childhood cancer survivor study. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **20,** 904–911 (2014).
2. Liu, X., Jian, X. & Boerwinkle, E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum. Mutat.* **34,** E2393–2402 (2013).
3. 1000 Genomes Project Consortium *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–1073 (2010).
4. Auer, P. L. *et al.* Imputation of exome sequence variants into population- based samples and blood-cell-trait-associated loci in African Americans: NHLBI GO Exome Sequencing Project. *Am. J. Hum. Genet.* **91,** 794–808 (2012).
5. Folsom, A. R. *et al.* Effect of 9p21 genetic variation on coronary heart disease is not modified by other risk markers. The Atherosclerosis Risk in Communities (ARIC) Study. *Atherosclerosis* **224,** 435–439 (2012).
6. Parker, M. *et al.* Assessing telomeric DNA content in pediatric cancers using whole-genome sequencing data. *Genome Biol.* **13,** R113 (2012).