



OPEN ACCESS

Original research

Targeted long-read sequencing identifies missing pathogenic variants in unsolved Werner syndrome cases

Danny E. Miller ^{1,2}, Lin Lee,³ Miranda Galey,² Renuka Kandhaya-Pillai,³ Marc Tischkowitz,⁴ Deepak Amalath,⁵ Avadh Vithlani,⁵ Koutaro Yokote,⁶ Hisaya Kato,⁶ Yoshiro Maezawa,⁶ Aki Takada-Watanabe,⁶ Minoru Takemoto,⁷ George M. Martin,³ Evan E. Eichler,^{2,8} Fuki M. Hisama,⁹ Junko Oshima^{3,6}

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/jmedgenet-2022-108485>).

For numbered affiliations see end of article.

Correspondence to

Dr Danny E. Miller, Department of Pediatrics, Division of Genetic Medicine, University of Washington, Seattle, WA 98195, USA; dm1@uw.edu and Dr Junko Oshima, Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA, USA; picard@uw.edu

Received 5 February 2022
Accepted 14 April 2022
Published Online First 9 May 2022

ABSTRACT

Background Werner syndrome (WS) is an autosomal recessive progeroid syndrome caused by variants in *WRN*. The International Registry of Werner Syndrome has identified biallelic pathogenic variants in 179/188 cases of classical WS. In the remaining nine cases, only one heterozygous pathogenic variant has been identified.

Methods Targeted long-read sequencing (T-LRS) on an Oxford Nanopore platform was used to search for a second pathogenic variant in *WRN*. Previously, T-LRS was successfully used to identify missing variants and analyse complex rearrangements.

Results We identified a second pathogenic variant in eight of nine unsolved WS cases. In five cases, T-LRS identified intronic splice variants that were confirmed by either RT-PCR or exon trapping to affect splicing; in one case, T-LRS identified a 339 kbp deletion, and in two cases, pathogenic missense variants. Phasing of long reads predicted all newly identified variants were on a different haplotype than the previously known variant. Finally, in one case, RT-PCR previously identified skipping of exon 20; however, T-LRS did not detect a pathogenic DNA sequence variant.

Conclusion T-LRS is an effective method for identifying missing pathogenic variants. Although limitations with computational prediction algorithms can hinder the interpretation of variants, T-LRS is particularly effective in identifying intronic variants.

INTRODUCTION

Segmental progeroid syndromes are a group of disorders that phenotypically resemble accelerated ageing.¹ The prototypical example of an adult-onset progeroid syndrome is Werner syndrome (WS; OMIM #277700), a rare autosomal recessive disorder caused by loss-of-function variants of the gene *WRN*.^{2,3} Individuals with WS typically do not show clinical signs until their early teens when the first clinical sign—lack of a growth spurt (often recognised retrospectively)—appears. An aged appearance (grey hair, atrophic skin) begins to develop in the 20s and 30s, which is followed by a series of common age-related diseases, including bilateral cataracts, gonadal atrophy, type II diabetes mellitus, osteoporosis and arteriosclerosis.^{3,4} Alzheimer-type dementia is generally not a feature

Key messages

What is already known on this topic

- ⇒ The burden of undiagnosed genetic disorders on individuals is high.
- ⇒ In some cases, a specific genetic disorder is suspected but no variants are found in genes associated with that disorder.
- ⇒ In other cases, testing identifies a single variant in a gene associated with the suspected disorder but no second variant.

What this study adds

- ⇒ This study reports the results of targeted long-read sequencing (T-LRS) in a cohort of individuals highly suspected to have a specific genetic disorder, Werner syndrome, but only one pathogenic variant identified in the causative gene, *WRN*.
- ⇒ This study demonstrates that T-LRS in a well-phenotyped cohort has a high yield.

How this study might affect research, practice or policy

- ⇒ The implications of this work are that in a well-defined cohort with a clear phenotype T-LRS represents an excellent next best test after non-diagnostic clinical testing.

of WS. The most common causes of death in WS are myocardial infarction and malignancies at a median age of 54 years.^{5,6} The most significant quality-of-life issue is development of deep ulcerations around the ankles and, occasionally, the elbows, which are associated with soft tissue calcifications and may eventually require lower limb amputation.⁷

WRN encodes a multifunctional nuclear protein with exonuclease and RecQ-type helicase domains.⁴ Molecular and cellular evidence suggest the involvement of *WRN* in a wide variety of functions, including DNA repair, recombination, replication and telomere maintenance.⁴ Cells derived from individuals with WS exhibit limited replicative lifespan, altered epigenetic signatures and mitochondrial dysfunction.^{4,8,9} More recently, *WRN* helicase activity was shown to be essential for the survival of mismatch repair-deficient cancer cell



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY. Published by BMJ.

To cite: Miller DE, Lee L, Galey M, et al. *J Med Genet* 2022;**59**:1087–1094.

lines, suggesting some functional overlap of these DNA repair pathways.¹⁰

To date, nearly 100 different pathogenic variants have been reported in individuals with WS worldwide.^{2,11} The majority of disease-causing variants in WS result in truncation of the WRN protein and the elimination of the nuclear localisation signal at the C-terminus and/or nonsense-mediated mRNA decay, making them functionally null.² That most variants result in little to no protein expression seems to be why all individuals with WS share similar phenotypes regardless of the causal variants. Several amino acid substitutions that abolish helicase activities have also been identified but only as compound heterozygous variants in combination with null variants.¹² Founder variants have been detected in specific populations, such as in Sardinia and Japan where carrier frequencies as high as 1:150 have been observed for specific variants.^{2,13} Possible founder mutations have also been reported in India, Pakistan, Turkey and the Netherlands.^{14,15} Differences in WS presentations among various populations have also been reported. For example, Indian/Pakistani individuals with WS tend to have earlier onset of cataracts, at a median age of 20 years compared with 31 years in non-Indian individuals with WS.¹⁴

Established in 1988, the International Registry of Werner Syndrome at the University of Washington recruits individuals with suspected progeroid syndromes from all over the world for molecular diagnosis and further biological study. Potential therapeutic approaches are being sought out in collaboration with the Japanese Werner Consortium.^{8,16,17} As of November 2021, the Registry has enrolled 179 individuals with classical WS with documented biallelic causal variants and 9 individuals with classical WS presentations in which only a single heterozygous causal variant has been identified. Recently, targeted long-read sequencing (T-LRS) on the Oxford Nanopore Technologies (ONT) platform was used to clarify complex structural variants and identify missing variants in cases that remained unsolved despite a complete clinical evaluation.¹⁸ We hypothesised that T-LRS could identify a second pathogenic variant in the unsolved WS cases. Of the nine molecularly unsolved cases from eight pedigrees in the registry, we identified a second pathogenic variant in eight. A second pathogenic variant was not identified in one case with known skipping of exon 20, despite long-read sequencing and phasing, demonstrating the limitation of both T-LRS and currently available prediction algorithms used to interpret DNA variants.

MATERIALS AND METHODS

Recruitment of study participants, sample processing and standard sequencing analysis

Individuals are referred to the International Registry of Werner Syndrome at the University of Washington by physicians who suspect a diagnosis of WS. Biological samples collected from individuals suspected to have WS who consent to be enrolled in the study are shipped to the International Registry.

Blood and skin sample processing was performed as previously described.⁵ Briefly, participant blood samples were processed immediately on arrival for cryopreservation of primary cells and plasma, isolation of DNA and RNA and establishment of lymphoblastoid cell lines (LCLs) using Epstein-Barr virus. Depending on the year of referral, different methods of DNA and RNA sequencing were done (table 1). In some cases, Sanger sequencing of some exons was done, while in others, sequencing was done on an exome backbone.^{2,5,15,19} RT-PCR sequencing was performed on total RNA isolated from blood or LCLs. Western blot analysis was done using total protein isolated from LCLs. Detailed protocols for Sanger sequencing and western blot analyses have been previously described.^{2,5,19}

Targeted long-read sequencing of the *WRN* locus

T-LRS was performed using ReadFish²⁰ on an ONT GridION as previously described.¹⁸ Briefly, for each sample 1–2 µg of genomic DNA was sheared using a Covaris g-TUBE by centrifuging at 6000 rpm for 2 min then inverting and centrifuging at 6000 rpm again for 2 min. DNA was prepared for sequencing using the ONT Ligation Kit (SQK-LSK109) following the manufacturer's instructions. Each library was loaded onto one or more R9.4.1 flow cells (FLO-MIN106D) and run for 72 hours with the goal of recovering at least 20× coverage per library (online supplemental table 1). For each sample, an approximately 3 Mbp region surrounding *WRN* was targeted along with two control regions (online supplemental table 2).

Long-read sequencing of RT-PCR products from individual SILV1010

To identify a second pathogenic variant in SILV1010, we performed overlapping RT-PCR of *WRN* mRNA using previously published primers.¹⁹ The RT-PCR product was prepared for ONT sequencing using the SQK-LSK109 ligation kit following the manufacturer's instructions. A single library was loaded

Table 1 Targeted long-read sequencing (T-LRS) identified candidate pathogenic variants in eight of nine individuals with Werner syndrome (BIA1010 and BIA1020 are siblings)

| Registry | Known variant (all heterozygous) | Additional workup | T-LRS result | Confirmation of T-LRS result |
|----------|---|---|-----------------------------------|------------------------------|
| WV | RT-PCR: c.2773delG, p.A925fs | Western: no protein | c.3961C>T, p.R1321* | PCR+Sanger |
| BIA1010 | Sanger: c.1105C>T, p.R369* | Western: no protein | c.3234-170A>G | RT-PCR |
| BIA1020 | Sanger: c.1105C>T, p.R369* | Western: no protein | c.3234-170A>G | RT-PCR |
| PD1010 | Exome sequencing: c.561A>G, p.K187* | Western: no protein | chr8:g.31135822_31474535delinsTCT | PCR+Sanger |
| CB4 | Sanger: c.3139-1G>C, p.G1047fs | Western: no protein SNP array: normal | c.1982-297A>G | RT-PCR |
| CB6 | Sanger: c.1105C>T, p.R369* | Western: no protein SNP array: normal | c.1982-297A>G | RT-PCR |
| FES | RT-PCR: c.1165delA, p.R389fs | Western: no protein | c.2103_2104delAC, p.Leu702fs | PCR+Sanger |
| SILV1010 | Sanger: c.2665C>T, p.R889* | Western: ~1% protein RT-PCR: skipping of exon 20 | No second variant found | n/a |
| EN1010 | Genome sequencing: c.2367_2368delAT, p.S790fs | None | c.839+1309T>G | Exon trap (figure 3) |

n/a, not available.

onto a Nanopore R9.4.1 Flongle flow cell (FLO-FLG001) and sequenced for 24 hours. Reads were aligned to GRCh38 using minimap2.²¹

Data analysis

Fast5 files were base called using the high accuracy model in Guppy 4.0.11 (ONT). Given no second candidate pathogenic variant was identified in SILV1010, it was re-base called using the superior model in Guppy 5.0.7 (ONT). For all samples, all FASTQ reads were aligned to GRCh38 using minimap2.²¹ Single-nucleotide and insertion/deletion variants (SNVs and indels) were called with Medaka (ONT), which was also used for phasing. Variant calling and phasing were redone using Guppy 5 data for SILV1010 using Clair3.²² Structural variants were identified using Sniffles,²³ SVIM²⁴ and CuteSV.²⁵ VEP V.103.1²⁶ was used to annotate VCF files produced by Guppy or Medaka and add splice predictions from SpliceAI,²⁷ CADD V.1.8 scores²⁸ and population allele frequency from the gnomAD V.2.1.1 genomes file.²⁹ Aligned reads were visualised using integrative genomics viewer.³⁰ Analysis was performed by filtering the VCF file by allele frequency and examining variants absent from population databases or those with allele frequencies <1%.

Exon trapping assay

The 1.8 kbp region that spans intron 7 to intron 8 (c.724-167 to c.839+1553) was amplified from genomic DNA of individual EN1010 and subcloned into the multicloning sites of the pSPL3 vector^{31,32} linearized with EcoRI and PstI, using the NEBuilder HiFi DNA assembly cloning kit (catalogue# E5520, New England Biolabs) to generate pSPL3-EN1010-Wt and pSPL3-EN1010-Mut. These constructs were transfected to immortalised control human fibroblasts, 82-6hTERT, using FuGENE 6 (catalogue# E2693, Promega). After 48 hours, total RNA was isolated and reverse-transcribed with High-Capacity cDNA Reverse Transcription Kit (catalogue# 4368814, Thermo Fisher Scientific). RT-PCR was done using primers: SD6 (5'-TCTGAGT-CACCTGGACAACC-3'), SA2 (5'-ATCTCAGTGGTATTTG

TGAGC-3') and one designed from WRN exon 8, J20 (5'-CTGA GGAAGTGATGGATCTGG-3').

RESULTS

Individuals sequenced in this report

Seven of the nine individuals with a single heterozygous variant in WRN have been previously described.² All nine individuals were referred to the Registry with a clinical diagnosis of WS (table 2), two of which are newly described below. Variants in WRN were initially identified by Sanger sequencing of coding exons, which, over time, was gradually replaced by clinical exome or genome sequencing. A western blot analysis for WRN was done in cases for which material was available. RT-PCR and SNP arrays were also used in selected cases.^{2,5,15}

Individual EN1010 is a man in his 50s born to non-consanguineous parents with a birth weight of 2.7 kg. He began to experience signs of premature ageing in his 20s with greying hair. He underwent bilateral cataract surgery in his early 30s and was subsequently diagnosed with hypothyroidism and type 2 diabetes mellitus in his 30s. He was diagnosed with hypogonadism and a prolactinoma in his 40s. His other medical history includes cardiac atherosclerosis, valve insufficiency and tinnitus.

On physical examination, his body mass index (BMI) was 20.2 kg/m², height of 168 cm (Z-score -1.2) and weight of 57 kg (Z-score of -1.34). He had a mildly elevated blood pressure (148/62) and appeared progeroid, with grey hair, vocal cord atrophy, tight, atrophic skin with a waxy appearance and ulcers of the right heel and left metatarsal. His limbs were thin and notable for cold fingers, reduced subcutaneous fat, muscle atrophy and flat feet. Dual-energy X-ray absorptiometry scan revealed osteoporosis, osteosclerosis of the fingers and toes and soft tissue calcification. He had elevated liver enzymes with hepatic steatosis observed on abdominal MRI.

His family history includes two unaffected siblings: a sister in her 40s who is 160 cm tall (Z-score -0.5) and a brother in his 50s who is 170 cm tall (Z-score -0.92); each has one child. The father is in his 80s; his height is 180 cm (Z-score 0.48). The

Table 2 Clinical signs of Werner syndrome in seven previously reported cases

| Registry# | EN1010 | PD1010 | BIA1010 | BIA1020 | SILV1010 | CB4 | WV |
|---|----------|----------|----------|----------|----------|----------|----------|
| Age at evaluation (years) | 42 | 42 | 45 | 38 | 68 | 40s | 40 |
| Cardinal signs | | | | | | | |
| Cataracts | Y | Y | Y | Y | Y* | Y | Y |
| Skin and facial feature | Y | Y | Y | Y | N | Y | Y |
| Short stature | N | Y | Y | Y | N | Y | Y |
| Greying or loss of hair | Y | Y | Y | Y | Y | Y | Y |
| Parental consanguinity or affected sibs | N | N | Y | Y | Y | N | NA |
| Further signs and symptoms | | | | | | | |
| Diabetes mellitus | N | Y | Y | Y | Y | Y | N |
| Hypogonadism | N | NA | Y | Y | NA | NA | Y |
| Osteoporosis | Y | Y | Y | Y | NA | NA | Y |
| Osteosclerosis | N | NA | Y | NA | NA | NA | NA |
| Soft tissue calcification | N | N | Y | Y | Y | Y | Y |
| Atherosclerosis | N | NA | Y | NA | NA | Y | NA |
| Neoplasms | N | NA | Y | NA | Y | N | NA |
| Voice change | Y | Y | Y | Y | Y | Y | Y |
| Flat feet | N | N | Y | Y | Y | Y | NA |
| Werner syndrome diagnostic criteria | Possible | Possible | Definite | Definite | Possible | Probable | Probable |

*Unilateral cataract.

N, no; NA, not available; Y, yes.

mother is in her 70s; her height is 160 cm (Z-score -0.5) and she has hypothyroidism and a pacemaker. Although he lacked short stature, this individual had three cardinal signs of WS (cataracts, dermatological pathology and premature greying) as well as additional clinical features, including type 2 diabetes mellitus and atherosclerosis; thus, this individual met criteria for possible diagnosis of WS diagnosis.³ Genome sequencing revealed a novel heterozygous *WRN* variant, c.2367_2368delAT, in exon 20, which results in the truncation of the *WRN* protein, p.Ser790fs (table 1). A second pathogenic variant was not reported.

Individual PD1010 is a South Asian man in his 40s born to non-consanguineous parents who was diagnosed with Addison disease in his teens, hypothyroidism and type 2 diabetes mellitus in his 20s (which later became insulin-dependent) and bilateral cataracts requiring surgery in his 30's. On physical examination, his BMI was 17.8 kg/m², height was 150 cm (Z-score -3.54) and weight was 40 kg (Z-score -4.3). He had a progeroid appearance, with sparse grey hair, a high-pitched squeaky voice and atrophic, thin skin. His limbs were thin, with loss of subcutaneous fat, muscle wasting and ulcers at both elbows. He had no siblings.

Exome sequencing revealed a heterozygous variant in exon 6 of *WRN*, c.561A>G (table 1). This synonymous variant is a known founder mutation seen in individuals of Indian or Pakistani descent and has been shown to activate a cryptic splice site in exon 6, resulting in p.Lys187fs.¹⁴ A second pathogenic variant was not identified by exome sequencing. Western blot analysis of protein from LCLs showed no detectable *WRN* protein, confirming the diagnosis of WS.

Detection of deep intronic splice variants by T-LRS

We performed T-LRS on eight of the nine molecularly unsolved cases with an average depth of coverage of $20\times$ (figure 1A, online supplemental table 1). In all eight sequenced cases, the known pathogenic variant was identified. In four of the eight sequenced cases, T-LRS identified an intronic variant predicted by SpliceAI²⁷ to alter splicing (figure 2, table 1, online supplemental table 3). One individual was not sequenced because a splice variant was found in his affected brother (BIA1010 and BIA1020). In all four cases, phasing predicted that the intronic splice variant is on a different haplotype than the previously identified pathogenic variant.

Individuals BIA1010 and BIA1020 were affected siblings from Poland carrying a heterozygous c.1105C>T, p.Arg369*, a known founder variant, accounting for approximately 20% of the mutant alleles in our registry.² T-LRS revealed a heterozygous c.3234-170A>G in intron 26, which was predicted by SpliceAI to create a cryptic splice acceptor site (AA→AG). RT-PCR of the region including exons 25 and 26 showed two abnormal transcripts, one with a 69 bp insertion (r.3233_3234ins69) and the other with a 169 bp insertion (r.3233_3234ins169) between exons 25 and 26 (online supplemental figure 1). Both cryptic exons started at c.3234-169, as predicted by in silico analysis. Within the family, PCR and Sanger sequencing confirmed that the two variants were from the parents: c.3234-170A>G was transmitted from the mother (BIA1040) and c.1105C>T from the father (BIA1050).

In two unrelated Japanese individuals with WS, CB4 and CB6, T-LRS detected a c.1982-297A>G in intron 17. In silico analysis

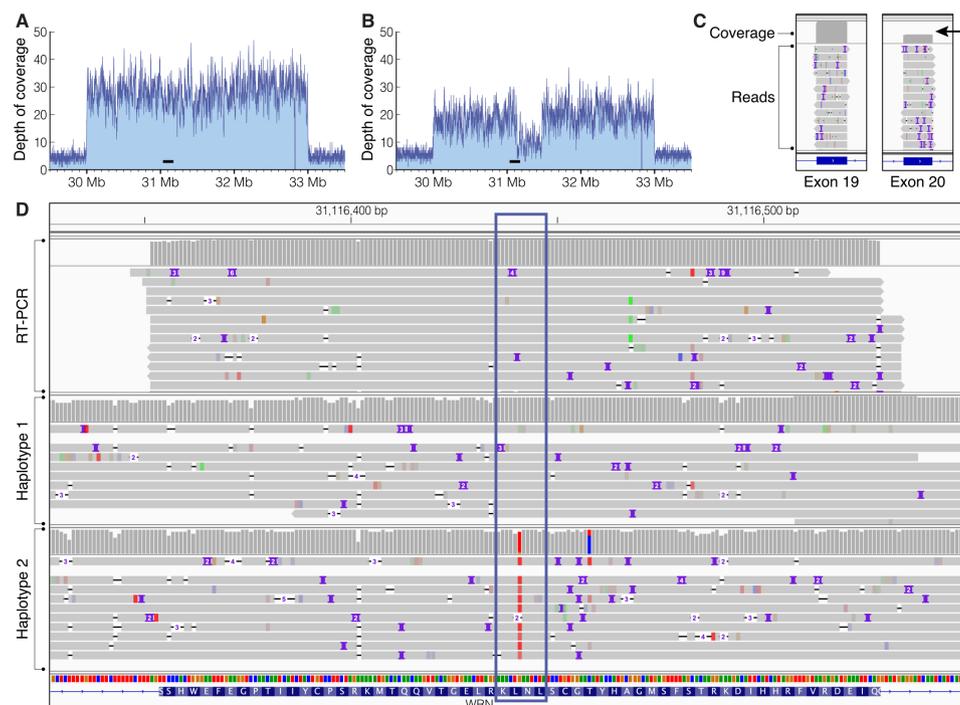


Figure 1 Targeted long-read sequencing (T-LRS) was used to identify missing variants. (A) T-LRS was used to target an approximately 3 Mbp region around *WRN*. Data shown for individual WV. Increased coverage (y-axis) represents the target region, decreased coverage represents background. The position of *WRN* is represented by the horizontal black bar around 31 Mbp (GRCh38). (B) T-LRS revealed a 339 kbp deletion in individual PD1010 that began within *WRN*. (C) Long-read sequencing of RT-PCR products from individual SIV1010 confirmed the absence of exon 20 in approximately half of the reads (arrow). Please see online supplemental figure 3 for an integrative genomics viewer screenshot of exons 19–21 that shows how reads are linked and that approximately half of reads skip of exon 20. (D) Long-read sequencing revealed RT-PCR products from individual SIV1010 did not carry the T allele present on haplotype 2 within exon 20, confirming that splicing was altered on haplotype 2. The previously known pathogenic variant (table 1) was predicted to be located on haplotype 1. No second pathogenic variant was identified in this individual.

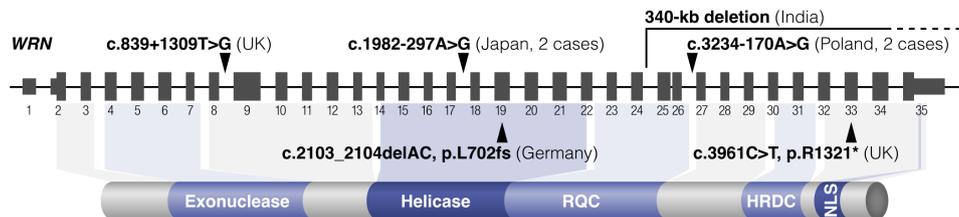


Figure 2 Diagram of *WRN* and the location of variants identified in this study. Pathogenic variants found in this study are shown with respect to the 35 exons within *WRN*. Functional domains of *WRN* proteins are shown below the corresponding exons. Those functional domains are the exonuclease and helicase domains, RecQ helicase conserved region (RQC), helicase RNaseD C-terminal conserved region (HRDC) and the nuclear localisation signal (NLS).

predicted that this substitution activates the adjacent GT to be the cryptic splice donor (A|GT→G|GT). RT-PCR sequencing revealed the presence of the corresponding 73 bp cryptic exon between exons 17 and 18 (r.1981_1982ins73) in CB6. RT-PCR was not done for individual CB4 because there was insufficient remaining mRNA.

For individual EN1010, T-LRS detected a c.839+1309T>G in intron 8, along with the previously identified c.2367_2368delAT in exon 20. This intron 8 variant has a CADD of 9.3 and SpliceAI predicts it creates a new donor site (T|GT→G|GT) with a score of 0.73. Unfortunately, we were unable to amplify RT-PCR products directly from the individuals materials or establish an immortalised cell line from the individual sample. We, therefore, performed exon trapping to test whether this variant altered splicing and generated a cryptic exon. Using the pSPL3 system,³¹ we observed that RT-PCR of the wildtype construct (pSPL3-EN1010-Wt) gave a single band with the expected exon 8 splicing, while the construct with the c.839+1309T>G variant (pSPL3-EN1010-Mut) gave an RT-PCR product with a 171 bp insertion following exon 8 (r.839_840ins171) (figure 3). The inserted 171 bp corresponded to c.893+1139 to c.893+1309. This insertion is predicted to cause premature termination of the *WRN* protein at position 280 (p.Arg280SerfsTer9).

T-LRS identified a previously unknown large deletion and missense variants

T-LRS of individual PD1010 revealed a 338 715 bp deletion that began within *WRN* and included exons 25 through 35 (figures 1B

and 2). PCR and Sanger sequencing of the predicted junction confirmed the deletion as chr8:g.31,135,822_31,474,535delinsTCT (online supplemental figure 2). Encouraged by the efficacy of T-LRS, two archived cases, individuals WV and FES, were then re-examined.³³ Western blot analysis of these cases showed no *WRN* protein, and RT-PCR in the mid-1990s revealed a single heterozygous variant in both cases.¹⁹ T-LRS identified a heterozygous pathogenic coding variant, c.3961C>T, p.Arg1321*, in exon 33 of individual WV, and c.2103_2104delAC, p.Leu702fs, in exon 19 of individual FES (figure 2). In all three cases, phasing predicted that the previously known pathogenic variant was on a different haplotype than the newly identified second variant. That we identified variants in coding sequence in these cases was not surprising as they were enrolled in the registry prior to the availability of PCR and Sanger sequencing.

T-LRS failed to identify a pathogenic variant in a case with known exon skipping

Individual SILV1010, a man from the USA whose presentation and phenotype is consistent with possible WS, was found to have a pathogenic heterozygous c.2665C>T, p.Arg889* in exon 22. Western blot analysis on lymphoblastoid cells demonstrated 1% *WRN* protein expression, consistent with WS. RT-PCR of material from a LCL derived from the affected individual showed heterozygous skipping of exon 20 (r.2274_2448del175) (figure 1C).² T-LRS of DNA from this individual detected the previously identified c.2665C>T, which was predicted to be on haplotype 1 (figure 1D), and a second variant (c.1269+36A>G)

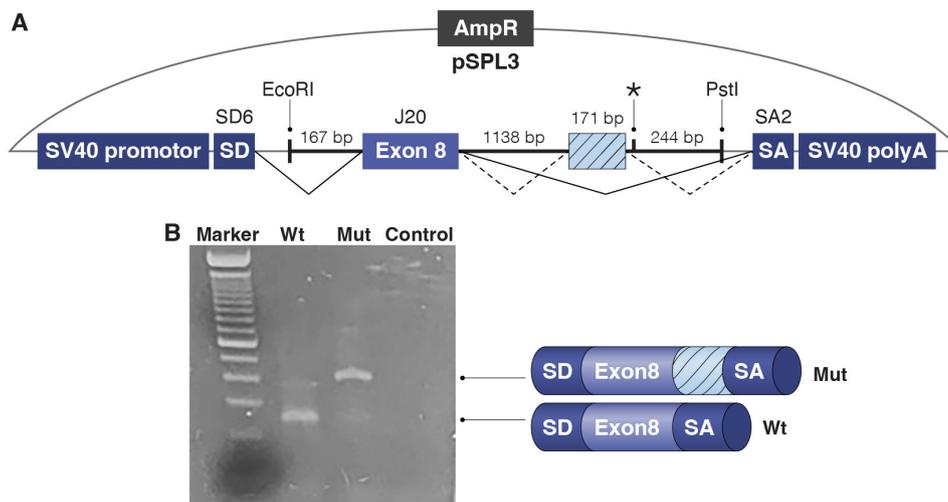


Figure 3 Exon trapping of the EN1010 variant. (A) The pSPL3 vector contains splice donor (SD) and splice acceptor (SA) exons and functional introns, with transcription beginning following the SV40 promoter and ending at the poly(A) site. The wildtype (Wt) construct, pSPL3-EN1010-Wt, and the mutant (Mut) construct, pSPL3-EN1010-Mut, contain the 1.8 kbp fragment derived from EN1010, with the c.839+1309T>G variant at the asterisk (*). (B) Agarose gel electrophoresis of RT-PCR products shows the larger amplicon in the Mut compared with Wt due to the 171 bp cryptic insertion.

located in the ninth intron predicted by SpliceAI to affect splicing by creating a new donor site with a score of 0.45 and loss of a donor site at c.1269 with a score of 0.12. However, we did not detect splice alteration involving exon 9, intron 9 or exon 10 in LCLs from the affected individual, nor did we find pathological significance linked to this variant in literature or databases. Thus, we felt this variant did not explain skipping of exon 20 in this individual. Because exon 20 contained a heterozygous c.2361G>T, we performed WGS of RT-PCR products on an ONT Flongle and found that approximately half of reads skipped exon 20 (online supplemental figure 3) and that those RT-PCR products which contained exon 20 were homozygous for the reference allele G and did not contain the T allele that was predicted to be located on haplotype 2 (figure 1D). A single SNV on haplotype 2 with an allele frequency <1% and absent from population databases was found in either intron 19 or 20 (c.2273+155A>G). Unfortunately, this variant is not computationally predicted to alter splicing. Thus, sequencing of RT-PCR products did not reveal a clear potentially pathogenic variant, and this case remains unsolved at the molecular level.

DISCUSSION

Here, we show that computational selection of specific genomic regions for sequencing using adaptive sampling on the ONT platform can be used to identify disease-causing variants in a cohort of individuals with WS. Different approaches for T-LRS have been described using both PCR and CRISPR/Cas-based methods.^{34–36} PCR-based methods allow for targeting by over-enrichment of target regions but remove epigenetic information from the target and are limited by the length of the fragment that can be reliably amplified by PCR. CRISPR/Cas-based methods work by first dephosphorylation of genomic DNA, then exposing new phosphorylation sites with targeted double-stranded breaks where sequencing libraries then anneal. The advantage of the CRISPR/Cas method is that epigenetic information is preserved while segments up to about 200 kbp can be targeted. The drawback is that multiple regions must be targeted if one wants to target several megabase pairs of genomic space, and recovery of large fragments may require gel-based separation of cut DNA fragments. In both cases, fragments can be sequenced on either an ONT or PacBio platform.

Recently, adaptive sampling on the ONT platform was shown to be an effective method to evaluate known structural variants and missing variant cases.¹⁸ This method works by computationally selecting DNA molecules for sequencing.²⁰ There are several advantages of adaptive sampling over CRISPR/Cas and PCR-based methods, including broader flexibility in selecting target sites, ability to target a larger amount of genomic space and simpler library preparation steps without the need for DNA fragment isolation by pulse-field gel if large segments are targeted. However, ONT sequencing has a higher per-read error rate than PacBio high-fidelity (HiFi) sequencing because each DNA molecule is read only one time by ONT instead of multiple times as with PacBio HiFi.³⁷ This error is most apparent in homopolymers, meaning that certain classes of variants are more likely to be missed by ONT sequencing.

Using adaptive sampling, we identified a second pathogenic variant (including missense, deep intronic splice and structural variants) in eight of nine individuals clinically diagnosed with WS and a single known pathogenic variant identified by prior testing. While all eight variants identified using T-LRS would likely also have been identified by short-read genome sequencing (WGS) and focused analysis of *WRN* there may be a

cost advantage of T-LRS over short-read WGS in missing variant cases such as these. Each individual was sequenced on a single ONT R9.4.1 flowcell after a single library preparation, representing a materials cost of approximately US\$600 per sample which is cost competitive with short-read WGS today. Exome sequencing, on the other hand, would likely identify the two missense variants but may have difficulty identifying the large deletion in individual PD1010 and would be unable to identify deep intronic variants because those regions are not typically targeted for capture or enrichment.^{38 39}

The major advantage of our approach over short-read WGS is the ability to resolve repetitive regions and phase samples for which parental samples may not be available. *WRN* contains two 1400 bp segmental duplications involving exons 10 and 11 as well as surrounding intronic sequence that are approximately 95% similar. Prior work using PCR to identify pathogenic indels and nonsense variants in exons 10 and 11 depended on a single-nucleotide difference between these two regions to design region-specific primers.² Thus, there is a technical advantage of T-LRS over short-read WGS to cover such regions.⁴⁰ Furthermore, unlike short-read WGS, T-LRS is able to more easily phase variants in the absence of parental samples for segregation, representing an additional advantage of T-LRS over short-read WGS. Thus, a strong argument exists for T-LRS the next best test for individuals with non-diagnostic clinical testing and either a compelling phenotype for a specific disorder or a single known variant in a gene associated with the suspected disorder.

That four of the missing variants identified in our study were splice variants demonstrate how difficult these can be to detect in the laboratory. Here, in silico analysis allowed us to identify candidate splice variants and allowed for targeted evaluation in our unsolved cases. Interestingly, the highest reported heterozygote frequency for a pathogenic WS variant is a deep intronic splice variant, c.2089-3024A>G, r.2089_2273del85 found in approximately 1:120 individuals in Sardinia.^{13 19}

Despite advances in the ability of computational prediction algorithms to identify intronic variants that may alter splicing, it is still necessary to confirm the pathogenicity of these variants in the laboratory. When material from an affected individual is unavailable, CRISPR may be used to generate a cell line containing the candidate variants. Intronic regions, however, generally contain repetitive sequence that makes CRISPR mutagenesis more challenging. Therefore, in this case we used exon trapping as an alternative approach to validate the candidate variant identified in individual EN1010.^{31 32} Although exon trapping has been widely used for this purpose, there is a formal possibility that natural introns and exons may have an unknown influence on cryptic splicing.

Worldwide, WS is thought to be clinically underdiagnosed because many of the symptoms are relatively non-specific and may be viewed as common age-related diseases. We believe that additional genetic testing, including long-read sequencing, will lead to an increased rate of diagnosis for individuals with WS. This increased diagnostic rate and awareness will help drive interest in research and development of therapeutics, thereby benefiting all individuals affected by this disorder.

Author affiliations

¹Department of Pediatrics, Division of Genetic Medicine, University of Washington, Seattle, Washington, USA

²Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington, USA

³Department of Laboratory Medicine and Pathology, University of Washington, Seattle, Washington, USA

⁴Department of Medical Genetics, National Institute for Health Research Cambridge Biomedical Research Centre, University of Cambridge, Cambridge, UK

⁵Department of Medicine, Jawaharlal Institute of Postgraduate Medical Education and Research, Puducherry, India

⁶Department of Endocrinology, Hematology and Gerontology, Chiba University Graduate School of Medicine, Chiba, Japan

⁷Department of Diabetes, Metabolism and Endocrinology, International University of Health and Welfare, Otawara, Japan

⁸Howard Hughes Medical Institute, University of Washington, Seattle, Washington, USA

⁹Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, Washington, USA

Twitter Danny E. Miller @danrdanny

Acknowledgements We thank A. Miller for help with figure preparation and editorial assistance and T. Brown for editorial assistance. We wish to thank Dr Anna Stears, National Severe Insulin Resistance Service, Addenbrookes Hospital, Cambridge, UK.

Contributors Conception: DEM, JO; data curation: DEM, JO; formal analysis: DEM, LL, MG, JO; funding acquisition: DEM, GMM, EEE, JO; investigation: DEM, LL, MG, RK-P., MT, DA, KY, HK, YM, AT-W, MT, AV; methodology: DEM; supervision: DEM, GMM, EEE, JO; writing—original draft: DEM, FMH, JO; writing—review and editing: DEM, EEE, FMH, JO. DEM and JO accepted full responsibility for the finished work and/or the conduct of the study, had access to the data, and controlled the decision to publish. All authors reviewed the final manuscript.

Funding This work was in part supported by the US National Institutes of Health (NIH) grants R01CA210916 (GMM/JO) and R01MH101221 (EEE), a Catalytic Collaborations Grant from the Brotman Baty Institute for Precision Medicine (DEM/EEE), a Trainee Award from the Brotman Baty Institute for Precision Medicine (DEM), JP20H00524 from the Japan Society for the Promotion of Science (JSPS) KAKENHI (KY) and JPMH21FC1016 from the Ministry of Health, Labour and Welfare (MHLW) of Japan (KY). MT is supported by the NIHR Cambridge BRC. Analysis of the sequence data was supported by NHGRI grants U01 HG011744 and UM1 HG006493 (University of Washington Center for Rare Disease Research). EEE is an investigator of the Howard Hughes Medical Institute. This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a non-exclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the author-accepted manuscript of this article can be made freely available under a CC BY 4.0 license immediately on publication.

Competing interests DEM has received travel support from Oxford Nanopore Technologies (ONT) to speak on their behalf. DEM is a paid consultant for and holds stock options in MyOme. DEM and EEE are engaged in a research agreement with ONT. EEE is a scientific advisory board (SAB) member of Variant Bio, Inc.

Patient consent for publication Not applicable.

Ethics approval Research participants were referred to the International Registry of Werner Syndrome by referring physicians. Individuals or their legal representative signed informed consent for the molecular and cell biological studies, publication of the results and deposition of data. Studies were performed in accordance with the Declaration of Helsinki protocol and were approved by the institutional review board (IRB) at the University of Washington, Seattle, Washington. Study number is STU00000233 and is approved through 16 October 2022. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Data that support the findings of this study are available on request from the corresponding authors.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

ORCID iD

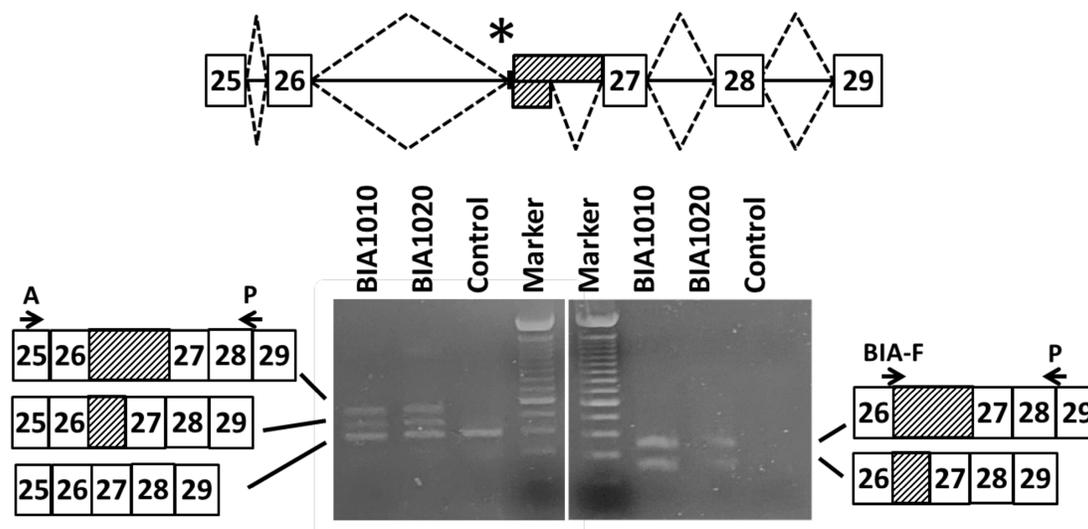
Danny E. Miller <http://orcid.org/0000-0001-6096-8601>

REFERENCES

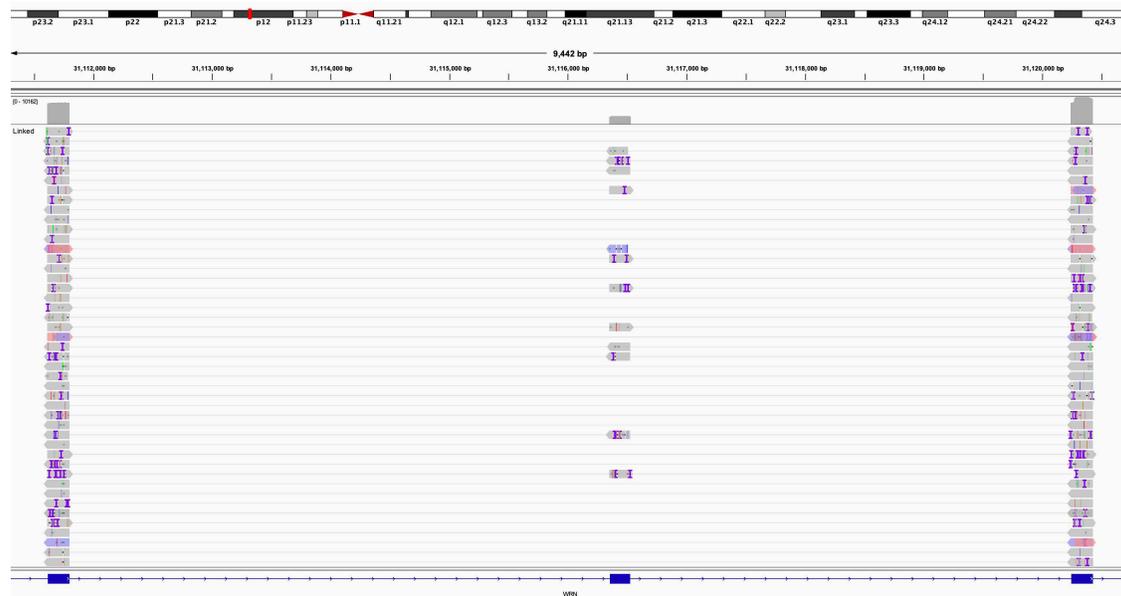
- Martin GM, Hisama FM, Oshima J. Review of how genetic research on segmental progeroid syndromes has documented genomic instability as a hallmark of aging but let us now Pursue *Antigeroid* syndromes! *J Gerontol A Biol Sci Med Sci* 2021;76:253–9.
- Yokote K, Chanprasert S, Lee L, Eirich K, Takemoto M, Watanabe A, Koizumi N, Lessel D, Mori T, Hisama FM, Ladd PD, Angle B, Baris H, Cefle K, Palanduz S, Ozturk S, Chateau A, Deguchi K, Easwar TKM, Federico A, Fox A, Grebe TA, Hay B, Nampoothiri S, Seiter K, Streeten E, Piña-Aguilar RE, Poke G, Poot M, Posmyk R, Martin GM, Kubisch C, Schindler D, Oshima J. Wrn mutation update: mutation spectrum, patient registries, and translational prospects. *Hum Mutat* 2017;38:7–15.
- Oshima J, Martin GM, Hisama FM. Werner syndrome. GeneReviews. Available: <https://www.ncbi.nlm.nih.gov/books/NBK1514/> [Accessed 13 May 2021].
- Oshima J, Sidorova JM, Monnat RJ. Werner syndrome: clinical features, pathogenesis and potential therapeutic interventions. *Ageing Res Rev* 2017;33:105–14.
- Huang S, Lee L, Hanson NB, Lenaerts C, Hoehn H, Poot M, Rubin CD, Chen D-F, Yang C-C, Juch H, Dorn T, Spiegel R, Oral EA, Abid M, Battisti C, Lucci-Cordisco E, Neri G, Steed EH, Kidd A, Isley W, Showalter D, Vittone JL, Konstantinow A, Ring J, Meyer P, Wenger SL, von Herbay A, Wollina U, Schuelke M, Huizenga CR, Leistriz DF, Martin GM, Mian IS, Oshima J. The spectrum of WRN mutations in Werner syndrome patients. *Hum Mutat* 2006;27:558–67.
- Goto M, Ishikawa Y, Sugimoto M, Furuichi Y. Werner syndrome: a changing pattern of clinical manifestations in Japan (1917–2008). *Biosci Trends* 2013;7:13–22.
- Takemoto M, Mori S, Kuzuya M, Yoshimoto S, Shimamoto A, Igarashi M, Tanaka Y, Miki T, Yokote K. Diagnostic criteria for Werner syndrome based on Japanese nationwide epidemiological survey. *Geriatr Gerontol Int* 2013;13:475–81.
- Fang EF, Hou Y, Lautrup S, Jensen MB, Yang B, SenGupta T, Caponio D, Khezri R, Demarest TG, Aman Y, Figueroa D, Morevati M, Lee H-J, Kato H, Kassahun H, Lee J-H, Filippelli D, Okur MN, Mangerich A, Croteau DL, Maezawa Y, Lyssiotis CA, Tao J, Yokote K, Rusten TE, Mattson MP, Jasper H, Nilsen H, Bohr VA. NAD⁺ augmentation restores mitochondria and limits accelerated aging in Werner syndrome. *Nat Commun* 2019;10:5284.
- Maierhofer A, Flunkert J, Oshima J, Martin GM, Poot M, Nanda I, Dittrich M, Müller T, Haaf T. Epigenetic signatures of Werner syndrome occur early in life and are distinct from normal epigenetic aging processes. *Aging Cell* 2019;18:e12995.
- Chan EM, Shibue T, McFarland JM, Gaeta B, Ghandi M, Dumont N, Gonzalez A, McPartlan JS, Li T, Zhang Y, Bin Liu J, Lazaro J-B, Gu P, Piatt CG, Apfel A, Ali SO, Deasy R, Keskuła P, Ng RWS, Roberts EA, Reznichenko E, Leung L, Alimova M, Schenone M, Islam M, Maruvka YE, Liu Y, Roper J, Raghavan S, Giannakis M, Tseng Y-Y, Nagel ZD, D'Andrea A, Root DE, Boehm JS, Getz G, Chang S, Golub TR, Tsherniak A, Vazquez F, Bass AJ. Wrn helicase is a synthetic lethal target in microsatellite unstable cancers. *Nature* 2019;568:551–6.
- Mojumdar A. Mutations in conserved functional domains of human RecQ helicases are associated with diseases and cancer: a review. *Biophys Chem* 2020;265:106433.
- Tadokoro T, Rybanska-Spaeder I, Kulikowicz T, Dawut L, Oshima J, Croteau DL, Bohr VA. Functional deficit associated with a missense Werner syndrome mutation. *DNA Repair* 2013;12:414–21.
- Masala MV, Scapaticci S, Olivieri C, Pirodda C, Montesu MA, Cuccuru MA, Pruneddu S, Danesino C, Cerimele D. Epidemiology and clinical aspects of Werner's syndrome in North Sardinia: description of a cluster. *European J Dermatology Ejd* 2007;17:213–6.
- Saha B, Lessel D, Nampoothiri S, Rao AS, Hisama FM, Peter D, Bennett C, Nürnberg G, Nürnberg P, Martin GM, Kubisch C, Oshima J. Ethnic-Specific WRN mutations in South Asian Werner syndrome patients: potential founder effect in patients with Indian or Pakistani ancestry. *Mol Genet Genomic Med* 2013;1:7–14.
- Friedrich K, Lee L, Leistriz DF, Nürnberg G, Saha B, Hisama FM, Eymann DK, Lessel D, Nürnberg P, Li C, Garcia-F-Villalta MJ, Kets CM, Schmidtke J, Cruz VT, Van den Akker PC, Boak J, Peter D, Compoginis G, Cefle K, Ozturk S, López N, Wessel T, Poot M, Ippel PF, Groff-Kellermann B, Hoehn H, Martin GM, Kubisch C, Oshima J. Wrn mutations in Werner syndrome patients: genomic rearrangements, unusual intronic mutations and ethnic-specific alterations. *Hum Genet* 2010;128:103–11.
- Koshizaka M, Maezawa Y, Maeda Y, Shoji M, Kato H, Kaneko H, Ishikawa T, Kinoshita D, Kobayashi K, Kawashima J, Sekiguchi A, Motegi S-I, Nakagami H, Yamada Y, Tsukamoto S, Taniguchi A, Sugimoto K, Shoda Y, Hashimoto K, Yoshimura T, Suzuki D, Kuzuya M, Takemoto M, Yokote K. Time gap between the onset and diagnosis in Werner syndrome: a nationwide survey and the 2020 registry in Japan. *Aging* 2020;12:24940–56.
- Takemoto M, Yokote K. Preface to management guideline for Werner syndrome 2020. *Geriatr Gerontol Int* 2021;21:131–2.
- Miller DE, Sulovari A, Wang T, Loucks H, Hoekzema K, Munson KM, Lewis AP, Fuerte EPA, Paschal CR, Walsh T, Thies J, Bennett JT, Glass I, Dipple KM, Patterson K, Bonkowski ES, Nelson Z, Squire A, Sikes M, Beckman E, Bennett RL, Earl D, Lee W, Allikmets R, Perlman SJ, Chow P, Hing AV, Wenger TL, Adam MP, Sun A, Lam C, Chang I, Zou X, Austin SL, Huggins E, Safi A, Iyengar AK, Reddy TE, Majoros WH, Allen AS, Crawford GE, Kishnani PS, King M-C, Cherry T, Chong JX, Bamshad MJ, Nickerson DA,

- Mefford HC, Doherty D, Eichler EE, University of Washington Center for Mendelian Genomics. Targeted long-read sequencing identifies missing disease-causing variation. *Am J Hum Genet* 2021;108:1436–49.
- 19 Oshima J, Yu CE, Piussan C, Klein G, Jabkowski J, Balci S, Miki T, Nakura J, Ogihara T, Ellis J, Smith M, Melaragno MI, Fraccaro M, Scappaticci S, Matthews J, Ouais S, Jarzebowicz A, Schellenberg GD, Martin GM. Homozygous and compound heterozygous mutations at the Werner syndrome locus. *Hum Mol Genet* 1996;5:1909–13.
- 20 Payne A, Holmes N, Clarke T, Munro R, Debebe BJ, Loose M. Readfish enables targeted nanopore sequencing of gigabase-sized genomes. *Nat Biotechnol* 2020:1–9.
- 21 Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;34:3094–100.
- 22 Luo R, Wong C-L, Wong Y-S, Tang C-I, Liu C-M, Leung C-M, Lam T-W. Exploring the limit of using a deep neural network on pileup data for germline variant calling. *Nat Mach Intell* 2020;2:220–7.
- 23 Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* 2018;15:461–8.
- 24 Heller D, Vingron M. SVIM: structural variant identification using mapped long reads. *Bioinformatics* 2019;35:btz041:2907–15.
- 25 Jiang T, Liu Y, Jiang Y, Li J, Gao Y, Cui Z, Liu Y, Liu B, Wang Y. Long-read-based human genomic structural variation detection with cuteSV, 2020. Available: <https://genomebiology.biomedcentral.com/track/pdf/10.1186/s13059-020-02107-y>
- 26 McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. The Ensembl variant effect predictor. *Genome Biol* 2016;17:122.
- 27 Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, Kosmicki JA, Arbelaez J, Cui W, Schwartz GB, Chow ED, Kanterakis E, Gao H, Kia A, Batzoglu S, Sanders SJ, Farh KK-H. Predicting splicing from primary sequence with deep learning. *Cell* 2019;176:e24:535–48.
- 28 Rentsch P, Witten D, Cooper GM, Shendure J, Kircher M. Cadd: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* 2019;47:D886–94.
- 29 Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferreira S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Neale BM, Daly MJ, MacArthur DG, Salinas CAA, Ahmad T, Albert CM, Ardissino D, Atzmon G, Barnard J, Beaugerie L, Benjamin EJ, Boehnke M, Bonycastle LL, Bottinger EP, Bowden DW, Bown MJ, Chambers JC, Chan JC, Chasman D, Cho J, Chung MK, Cohen B, Correa A, Dabelea D, Darbar D, Duggirala R, Dupuis J, Ellinor PT, Elosua R, Erdmann J, Esko T, Färkkilä M, Florez J, Franke A, Getz G, Glaser B, Glatt SJ, Goldstein D, Gonzalez C, Groop L, Haiman C, Hanis C, Harms M, Hiltunen M, Holi MM, Hultman CM, Kallela M, Kaprio J, Kathiresan S, Kim B-J, Kim YJ, Kirov G, Koener J, Koskinen S, Krumholz HM, Kugathasan S, Kwak SH, Laakso M, Lehtimäki T, Loos RJF, Lubitz SA, Ma RCW MDG, Marrugat J, Mattila KM, McCarroll S, McCarthy MI, McGovern D, McPherson R, Meigs JB, Melander O, Metspalu A, Nilsson PM, O'Donovan MC, Ongur D, Orozco L, Owen MJ, Palmer CNA, Palotie A, Park KS, Pato C, Pulver AE, Rahman N, Remes AM, Rioux JD, Ripatti S, Roden DM, Saleheen D, Salomaa V, Samani NJ, Scharf J, Schunkert H, Shoemaker MB, Sklar P, Soininen H, Sokol H, Spector T, Sullivan PF, Suvisaari J, Tai ES, Teo YY, Tiinamäija T, Tsuang M, Turner D, Tusie-Luna T, Vartiainen E, Ware JS, Watkins H, Weersma RK, Wessman M, Wilson JG, Xavier RJ, Neale BM, Daly MJ, Genome Aggregation Database Consortium. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;581:434–43.
- 30 Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 2013;14:178–92.
- 31 Church DM, Stotler CJ, Rutter JL, Murrell JR, Trofatter JA, Buckler AJ. Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. *Nat Genet* 1994;6:98–105.
- 32 Wang X, Zhao X, Wang X, Yao J, Zhang F, Lang Y, Tuffery-Giraud S, Bottillo I, Shao L. Two novel HOGA1 splicing mutations identified in a Chinese patient with primary hyperoxaluria type 3. *Am J Nephrol* 2015;42:78–84.
- 33 Gibbs DD. Werner's Syndrome ('Progeria of the Adult'). *Proc R Soc Med* 1967;60:135–6.
- 34 Gilpatrick T, Lee I, Graham JE, Raimondeau E, Bowen R, Heron A, Downs B, Sukumar S, Sedlazeck FJ, Timp W. Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat Biotechnol* 2020:1–6.
- 35 Walsh T, Casadei S, Munson KM, Eng M, Mandell JB, Gulsuner S, King M-C. CRISPR-Cas9/long-read sequencing approach to identify cryptic mutations in *BRCA1* and other tumour suppressor genes. *J Med Genet* 2021;58:850–852.
- 36 Yamaguchi K, Kasajima R, Takane K, Hatakeyama S, Shimizu E, Yamaguchi R, Katayama K, Arai M, Ishioka C, Iwama T, Kaneko S, Matsubara N, Moriya Y, Nomizu T, Sugano K, Tamura K, Tomita N, Yoshida T, Sugihara K, Nakamura Y, Miyano S, Imoto S, Furukawa Y, Ikenoue T. Application of targeted nanopore sequencing for the screening and determination of structural variants in patients with Lynch syndrome. *J Hum Genet* 2021;66:1053–60.
- 37 Logsdon GA, Vollger MR, Eichler EE. Long-Read human genome sequencing and its applications. *Nat Rev Genet* 2020:1–18.
- 38 Seaby EG, Pengelly RJ, Ennis S. Exome sequencing explained: a practical guide to its clinical application. *Brief Funct Genomics* 2016;15:374–84.
- 39 Harel T, Lupski JR. Genomic disorders 20 years on—mechanisms for clinical manifestations. *Clin Genet* 2018;93:439–49.
- 40 Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE. Recent segmental duplications in the human genome. *Science* 2002;297:1003–7.

SUPPLEMENTARY MATERIALS



Supplemental Figure 1. A splice variant results in a cryptic exon within *WRN* in the BIA pedigree. RT-PCR of BIA1010 and BIA1020, both of whom are heterozygous for c.3234-170A>G in intron 26 (as shown by the asterisk (*) on top diagram), showed larger products using the primer design from exon 25 (A: 5'-ATCGGCAAGGATCAAACAGAGAG-3') and exons 28-29 (P:5'-TTCTGGTACTGTACCATGATAC-3') (left panel). Using a primer designed from exon 26 and cryptic splice acceptor site (BIA-F 5'-GCTTCTGCCTAGACATCTTAAAGC-3'), two unique RT-PCR products were detected in both affected patients (right panel).



Supplemental Figure 3. IGV view of long-read whole-genome sequencing of rtPCR

products from SILV1010. Exons 19–21 are shown at the bottom and reads are represented by gray boxes. Linked reads are shown by thin lines connecting the gray boxes over the exons. Note that coverage is approximately half for exon 20 when compared to exons 19 and 21. Exon skipping of 20 was known in this sample.

| Sample | Estimated coverage of target region |
|---------------|--|
| WV | 28.3 |
| BIA1010 | n/a |
| BIA1020 | 25.6 |
| PD1010 | 18.4 |
| CB4 | 13.3 |
| CB6 | 20.2 |
| FES | 15.4 |
| SILV1010 | 22.9 |
| EN1010 | 14.3 |

Supplemental Table 1. Estimated coverage of each sample.

| Gene or Region | Chromosome | Start | End |
|-----------------------|-------------------|--------------|-------------|
| FMR1 | X | 147,400,000 | 148,500,000 |
| WRN | 8 | 29,900,000 | 33,700,000 |
| COL1A1 | 17 | 49,650,000 | 50,750,000 |

Supplemental Table 2. Regions targeted for adaptive sampling on the Oxford Nanopore Technologies (ONT) platform using GRCh38 coordinates. Target region is WRN, while control regions are FMR1 and COL1A1.

| Registry | Variant | CADD score | SpliceAI prediction | SpliceAI score |
|------------------|--------------------|------------|---------------------|----------------|
| BIA1010, BIA1020 | c.3234-170A>G | 9.556 | Acceptor site gain | 0.32 |
| CB4, CB6 | c.1982-297A>G | 11.89 | Donor site gain | 0.84 |
| EN1010 | c.839+1309T>G | 9.358 | Donor site gain | 0.73 |
| WV | c.361C>T, p.R1321* | 44 | NA | NA |

Supplemental Table 3. CADD and SpliceAI scores for single nucleotide variants identified in this study. NA, not applicable.

Supplemental File 1. Sequences of the intronic variants and activated splice sites found in this study.

Intronic variants (red), activated splice sites (blue) and region of cryptic insertion (gray) are shown with respect to WRN exons (yellow). See text for details.

c. 839+1309T>G (Registry#EN1010)

```

42841 ttaatgaagt ggctaaatga atatctctgc tttgtggttt gaaaattaat attgattttt
42901 tttcccccta gaggaagaaa tcctacttag cgacatgaac aaacagttga cttcaatctc
E8 42961 tgaggaagtg atggatctgg ctaagcatct tcctcatgct ttcagtaaaa tggaaaaccc
43021 acggaggtta aatattacct tttttttttt taacttaaat caattctggt tattttttta
43081 tcacattttc ctatatgtga agaatactat tcattaaggc tgactcatag aaatttgctt
43141 gaatacacac acacagactg atgcataaat ttgagaaata tctttccagt tgttgaatag
43201 tgaatccaaa ggtttaagcc ttttctgtct gctttgtagg atgttttatg aaacaaagga
43261 atagaaattg tattcatatt atacattttg aaattatata tatattatat tttttgtcaa
43321 agcaaaataa attatacatt aagggaaaata ttttaagattc aaagtgtgtca ttttctttct
43381 gttaatgagt atctttgatt taataacatt tatagtttg tcaactgaaa tctatagtga
43441 ccaaagcaa atgcatatatt agtagcctca tttctgttac tatctgatta ttacttaaga
43501 atttacttta attaatagat tattagaaca aaaggcattc caattttaga gattctttta
43561 tggttaaagc actaggtaac ttaaaaaaatt tggatcaaat gttatacaga ttaatagacc
43621 ttggaataca ggggctatgt ctttttgtaa tcagaccttg tcctggcctt aattgcagga
43681 tgacttctct attaaatctc taaaagcagg aatgctctgg aacaatgaac aattagaatg
43741 tggtaatgta attttgggcc aaaaaaactt ttgtagaaat gtgattctct tacttggaa
43801 agaaagaatt ctgactcgaa aaagtgcact taatgtatgc ttttctgggt aagattttgt
43861 atagattttg attgagattt aaaaccatga tatataaatt tagctttttt gtcaaagtaa
43921 tttgccaccc tagaatttgt gttggtaatt tttaatgcaa cttggacaat ttctgaaaga
43981 tgtaactttc ttttttttga gacgagtctc gctttgttgc ccaggctgga gtgcagtgac
44041 atgatctcag gtcaactgaa gctccgcctc ccgggttcac accattatcc tgcctcagcc
44101 tcccagtagt ctgggactac cggcgcccgc caccacacct ggctaatttt ttttttattt
44161 ttagtagaga cggggtttca ccgtgtttag caggatcgtc tcgatctccg gacctcgtga
44221 tctgcccgcc tcagcctccc aaagtgcctgg gattacaggc gttagccacc ggcgccagcc
44281 tgaagatga actttaaaat ggggaaggat actagtaaaag ccatttctta ctcatgtata
44341 catttttccc tccatgaaat tagagcagcc accttttaca tgacttttac atgcatttca
44401 gatttttgta ctagcaatta aaatatctac tggatgtaaa tcacatgtgc ttaatgcttt
44461 gttcataaat atctgttgaa gtcaattaaa atatagagac agatctttta attgaacatg
44521 ttttatttgg gaagcaagaa ttaaatttga ggcatataca cagactgggt gtccttggtta
44581 tctctgaaga acaaagagaa gtttggagggt tatagaaaaa ggaaaatgtg acattgtttt
44641 tccagaaagt tctttggcac tagtaaatth tggggagttg gaagctctga tgggttagtg
44701 actgtggttg gtaaaaactag tcttagagtc acaacagggt gtttcagtat ctattagata
44761 aacaggttt taggttacia taggcagttt cagcagctag gcttgcagaa aattacattc
44821 ttagagtaat gtatgtgcc ttgagtgctt tttcccctt gcccttgac tctgatttag
44881 ttgggtatga caagaattat caatttatat aattactttt tacatattga agctgaatag
44941 gttttcttaa aacttgaaa tgggaaatac tgggtcttgt gaagttttag tgtgaaaacg
45001 ataactagtc tcttatttca tgctctttgt aacaagtcta cagtaatacc taatgtatct
45061 agatgacacc tcagaaccag gagggaaacaa aaaagacctc tgagcagtc aataatgtct
45121 ataagcatgt attccagctt actttcggca tttctgtgga gtgcaggagg ccacattttg
45181 tcatagatc cagtttgaca agcttgggta ctgagttaaa gtgaaaatgg tgtaaataag
45241 taatcatatt cacatgacag taggtggaag tgacagctga caatctgaaa gagtggacaa
45301 aactataaaa attagaacta agaactctta aaatagaacg gtttgaagt tgttacagaa
45361 gggagcagct ttatgcttca gtggactctg aaggtattgg cagtgtatag tactgcttag
45421 tcattgattt aaattcggta tatatctctt ctattgctgt catgtacttg acattctgct
45481 ggggtgctaag gctgtctgca ggctactaat atgcataata acaatcctga aaaagaagga
45541 aaagtatact gtgccttact gaagaaggca ggacagccgt tggatattgt tagagtaaat
45601 attagttatc ttacataaag catgttatat tcctgatgt cagggcctga taaatgagcc
45661 cttacactct tgatagtcaa tctgtgccac tgttgctgtg tttgtatggt acaccatttc
45721 ctgtccatgt aacagttaaa ctgtatcaca tatctgttcc tactaaacta tttgcatata
45781 attttaacat ttgagtggtc gtatataaag ctaaaagttc tttataagat ttatagcatc
45841 ttagaatggt taaattagat aacattagtg tgcttttagt ttttgaattt aattttgctc
45901 ttgattttgt ttgtgttaat gtagactttg tctacgtaag cttgagccca actttggagt
45961 ggtgttttca gtgatggaag taactggaaa taactcatga aaccacttag tacttttgat
46021 agtttattcc catgaaggat caaattttta cttgagtcac taatataaaa atttaaaaga
46081 catgtcagat tttattaagc atggatttta aatatataaa tattagaaga cctaaacaaa
46141 atcgctaatg aaagagtaaa acatgtccct gcctatcata ctcccttcac tttcactgat
46201 ttctgctttt atgaagtact ttaataagct ctacaacaaa taacaagac attgtgcctc
46261 attttacact ctagcagtggt ttttccattc caactaacta gattgggttt ataatatag

```

46321 catttaaact tacttagtta aaatgttata cattagcaac tgcaattaag gagctaagaa
 46381 attcagctcg tatttattta tatgcatctt atccatagtt tctcacttgt tatattttgcc
 46441 tactgatggc aaattgattt tttctgttcc aaaaatgtag tgttttgaga tttttggaga
 46501 tgcaaaacag catcattttc aaattagatc acaaaacaatc ttacttgttg cagttgagac
 46561 tagtaattga tcagtgaggt gaaaaaatta ggcaaaacgg tatcatcaaa attttctttc
 46621 tttttttttt tgaacagag tattgctctt gtcgccagg ctggagtgca atggcacgat
 46681 ctacgctcag ctactgggg ccctaccta cgggattcaa gggattctcc tgcctcagcc
 46741 tcctgagtag ctgagattac aggcgatgc caccaagccc agctaatttt tgtattttta
 46801 gtagagacag ggttttgcca tgttggccag gctggctctg aacttcagac ctcaagtgat
 46861 tcacctgcct cagcctccca aagtgtctgg attacaggca tgaaccaccg tgcccgcccc
 46921 aaattttctt ttaacttaaa gtgttaatgt ttgtttgttc ccaatatttt tatatatgtc
 46981 aaggcttttt cattatatac aggcagctc tttggagttc tgcatagtgt ttcttgttta
 47041 aaccacacct ataatgtacc atctacaatt tccaattctt taagacagag tgaaaaatta
 47101 agacacctga agcaatctga gtgcggagtc ctaggcatta tgattttcct tgacactttt
 47161 agggttcttt tatcacagtt tatcaaacac cttttttttt ttaacaaaaa acaaatttac
 47221 aaaaactttc catgtattca acttagaatt taatgaaata gggttttttt ttttttggtg
 47281 gtcatgtatc atcaactttc tttacattta attaaattaa tcaacaaaaa acatatagtt
 47341 gacataaaca aatttgggac caagcactac tgactcttag taaacataca agtcaactgt
 47401 tggtcagaa gtgccttgat gctagagagt agcctgcaag ctgtaattat atatggcat
 47461 tagggaaatga agaacagcct taataactatt gatcttttaa gtgaaggcca gcttagaag
 47521 cttttacttg ttaaaaagct tcacagtttg tccttgtagt taatgcaatt gaagttgaat
 47581 taacttttct taattttttt ttttagggttt ctatcttact aaaggatat tcagaaaaatc
 47641 tatattcact gaggaggatg ataattgggt ctactaacat tgagactgaa ctgaggccca
 47701 gcaataattt aactttatta tcctttgaag attcaactac tgggggagta caacagaaac
 47761 aaattagaga acatgaagtt ttaattcacg ttgaagatga aacatgggac ccaacacttg
E9 47821 atcatttagc taaacatgat ggagaagatg tacttggaaa taaagtggaa cgaaaagaag
 47881 atggatttga agatggagta gaagacaaca aattgaaaga gaatatggaa agagcttgtt
 47941 tgatgtogtt agatattaca gaacatgaac tccaaaatttt ggaacagcag tctcaggaag
 48001 aatatcttag tgatattgct tataaatcta ctgaggtact aaataaagag gaagcacatt
 48061 tttagttatt agtaggttct ggcagacttt attcccgtaa agagacagat agtaaatttt

c.1982-297A>G (Registry#CB4 and CB6)

63421 tccttgagat gattgttttc tttattgtta atatgtttcc ctctctgttt tttttttttt
 63481 ctcttttctt ttgtttgttt ttacagaggt aaataaccgga ttgtatacgt aactccagaa
E17 63541 tactgttcag gtaacatggg cctgctccag caacttgagg ctgatattgg taagtataaa
 63601 agaagatct ctgtaaatatc ttactgagtt aatatttaaa gttaaacctc tggatggaca
 63661 ctggatttca ctctgttaa agtttatttc aaacattact tcctccagga aatctctgac
 63721 tctctaactt gctgtttcac ctatcttacc ccatgatact ctataattcc ttttgctata
 63781 gaaattagtc tctgttatca taattacctg tttatataatc tgttttcttc tgctaggcta
 63841 taaacttttc aagggaaaaa tatcactttt taaaatttat gtttccccag tgcagtgttt
 63901 ggcataatag cacttaaaca tattgttgaa tgaatgttga atcctaggtt ctgtgaaatt
 63961 ttaatttact tgtttttcat gttctgtatc taaataaagt cagtgtagtg gttttgtacc
 64021 ctctttgttc ataagaatca cttttcaaac tttaaaaaaa atgtaggtaa ccaacacca
 64081 actctgaaaa ttctgattta acaggcctgg gttggcacct ggcattgggt gtttctcaca
 64141 ggtgattctg atatatagcc cgggttgaa accacagtac ctttgattag ttgccactag
 64201 tgagtatgat ttattacaaa aaccaaaaaac ttagataagt cacagtttat ttctagaac
 64261 gtgtaaatat agaattagtc aaaagctgat agcacttttt ttgtacttaa aaatatttta
 64321 agcctcagta tagtgagacc tcatctctac aaataatttt ttaaaaaatt ggcgagcct
 64381 ggtggtgcat gcccatagtc caagctactt gggaggctga ggtaggagga ttgtttgagc
 64441 ctgggagatg aaggctgagg tgagccaaga tcatgctact gcaactccatc ctgggcaaca
 64501 gagcgagacc atgcctcaaa aaaaattttt ttttcttctt ggtggtttag cttaactatc
 64561 caagaattgc aatcatggct tctattgcaa atgcccctat ttcccacctt cagttggcct
 64621 aaccaaaatt aatttttcaa tatcatatga tctttttgta taattatgta ttcaaataat
 64681 tccagatagg ttaggttttc gaattgcaac gtctcattta atggttactt tcataact
 64741 tagggaagc atttgataat ttttocaaca atgtaatgaa acaaaacttg tatctttttt
 64801 attatgtgtt aaaaccaaca aaagaatgag ggcagcaatt aacatttcat taaatatatt
 64861 agcaatgagt acatttatga gctcttaaac tcccttattg acatactgta ttttttgctt
 64921 ttaatggaca atgaatttga tatgtttcct aataggttgt atgaagttat tgtgaatatt
 64981 tcacttttga ttaagaataa ctaccctata tataaatttc atggttgtat tgcaaaagct
 65041 ggggtctatg tccatagtgg tttgggtacg tttttactat ttgctttgtt tctgtgatag
 65101 tgtaaatga ataataagct taagtattta agcaaaact taatacttct gacaagtagt
 65161 tagaactctc tatttttggt tttttgtaat ggtatgagtt tggacactgc aattattaga

```

65221 gcagcgtact ttgcttatga tggatgaaaa tgttctctgg ttgactcagt gttgaataaa
65281 aagggataca tgaataaaaa attatTTTTTg aaatattttc ctttgccttt tgttgcggga
65341 aattgtacaa tgcttgcaaa aatgaatata cgcaggatat tatttgtaca gcagtcctgac
65401 atttagctaa tcagattcct catTTTTtagg taaatgttaa attacttaag tagtctggat
65461 ttttaacaga aacagcaaac atagTTTTat agggaggctc cttaccaacc agacttttaa
65521 cttaagtgaa tgttggcaaa agcttTctta aactgatttc cagggataag gtttagctcc
65581 attaaaagct agtttgtaa acctgtaact aaagctTTTT aaaaaatgaa agttttaaaa
65641 gagaccatca ttctgtttta acctatttaa tttaaaatct gggctgaagt aggagtattg
65701 cttgaacca ggagttcgag accagcctgg gcaacacagt gagaccccat ctctgaaaaa
65761 aaaaaatgag ctgagcatgg tacctgttac agtgctagtt atgaaagaaa gaaagagaga
65821 gagagagaag gaaggaaggg agaaaggaag gaaggggctg ggagtgggtg ctcacacctg
65881 taatcccagc actttgggag gccgagggcg gtggatcatg aggtcaggag atcagacca
65941 cctggcgaa catggtgaaa cctgtctct actaaaaata caaaaagaaa aaattagccg
66001 ggcatgggtg tgggtgcctg tagtcccagc tactggggag ggtaaggcag gagaatggca
66061 tgaaccoggg aggcagagct tgcagtgagc cgagattgag ccactgcact ttagcctaag
66121 aaaaggaag aaggaaggaa ggaaggaag aaggaaaaat cagtcactca ctcaggTTta
66181 gtttcaaaag ccaaaacagt gataatattg aattTTTTcc agccactcag tagtaccag
66241 agaattttga tacaaataac tgactttgac cccatgttta gtaggttttt tttttgtttg
66301 tttgtttgtt tttttttttt gtctttcttc tccacactct cttaggactg ttgtagctg
66361 taatgttagt gaattgtaat tgtttatttg catgtctgtc tcctctTTta gaattgtgaa
66421 ctctcaggt catggattgt attttattag tctttttatt cctagaaaag tgtttggata
66481 ataatagata tttattaacc ataataggag tagtagcatt tttgtgtgtg tgtgccaggc
66541 cctatttttag cattttatgt gaatgatata atttaatttt tagagcatct agaaaatgtt
66601 aggtaccatt attatccaca ttttacttgt gagaggactg agacttgtaa agcttaagta
66661 acttgcttaa ggtaatacac atggcagatg gtggagccag gttctgcatc tgggcagctc
66721 gacttttagag ctttttaatc atcctcctcc tagtgtccga acattattaa tgcctagaga
66781 agtatcaaag aatgatgata ctcatggagg gaaaaaatta ttcttaagta tcaatttTtt
66841 aattgaggac aaaagcacta ctgttactac tgtgaaaaat ctgaaaacat tgtcacggat
66901 gattttacaa taacatcctt ttctgacatg cattacaaca cactattgat tccgtttcaa
66961 tgtttTgtgtc tggatgggg gcggagtaga gaggctacta caggaagaaa gaacaaataa
67021 catttcttct tactactttg tgcacatgag gatthagctc ctcccatacc attttacatt
67081 agaaaatgtg gctgtctcc tgtgtctag tactggcatt tcagaggact ctttgcataa
67141 tgatagtga tcaatcatgg atggcagatg agttaaatttc agtagaaatt tgcctcactc
67201 tttatctct tttaatagct aatgcatgag ataaaattga aggtattacc aaggagaaat
67261 gcaggtggat agttttggag gcaaagtcac agtTaaagcaat tcatctctcc aaggaatagt
67321 gatgaaagct gagatccaat aaatgtcttg tttggattaa tacataaata tttcctTTta
67381 aaattcttaa ggcttttatt catatatctc ccctctccct atttccaata atgatttttg
67441 gtgttaatat cgtgatattt atgcttctgt ttttatagga accaaatact ttgattttgtg
67501 tatttatctt ccattggaga tgtagatgag tttattTTTT cctttcagagc tttatctttt
67561 cctttatgtg ttttctttt tttacaggtta tcacgctcat tgctgtggat gaggctcact
E18 67621 gtattttctga gtgggggcat gattttaggg attcattcag gaagtTgggc tccctaaaga
67681 cagcactgcc aatgTtaagc tttgccaagt ctgatgtccc gaaattacat tcttaataag
67741 gagagcattc aggattgggg agtggtaaag aagctgaaga cttcactata aaagagcaaa

```

c.3234-170A>G (Regitsry#BIA1010, BIA1020

```

108361 ctgtttgact taattttgtt tcccactcca cattaaaaga tcctTTTTgc ttttaatatg
108421 gtagaaattg gcttcataaa gctaatacag aatctcagag cctcatcctt caagctaattg
E26108481 aagaattgtg tccaaagaag ttgcttctgc ctaggttcat ttttcagttt ttttcttgta
108541 acttctgcat tttttgttgc tatttatgtg attcaaatta taccagtTta taggcctctc
108601 acaagtaaaa tgaattacct gtttgttttt gtatgcctat ttttagtcagt ttgggggaag
108661 ggatctgtga gaaaggata agtcatagag cacttttctt ttttaagaga cagagtctct
108721 ctgtgttgct caagctggag tgcagtgggt cgatcatagc ttactgcagc ctcgatctcg
108781 tgggcccaag taatcctcag ccacctgagt agatgggact acagacatgc actactatgc
108841 ccagctaata tattttaatt ttttgatata agacagggct tctagtgtct tctaggctg
108901 cctctgaact cctgagctca agtgcctcct ctgectcagc ctoccaaact actgggatta
108961 caggcatgat ccaccgctcc cagccagaac attttcttgg ttgatgggaa gttagctgacc
109021 atggtattta gaaaacttct ttctcatcga ttaaagaagc agtactgaaa tcaatgcgga
109081 ggaatccata tatcatattt acttctgggt tgtagaagtg gaaagggaaT acattttgtg
109141 cttacttttt tgtaccttta catgtgattg atcacttTgt agttttttct ttcaaacatc
109201 ttaaagcttc cagagctttt tctagaaaaa aaaatcagtt ttaagaatca ccagttctaa
109261 aaggTtaata tcttattcat ctttctgaga atggagtatc atgattcatg aattagatc
109321 ttgcatctta acatttgaaa taatttaatt ttattatttt ttagttcgaa aactgtatct

```

E27109381 tcgggcacca aagagcattg ttataatcaa gtaccagttg aattaagtac agagaagaag
109441 gtttgtttta aagaaattgt tctgatttat ttcattcttt attgattcaa attctgttta