HHS Public Access

Author manuscript

Hum Mutat. Author manuscript; available in PMC 2023 December 01.

Published in final edited form as:

Hum Mutat. 2022 December ; 43(12): 1856–1859. doi:10.1002/humu.24469.

Next generation sequencing errors due to genetic variation in *WRAP53* encoding TCAB1 on chromosome 17

Sharon A. Savage[1], **Kristine Jones**[2,3], **Kedest Teshome**[2,3], **Adriana Lori**[4], **Lisa J. McReynolds**[1], **Marena R. Niewisch**[1]

[1]Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA

[2]Cancer Genomics Research Laboratory, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA

[3]Leidos Biomedical Research, Inc. Frederick National Laboratory for Cancer Research, Frederick, MD, USA

[4]American Cancer Society, Atlanta, GA, USA

Abstract

Next generation sequencing (NGS) is a valuable tool, but has limitations in sequencing through repetitive runs of single nucleotides (homopolymers). Pathogenic germline variants in *WRAP53* encoding telomere cajal body protein 1 (TCAB1) are a known cause of dyskeratosis congenita. We identified a significant NGS error in *WRAP53*, c.1562dup, p.Ala522Glyfs*8 (rs755116516 G>-/GG/GGG) that did not validate by Sanger sequencing. This error occurs because rs755116516 G>-/GG/GGG (Chr17:7,606,714) is polymorphic and variants at this site challenge the ability of NGS to accurately call the correct number of nucleotides in a homopolymer run. This was further complicated by the fact that chr17:7,606,721 (rs769202794) is multiallelic G>A, C, T, and that chr17:7,606,722 is immediately adjacent and also multi-allelic (rs7640C>A/G/T and rs373064567C>delC). In addition to expert interpretation of potentially clinically actionable variants, it recommended that all variants in regions of the genome with homopolymers be validated by Sanger sequencing prior to clinical action.

Keywords

next generation sequencing; *WRAP53* ; sequencing errors

BRIEF REPORT

WRAP53 encodes telomere Cajal body protein 1 (TCAB1), a protein that interacts with telomerase, telomerase RNA component, and small Cajal body RNAs. Biallelic pathogenic germline variants in *WRAP53* are a known cause of with dyskeratosis congenita (DC),

Corresponding author: Sharon A. Savage, M.D., Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 9609 Medical Center Drive, Room 6E456, Bethesda, MD 20892-6772, Tel: 240-276-7252, savagesh@mail.nih.gov.

Conflict of interest statement: The authors declare no conflicts of interest.

a telomere biology disorder (TBD) associated with high rates of bone marrow failure, pulmonary fibrosis, cancer, and other medical problems (Bergstrand et al., 2020; Niewisch et al., 2021; Zhong et al., 2011). Heterozygous variants in *WRAP53* have not been reported in association with human disease. Given the rarity of *WRAP53*-associated disease, validation of variants and careful curation are essential prior to assigning clinical relevance (Richards et al., 2008).

Next generation exome sequencing (NGS) was performed on 4,688 individuals as a part of several previously reported familial and population-based exome sequencing studies (McReynolds et al., 2022; Mirabello et al., 2017; Shi et al., 2014). Briefly, exon-enriched libraries generated with NimbleGen v3 or v3+UTR capture kits were sequenced with Illumina MiSeq or HiSeq to an average depth of ~55x and minimum coverage of >80% at 15x. Reads were aligned to the hg19 reference genome (Novoalign, Picard). GATK Version 3.8 UnifiedGenotyper, HaplotypeCaller, and Freebayes were used to call variants with variants only used if called by at least two of three callers. Variant calling was limited to the intersection of the v3 and v3+UTR capture kits.

We identified 54 unrelated individuals with a *WRAP53* c.1562dup, p.Ala522Glyfs*8 (rs755116516 G>-/GG/GGG) variant by NGS. This finding was not associated with any specific cohort. This region was further assessed by Sanger sequencing 29 of these individuals and found that the NGS sequencing calls were errors (Figure 1). Notably, 27 individuals had the GG SNP at rs7640, two had one C and G at rs7640. Four individuals had dupG rs755116516 in addition to G at rs7640, but 25 showed no evidence of the rs755116516 indel in the Sanger sequencing data. This erroneous call in NGS data occurs because when added Gs at rs755116516 or rs7640 occur, a longer run of Gs is created, leading to errors in base calling due to the known challenges of accurately calling homopolymer runs on next generation sequencing platforms (Foox et al., 2021; Mu, Lu, Chen, Li, & Elliott, 2016).

Review of publicly available databases using human genome build GRCh37 show that this region of *WRAP53* is prone to errors in next generation sequencing due to the number of G and C nucleotides (Figure 1). rs755116516 G>-/GG/GGG at Chr17:7,606,714 is polymorphic and variants at this site will challenge the ability of NGS platforms to accurately call the correct number of nucleotides in a homopolymer run. This is further complicated by the fact that chr17:7,606,721 (rs769202794) is multiallelic G>A, C, T, and that chr17:7,606,722 is immediately adjacent and also multi-allelic (rs7640C>A/G/T and rs373064567C>delC).

It is important for clinicians and researchers to understand that NGS is a valuable diagnostic tool but has known limitations in generating repetitive runs of single nucleotides (homopolymers), throughout the genome (Foox et al., 2021; Mu et al., 2016; Samorodnitsky et al., 2015; Slatko, Gardner, & Ausubel, 2018). In addition to expert interpretation of potentially clinically actionable variants, it recommended that all variants in regions of the genome with homopolymers be validated by Sanger sequencing prior to clinical action.

Acknowledgements

This work was supported by the intramural research program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health. The Cancer Genomics Research Laboratory is funded with federal funds from the NCI, NIH, under NCI Contract 75N910D00024. This work does not necessarily represent the views of the National Institutes of Health. Blood-derived DNA samples used in this study are from the Transplant Outcomes in Aplastic Anemia study, an NCI collaboration with the Center for International Blood and Marrow Transplant Research (CIBMTR), and from the American Cancer Society's Cancer Prevention Study-II (CPS-II) and the NCI Prostate, Lung, Colorectal and Ovarian Cancer study (PCLO) (Calle et al., 2002; "PLCO: National Cancer Institute, Cancer Data Access System,"). CIBMTR is supported primarily by Public Health Service U24CA076518 from the NCI, the National Heart, Lung and Blood Institute and the National Institute of Allergy and Infectious Diseases; HHSH250201700006C from the Health Resources and Services Administration; and N00014-21-1-2954 and N00014-20-1-2832 from the Office of Naval Research. The American Cancer Society funds the creation, maintenance, and updating of the CPS-II cohort. The NCI PLCO study is supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics and by contracts from the Division of Cancer Prevention, NCI, NIH.

The authors thank Dr. Shahinaz Gadalla and other members of the Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute for valuable discussions related to this manuscript.

Data availability:

De-identified sequencing data will be available through the dbGAP-controlled access database accession number dbGaP: phs001710.v1.p1 and/or via direct request to Dr. Sharon Savage after establishment of appropriate data transfer agreements.

References

Bergstrand S, Bohm S, Malmgren H, Norberg A, Sundin M, Nordgren A, & Farnebo M (2020). Biallelic mutations in WRAP53 result in dysfunctional telomeres, Cajal bodies and DNA repair, thereby causing Hoyeraal-Hreidarsson syndrome. Cell Death Dis, 11(4), 238. doi:10.1038/s41419-020-2421-4 [PubMed: 32303682]

Calle EE, Rodriguez C, Jacobs EJ, Almon ML, Chao A, McCullough ML, … Thun MJ (2002). The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. Cancer, 94(9), 2490–2501. doi:10.1002/cncr.101970 [PubMed: 12015775]

Foox J, Tighe SW, Nicolet CM, Zook JM, Byrska-Bishop M, Clarke WE, … Mason CE (2021). Performance assessment of DNA sequencing platforms in the ABRF Next-Generation Sequencing Study. Nat Biotechnol, 39(9), 1129–1140. doi:10.1038/s41587-021-01049-5 [PubMed: 34504351]

McReynolds LJ, Rafati M, Wang Y, Ballew BJ, Kim J, Williams VV, … Gadalla SM (2022). Genetic testing in severe aplastic anemia is required for optimal hematopoietic cell transplant outcomes. Blood. doi:10.1182/blood.2022016508

Mirabello L, Khincha PP, Ellis SR, Giri N, Brodie S, Chandrasekharappa SC, … Savage SA (2017). Novel and known ribosomal causes of Diamond-Blackfan anaemia identified through comprehensive genomic characterisation. J Med Genet, 54(6), 417–425. doi:10.1136/jmedgenet-2016-104346 [PubMed: 28280134]

Mu W, Lu HM, Chen J, Li S, & Elliott AM (2016). Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing. J Mol Diagn, 18(6), 923–932. doi:10.1016/j.jmoldx.2016.07.006 [PubMed: 27720647]

Niewisch MR, Giri N, McReynolds LJ, Alsaggaf R, Bhala S, Alter BP, & Savage SA (2021). Disease Progression and Clinical Outcomes in Telomere Biology Disorders. Blood. doi:10.1182/blood.2021013523

PLCO: National Cancer Institute, Cancer Data Access System. Retrieved from https://cdas.cancer.gov/plco/

Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, … Molecular Subcommittee of the, A. L. Q. A. C. (2008). ACMG recommendations for standards for interpretation and

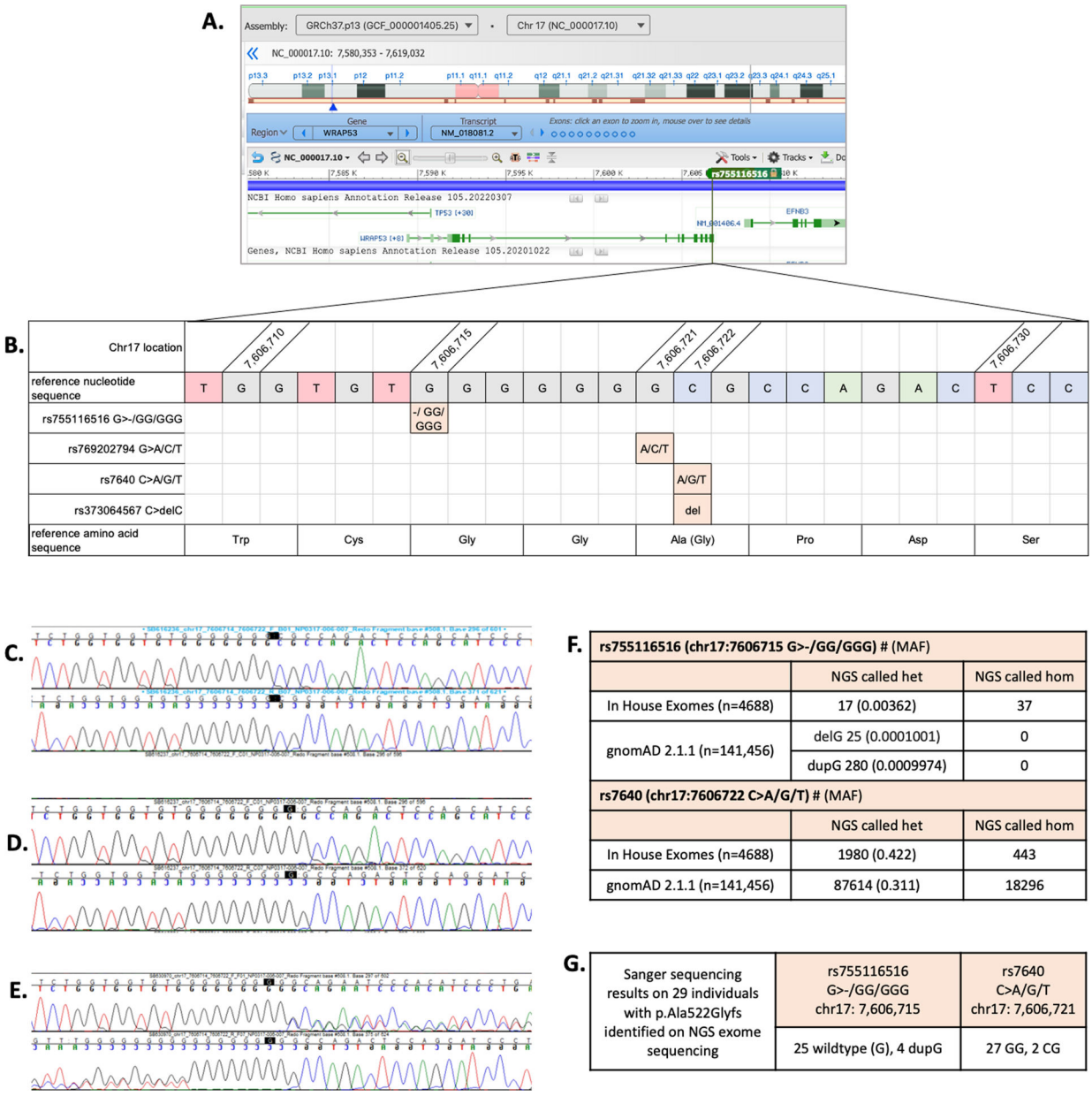
reporting of sequence variations: Revisions 2007. Genet Med, 10(4), 294–300. doi:10.1097/GIM.0b013e31816b5cae [PubMed: 18414213]

Samorodnitsky E, Jewell BM, Hagopian R, Miya J, Wing MR, Lyon E, … Roychowdhury S (2015). Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing. Hum Mutat, 36(9), 903–914. doi:10.1002/humu.22825 [PubMed: 26110913]

Shi J, Yang XR, Ballew B, Rotunno M, Calista D, Fargnoli MC, … Grp, F. F. M. S. (2014). Rare missense variants in POT1 predispose to familial cutaneous malignant melanoma. Nature Genetics, 46(5), 482–486. doi:10.1038/ng.2941 [PubMed: 24686846]

Slatko BE, Gardner AF, & Ausubel FM (2018). Overview of Next-Generation Sequencing Technologies. Curr Protoc Mol Biol, 122(1), e59. doi:10.1002/cpmb.59 [PubMed: 29851291]

Zhong F, Savage SA, Shkreli M, Giri N, Jessop L, Myers T, … Artandi SE (2011). Disruption of telomerase trafficking by TCAB1 mutation causes dyskeratosis congenita. Genes Dev, 25(1), 11–16. doi:10.1101/gad.2006411 [PubMed: 21205863]



F.

rs755116516 (chr17:7606715 G>-/GG/GGG) # (MAF)		
	NGS called het	NGS called hom
In House Exomes (n=4688)	17 (0.00362)	37
gnomAD 2.1.1 (n=141,456)	delG 25 (0.0001001)	0
	dupG 280 (0.0009974)	0
rs7640 (chr17:7606722 C>A/G/T) # (MAF)		
	NGS called het	NGS called hom
In House Exomes (n=4688)	1980 (0.422)	443
gnomAD 2.1.1 (n=141,456)	87614 (0.311)	18296

G.

Sanger sequencing results on 29 individuals with p.Ala522Glyfs identified on NGS exome sequencing	rs755116516 G>-/GG/GGG chr17: 7,606,715	rs7640 C>A/G/T chr17: 7,606,721
	25 wildtype (G), 4 dupG	27 GG, 2 CG

Figure 1. Exon 10 of *WRAP53* on chromosome 17 includes multi-allelic variants, insertions, deletions, and homopolymer runs of guanines making next generation sequencing prone to errors.

A. Image from Variation Viewer showing the location of *WRAP53* on chromosome 17, https://www.ncbi.nlm.nih.gov/variation/view/?assm=GCF_000001405.39

B. Schematic of sequence based on human genome build GRCh37/hg19 illustrating the run of guanine nucleotides and locations of single nucleotide polymorphisms (SNPs) leading to next generation sequencing errors. rs7640 (chr17:7,606,722 C>G) often causes an indel G call at 7,606,722 or at 7,606,714 in next generation sequencing that is false due to the long run of Gs created by the SNPs. Of the 29 subjects with an exome sequencing indel call at chr17:7,606,715 (rs755116516, c.1562dup,), all had an rs7640 C>A/G/T alternate allele (27 homozygous G and 2 heterozygous G) by Sanger sequencing.

C. Reference Sanger sequencing trace with 7 guanines at chr17:7,606,715–7,606,721.
D. Example of Sanger sequencing trace of subject with rs7640 G allele
E. Example of Sanger sequencing trace seen in 4 subjects with an exome indel call at chr17:7,606,715, and both the C>G variant at chr17:7,606,722 AND a single base G insertion. The insertion could be between chr17:7,606,714/15, or between chr17:7,606,723/724 but this cannot be resolved as it is not possible to determine which end the G is inserted into.
F. Next generation exome sequencing results from gnomAD 2.2.1 and in house exome sequencing.
G. Sanger sequencing results of the same region.