

ORIGINAL ARTICLE

Deleterious mis-splicing of *STK11* caused by a novel single-nucleotide substitution in the 3' polypyrimidine tract of intron five

Thorkild Terkelsen¹  | Ole H. Larsen²  | Søren Vang²  | Uffe B. Jensen¹  |
Friedrik Wikman² 

¹Department of Clinical Genetics, Aarhus University Hospital, Aarhus, Denmark

²Department of Molecular Medicine, Aarhus University Hospital, Aarhus, Denmark

Correspondence

Thorkild Terkelsen, Department of Clinical Genetics, Aarhus University Hospital, Brendstrupgaardsvej 21C, DK-8200 Aarhus, Denmark.
Email: thorkild.terkelsen@clin.au.dk

Funding information

Aarhus Universitetshospital

Abstract

Background: Pathogenic variants in *STK11*, also designated as *LKB1*, cause Peutz–Jeghers syndrome, which is a rare autosomal dominant disorder characterized by mucocutaneous pigmentation changes, polyposis, and a high risk of cancer.

Methods: A male meeting the clinical diagnostic criteria for Peutz–Jeghers syndrome underwent next-generation sequencing. To validate the predicted splicing impact of a detected *STK11* variant, we performed RNA-Seq on mRNA extracted from patient-derived Epstein-Barr virus-transformed lymphocytes treated with cycloheximide to inhibit nonsense-mediated decay *ex vivo*.

Results: Blood testing identified a novel single-nucleotide substitution, NM_000455.4:c.735-10C>A, at the end of the 3' polypyrimidine tract of intron five in *STK11*. RNA-Seq confirmed a predicted eight base pair insertion in the mRNA transcript. Following inhibition of nonsense-mediated decay, the out-of-frame insertion was detected in 50% of all RNA-Seq reads. This confirmed a strong, deleterious splicing impact of the variant.

Conclusion: We characterized a novel likely pathogenic germline variant in intron five of *STK11* associated with Peutz–Jeghers syndrome. The study highlights RNA-Seq as a useful supplement in hereditary cancer predisposition testing.

KEYWORDS

germline mutation, *LKB1*, Peutz–Jeghers syndrome, RNA-Seq, splice variant, *STK11*

1 | INTRODUCTION

Peutz–Jeghers syndrome (OMIM #175200) is a rare autosomal dominant disorder characterized by mucocutaneous pigmentation, gastrointestinal hamartomatous polyposis, and a high risk of various neoplasms (Jeghers, Mc, & Katz, 1949;

van Lier et al., 2010). Germline pathogenic variants in the serine/threonine protein kinase 11 gene (*STK11*, also designated as *LKB1*; OMIM *602216) cause Peutz–Jeghers syndrome (Hemminki et al., 1998; Jenne et al., 1998). Traditional testing methods detect pathogenic variants in *STK11* in approximately 80% of individuals with Peutz–Jeghers syndrome

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Molecular Genetics & Genomic Medicine* published by Wiley Periodicals LLC.

(Volikos et al., 2006). Techniques focused on the transcriptome, such as whole RNA-sequencing (RNA-Seq), could add to the diagnostic yield by detecting deep intron variants (Abed, Gunther, Kraus, Hohenberger, & Ballhausen, 2001). We report an individual with Peutz–Jeghers syndrome due to a novel single-nucleotide substitution c.735-10C>A in the 3' polypyrimidine tract of intron five in *STK11*. The study highlights RNA-Seq as an efficient tool to assess potential splice variants.

2 | MATERIALS AND METHODS

2.1 | Ethical compliance

The patient gave informed consent to participate in the study. The study did not require approval from the ethics committee system in Denmark.

2.2 | Clinical data

A 60-year-old male underwent a screening colonoscopy in the biannual population screening program for colorectal cancer. An adenocarcinoma was identified in the sigmoid colon along with multiple polyps throughout the colon and rectum. Pathological assessment of a polyp revealed the classical Peutz–Jeghers appearance of thick smooth muscle bundles between nonneoplastic epithelial crypts. Multiple additional polyps in the upper and lower bowel demonstrated the same hamartomatous morphology, which confirmed the clinical diagnosis of Peutz–Jeghers syndrome. The patient had no family history of Peutz–Jeghers syndrome (Figure 1a).

2.3 | DNA analysis

DNA was extracted from a blood sample of the patient and analyzed by next-generation sequencing of a gene panel for hereditary polyposis and colorectal cancer (Table S1). All exons plus 20 base pairs (bp) of adjacent intronic sequence were analyzed. Copy number variants were identified by bioinformatic analysis of next-generation sequencing data using ExomeDepth (Plagnol et al., 2012) and Delly2 (Rausch et al., 2012). Multiplex ligation dependent probe amplification (MLPA) was also performed on the three genes: *MLH1*, *MSH2*, and *PMS2*. Splice effects were predicted in silico using SpliceSiteFinder-like (Zhang, 1998), MaxEntScan (Yeo & Burge, 2004), NNSPLICE (Reese, Eeckman, Kulp, & Haussler, 1997), and GeneSplicer (Pertea, Lin, & Salzberg, 2001). Population allele frequencies were extracted from the Genome Aggregation Database (gnomAD entry April 2020) (Karczewski et al., 2020) and variants were looked up in the Human Gene Mutation Database (Stenson et al., 2017) and ClinVar records (Landrum et al., 2018). The variants were classified as recommended by the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) (Richards et al., 2015). Only variants classified as C3-uncertain significance, C4-likely pathogenic, or C5-pathogenic were reported by the laboratory.

Sequence variant descriptions followed the guidelines of the Human Genome Variation Society (den Dunnen et al., 2016), and the nomenclature was checked using Mutalyzer (Wildeman, van Ophuizen, den Dunnen, & Taschner, 2008). The locus-specific database Leiden Open Variation Database (LOVD) was used for variant submission (Fokkema et al., 2011).

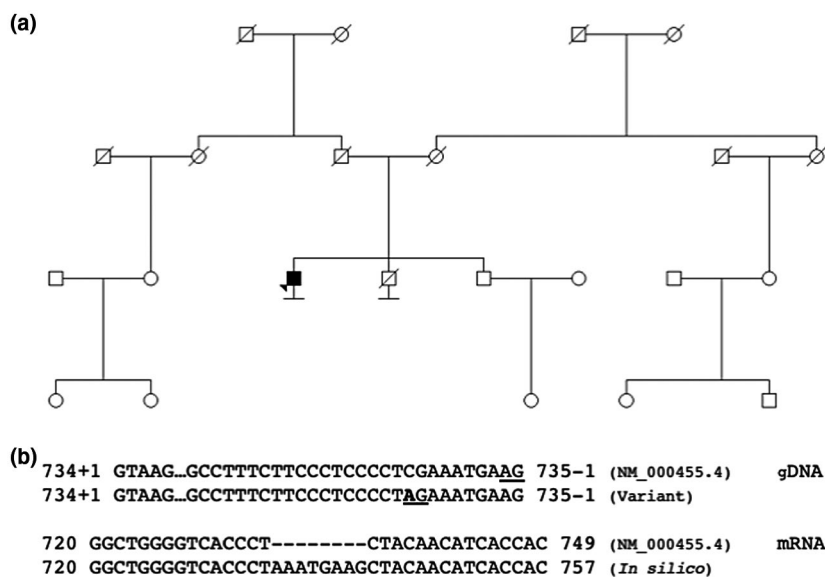


FIGURE 1 Summary of the genetic findings. (a) The pedigree of the male affected by Peutz–Jeghers syndrome and his unaffected family members. (b) The novel genomic variant in intron five of *STK11*, NG_007460.2(NM_000455.4):c.735-10C>A, created a strong potential splice acceptor site (upper alignment) and was predicted to result in an aberrant *STK11* transcript that included the last eight bp of the intron (lower alignment)

2.4 | Cell culture

Treatment of human cells with the translation elongation inhibitor cycloheximide (CHX) impairs the nonsense-mediated decay regulatory mechanism (Carter et al., 1995), which otherwise eliminates mRNA transcripts harboring premature termination codons (Khajavi, Inoue, & Lupski, 2006). An Epstein-Barr virus-transformed lymphocyte culture was established from a blood sample of the patient. Immortalized lymphocytes were grown in three 75 cm² culture flasks containing 19 ml culture medium added 1 ml CHX solution to final concentrations of 0, 10, or 40 µg/ml. Cells were incubated for 8 hr at 37°C, 5% CO₂ before RNA extraction using the RNeasy Mini Kit Cat. no. 74106 (QIAGEN Nordic).

2.5 | RNA-Seq

Paired de-multiplexed FASTQ files were generated using the software bcl2fastq (Illumina) and quality checked using FastQC (Andrews, 2010) and FastQ Screen (Wingett & Andrews, 2018). Adapter sequences were trimmed using TrimGalore (Krueger, 2012) and Cutadapt (Martin, 2011). Statistics were generated for the raw and trimmed FASTQ sequences for each sequence library within each sample. The trimmed reads were mapped to the human reference genome (hg19) with masked PAR1 and PAR2 regions on the Y chromosome to better estimate the expression of pseudoautosomal genes on the X chromosome. TopHat (Trapnell, Pachter, & Salzberg, 2009) was run in strand-aware mode using the “—library-type fr-secondstrand” option.

To find the reads containing the predicted novel transcript that would not align to the normal RNA sequence for *STK11* (NM_000455.4), a sequence was constructed from the 10 last bases of exon 5, the 8 last bases of intron 5 (the predicted inserted sequence), and the 20 first bases of exon 6. This sequence was used as search string for agrep (Wu & Manber, 1992), allowing a maximum of 2 misaligned bp, to search the unmapped reads for reads originating from the new transcript. The findings were confirmed using the hg38 genome build and an alternative mapper STAR v2.7.2b (Dobin et al., 2013).

The RNA-Seq reads were assumed to follow a binomial distribution between the two transcripts. Two-sided exact binomial confidence intervals and test statistics were reported. Statistical significance was considered at the 95% confidence level.

The statistical analysis was performed in Stata v12.1 (StataCorp).

3 | RESULTS

In the 60-year-old male meeting the clinical diagnostic criteria for Peutz–Jeghers syndrome, a germline variant was

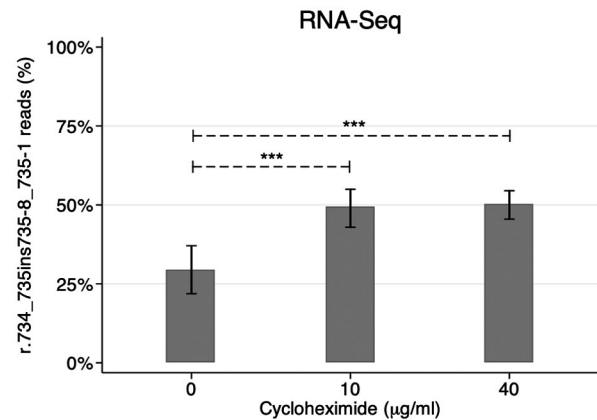


FIGURE 2 RNA-Seq validation of the splicing impact. The figure shows the proportion of *STK11* (NM_000455.4) RNA-Seq read counts (%) of the variant transcript with the error bars indicating the 95% confidence interval. A higher frequency of the variant transcript was observed when the nonsense-mediated decay mechanism was inhibited with cycloheximide ($p < .001$, exact binomial test)

identified in *STK11*, NG_007460.2(NM_000455.4):c.735-10C>A. The variant lead to a substitution of a cytosine with an adenine base at the end of the terminal polypyrimidine tract of intron five, which could potentially create a strong consensus splice acceptor site (Coolidge, Seely, & Patton, 1997). All four in silico tools predicted the variant to create a novel strong consensus splice acceptor site eight bp upstream of the natural splice acceptor site in intron five. This would lead to a splice product with an out-of-frame insertion of the last eight bp of intron five, r.734_735ins735-8_735-1, predicted to be a target of nonsense-mediated decay due to a premature termination codon in exon 6, p.Tyr246Asnfs*44 (Figure 1b).

The variant was not observed previously in more than 120,000 healthy individuals in the Genome Aggregation Database. According to the Human Gene Mutation Database, the variant had been reported once as disease-causing in an individual with a confirmed diagnosis of Peutz–Jeghers syndrome, yet, further data were not available about the family (Lim et al., 2004). At the same position, thymine or guanine substitutions were observed at low frequencies in the Genome Aggregation Database and not predicted to affect splicing in silico.

To investigate whether the variant affected splicing of the transcript, we performed RNA-Seq on immortalized lymphocytes from the patient. The predicted deleterious splice product was detected at a low frequency in the absence of CHX. Yet, when the culture medium was supplemented with CHX to inhibit nonsense-mediated decay, the RNA-Seq read frequency of the frameshift transcript increased to 50%. This indicated a strong splicing impact of c.735-10C>A leading to complete mis-splicing of the transcript (Figure 2).

Interestingly, most of the aberrant transcript reads did not readily map to the reference genome but had to be identified

TABLE 1 *STK11* RNA-Seq read counts

Cycloheximide (μg/ml)	RefSeq ^a			r.734_735ins735-8_735-1 ^a		
	Mapped	Unmapped	Total	Mapped	Unmapped	Total
0	98 (93%)	7 (7%)	105	3 (7%)	40 (93%)	43
10	134 (94%)	9 (6%)	143	33 (24%)	104 (76%)	137
40	242 (98%)	5 (2%)	247	17 (7%)	230 (93%)	247

^aGenBank reference sequence for *STK11* (NM_000455.4).

using an agrep search string in the pool of unmapped reads (Table 1). Yet, these findings were confirmed using the hg38 genome build and the alternative mapper STAR.

We have submitted the *STK11* sequence variant to the locus-specific database Global Variome shared LOVD (<https://databases.lovd.nl/shared/variants/0000647291>).

4 | DISCUSSION

We characterized a novel intron five variant in *STK11* associated with Peutz–Jeghers syndrome. Located between the natural splice acceptor site and a cytosine-rich polypyrimidine tract, the variant could potentially create a strong splice acceptor site (Coolidge et al., 1997). RNA-Seq confirmed that the variant resulted in a novel transcript with a deleterious out-of-frame insertion, which underwent selective degradation as a target of NMD (Khajavi et al., 2006). Peutz–Jeghers syndrome is caused by heterozygous loss of function of *STK11* (Ylikorkala et al., 1999), hence, we may assume the cytosine to adenine substitution, c.735-10C>A, to be a causative variant. Although it might have been interesting to compare the expression of *STK11* between normal and affected tissue, immunohistochemical analysis of *STK11* expression was not part of the routine clinical work-up for Peutz–Jeghers syndrome. Analyzing lymphoblastoid cells instead of colorectal tissue may have affected the distribution of alternatively spliced transcripts in a potential limitation of the RNA-Seq analysis.

Since the parents were not known to be affected by Peutz–Jeghers syndrome, and because the variant was not detected in any of the relatives undergoing testing, we suspect that the variant had most likely occurred de novo. However, no tissue was available for testing from the diseased parents, which meant that we were unable to confirm this.

RNA-Seq, as a supplement to diagnostic germline testing, has been shown to improve the outcome of hereditary cancer predisposition testing (Karam et al., 2019). Furthermore, our analysis highlights the importance to explore the unmapped RNA-Seq reads. A low fraction of the mapped RNA-Seq reads demonstrated mis-splicing, yet, only the unmapped RNA-Seq reads revealed the full impact of the splice variant.

In conclusion, this study establishes a deleterious splicing impact of c.735-10C>A. In lack of additional clinical data to confirm its pathogenicity, we classify the variant as likely pathogenic following the guidelines of the ACMG-AMP (Richards et al., 2015).

ACKNOWLEDGMENTS

The open access article publication charge (APC) was funded by Aarhus University Hospital.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

The study was conceived and designed by all authors. UBJ and TT collected the clinical data. FW, OHL, and SV lead the sequencing experiments. FW and SV performed the bioinformatic analyses. FW and TT analyzed the results and wrote the first draft of the manuscript. All authors revised the manuscript for intellectual content, gave approval of the final version to be published, and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Thorkild Terkelsen  <https://orcid.org/0000-0003-2435-1043>

Ole H. Larsen  <https://orcid.org/0000-0001-5213-7349>

Søren Vang  <https://orcid.org/0000-0002-1899-4205>

Uffe B. Jensen  <https://orcid.org/0000-0002-6205-6355>

Friedrik Wikman  <https://orcid.org/0000-0001-9038-2003>

REFERENCES

- Abed, A. A., Gunther, K., Kraus, C., Hohenberger, W., & Ballhausen, W. G. (2001). Mutation screening at the RNA level of the *STK11/LKB1* gene in Peutz–Jeghers syndrome reveals complex splicing

- abnormalities and a novel mRNA isoform (STK11 c.597(insertion mark)598insIVS4). *Human Mutation*, 18(5), 397–410. <https://doi.org/10.1002/humu.1211>
- Andrews, S. (2010). FastQC [Computer software]. Retrieved from <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Carter, M. S., Doskow, J., Morris, P., Li, S., Nhim, R. P., Sandstedt, S., & Wilkinson, M. F. (1995). A regulatory mechanism that detects premature nonsense codons in T-cell receptor transcripts in vivo is reversed by protein synthesis inhibitors in vitro. *Journal of Biological Chemistry*, 270(48), 28995–29003. <https://doi.org/10.1074/jbc.270.48.28995>
- Coolidge, C. J., Seely, R. J., & Patton, J. G. (1997). Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Research*, 25(4), 888–896. <https://doi.org/10.1093/nar/25.4.888>
- den Dunnen, J. T., Dalgleish, R., Maglott, D. R., Hart, R. K., Greenblatt, M. S., McGowan-Jordan, J., ... Taschner, P. E. M. (2016). HGVS recommendations for the description of sequence variants: 2016 update. *Human Mutation*, 37(6), 564–569. <https://doi.org/10.1002/humu.22981>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Fokkema, I. F. A. C., Taschner, P. E. M., Schaafsma, G. C. P., Celli, J., Laros, J. F. J., & den Dunnen, J. T. (2011). LOVD vol 2.0: the next generation in gene variant databases. *Human Mutation*, 32(5), 557–563. <https://doi.org/10.1002/humu.21438>
- Hemminki, A., Markie, D., Tomlinson, I., Avizienyte, E., Roth, S., Loukola, A., ... Aaltonen, L. A. (1998). A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. *Nature*, 391(6663), 184–187. <https://doi.org/10.1038/34432>
- Jeghers, H., Mc, K. V., & Katz, K. H. (1949). Generalized intestinal polyposis and melanin spots of the oral mucosa, lips and digits; A syndrome of diagnostic significance. *New England Journal of Medicine*, 241(25), 993, illust; passim. <https://doi.org/10.1056/NEJM194912222412501>
- Jenne, D. E., Reomann, H., Nezu, J.-I., Friedel, W., Loff, S., Jeschke, R., ... Zimmer, M. (1998). Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. *Nature Genetics*, 18(1), 38–43. <https://doi.org/10.1038/ng0198-38>
- Karam, R., Conner, B., LaDuca, H., McGoldrick, K., Krempely, K., Richardson, M. E., ... Chao, E. (2019). Assessment of diagnostic outcomes of RNA genetic testing for hereditary cancer. *JAMA Network Open*, 2(10), e1913900. <https://doi.org/10.1001/jamanetworkopen.2019.13900>
- Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alfoldi, J., Wang, Q., ... MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809), 434–443. <https://doi.org/10.1038/s41586-020-2308-7>
- Khajavi, M., Inoue, K., & Lupski, J. R. (2006). Nonsense-mediated mRNA decay modulates clinical outcome of genetic disease. *European Journal of Human Genetics*, 14(10), 1074–1081. <https://doi.org/10.1038/sj.ejhg.5201649>
- Krueger, F. (2012). Trim Galore [Computer software]. Retrieved from http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
- Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., ... Maglott, D. R. (2018). ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, 46(D1), D1062–D1067. <https://doi.org/10.1093/nar/gkx1153>
- Lim, W., Olschwang, S., Keller, J. J., Westerman, A. M., Menko, F. H., Boardman, L. A., ... Houlston, R. S. (2004). Relative frequency and morphology of cancers in *STK11* mutation carriers. *Gastroenterology*, 126(7), 1788–1794. <https://doi.org/10.1053/j.gastro.2004.03.014>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, 17(1), 3. <https://doi.org/10.14806/ej.17.1.200>
- Pertea, M., Lin, X., & Salzberg, S. L. (2001). GeneSplicer: A new computational method for splice site prediction. *Nucleic Acids Research*, 29(5), 1185–1190. <https://doi.org/10.1093/nar/29.5.1185>
- Plagnol, V., Curtis, J., Epstein, M., Mok, K. Y., Stebbings, E., Grigoriadou, S., ... Nejentsev, S. (2012). A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics*, 28(21), 2747–2754. <https://doi.org/10.1093/bioinformatics/bts526>
- Rausch, T., Zichner, T., Schlattl, A., Stutz, A. M., Benes, V., & Korbel, J. O. (2012). DELLY: Structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*, 28(18), i333–i339. <https://doi.org/10.1093/bioinformatics/bts378>
- Reese, M. G., Eeckman, F. H., Kulp, D., & Haussler, D. (1997). Improved splice site detection in Genie. *Journal of Computational Biology*, 4(3), 311–323. <https://doi.org/10.1089/cmb.1997.4.311>
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., ... Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, 17(5), 405–424. <https://doi.org/10.1038/gim.2015.30>
- Stenson, P. D., Mort, M., Ball, E. V., Evans, K., Hayden, M., Heywood, S., ... Cooper, D. N. (2017). The Human Gene Mutation Database: Towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Human Genetics*, 136(6), 665–677. <https://doi.org/10.1007/s00439-017-1779-6>
- Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*, 25(9), 1105–1111. <https://doi.org/10.1093/bioinformatics/btp120>
- van Lier, M. G., Wagner, A., Mathus-Vliegen, E. M., Kuipers, E. J., Steyerberg, E. W., & van Leerdam, M. E. (2010). High cancer risk in Peutz-Jeghers syndrome: A systematic review and surveillance recommendations. *American Journal of Gastroenterology*, 105(6), 1258–1264; author reply 1265. <https://doi.org/10.1038/ajg.2009.725>
- Volikos, E., Robinson, J., Aittomaki, K., Mecklin, J. P., Jarvinen, H., Westerman, A. M., ... Aaltonen, L. A. (2006). LKB1 exonic and whole gene deletions are a common cause of Peutz-Jeghers syndrome. *Journal of Medical Genetics*, 43(5), e18. <https://doi.org/10.1136/jmg.2005.039875>
- Wildeman, M., van Ophuizen, E., den Dunnen, J. T., & Taschner, P. E. (2008). Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. *Human Mutation*, 29(1), 6–13. <https://doi.org/10.1002/humu.20654>
- Wingett, S. W., & Andrews, S. (2018). FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res*, 7, 1338. <https://doi.org/10.12688/f1000research.15931.2>

- Wu, S., & Manber, U. (1992). Agrep - A fast approximate pattern-matching tool (pp. 153–162). In *Proceedings of the Winter 1992 USENIX Conference*.
- Yeo, G., & Burge, C. B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *Journal of Computational Biology*, *11*(2–3), 377–394. <https://doi.org/10.1089/1066527041410418>
- Ylikorkala, A., Avizienyte, E., Tomlinson, I. P., Tiainen, M., Roth, S., Loukola, A., ... Aaltonen, L. A. (1999). Mutations and impaired function of LKB1 in familial and non-familial Peutz-Jeghers syndrome and a sporadic testicular cancer. *Human Molecular Genetics*, *8*(1), 45–51. <https://doi.org/10.1093/hmg/8.1.45>
- Zhang, M. Q. (1998). Statistical features of human exons and their flanking regions. *Human Molecular Genetics*, *7*(5), 919–932. <https://doi.org/10.1093/hmg/7.5.919>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Terkelsen T, Larsen OH, Vang S, Jensen UB, Wikman F. Deleterious mis-splicing of *STK11* caused by a novel single-nucleotide substitution in the 3' polypyrimidine tract of intron five. *Mol Genet Genomic Med*. 2020;8:e1381. <https://doi.org/10.1002/mgg3.1381>