



Lessons learned from the DFNA37 gene discovery odyssey

Zohreh Talebizadeh, PhD^{1,2}

Genetics in Medicine (2019) 21:1481–1482; <https://doi.org/10.1038/s41436-018-0395-8>

Despite significant advances in genetic technology, variant detection is still not as successful as it might be, even for cases and conditions that are clearly Mendelian. One way to increase the detection rate is by learning from successful discoveries. However, a successful story may include lengthy troubleshooting and failed attempts that are rarely described in scientific papers. Often, information provided in most genetic publications is only the tip of the iceberg, lacking behind-the-scenes stories about the bumpy road that led to the final results. While most troubleshooting steps may be routine and do not present new lessons, sharing those that do with the research community could be beneficial, particularly those related to false-negative variant discoveries or failed attempts caused by the complexity of the genome and exacerbated by technical limitations and human errors.

In a collaborative paper,¹ we reported on the identification of a splice-site variant in the *COL11A1* gene resulting in exon 5 skipping as a causative variant in DFNA37, a type of autosomal dominant nonsyndromic hearing loss (ADNSHL). As noted in the paper, the DFNA37 gene discovery odyssey started almost 20 years ago with the initial mapping of the DFNA37 locus, a process in which I was involved. Lessons learned from the long journey toward identification of the *COL11A1* pathogenic variant encouraged me to reconnect pieces of a puzzle, to figure out why this variant was missed earlier by routine molecular techniques.

Twenty years ago, as a graduate student, I was working on mapping the deafness genes in multiple ADNSHL families. Hearing loss in one of the examined families was mapped to a novel locus on chromosome 1p21, which was designated DFNA37. The linked region was large, containing a great number of genes. A paper published around that time on *COL11A1* variants in Marshall syndrome² made it our top candidate gene for DFNA37. So why was the *COL11A1* variant not identified at that time?

Screening of individual exons from genomic DNA (gDNA) using Sanger sequencing was a standard approach two decades ago. In addition to isolating DNA from all the DFNA37 family members, a small amount of RNA was isolated from the proband and one unaffected family member,

and complementary DNA (cDNA) primers were used in an attempt to quickly scan the *COL11A1* transcript for the causative variant. In the proband's RNA sample, but not the unaffected family member, sequencing of a cDNA fragment covering the 5' end of the gene produced a messy electropherogram pattern (data not shown). This pattern started at the junction between exons 4 and 5, indicative of a heterozygous transcript (i.e., with and without exon 5). Absence of this heterozygous sequencing product in the unaffected subject prompted us to further sequence the intron/exon junction in gDNA from these two individuals. Surprisingly, sequencing the genomic area spanning exon 5 with a forward primer inside intron 4 resulted in an out-of-phase electropherogram starting upstream of the intron 4/exon 5 junction, in both the proband and the unaffected subject. This unexpected finding in gDNA diminished our enthusiasm for pursuing the potential exon 5 skipping of *COL11A1* transcript in the proband's cDNA. The findings in cDNA and gDNA could not have been fully explained at that time. So, considering that the region harbors a rather highly polymorphic and complex sequence, we assumed that the findings were due to polymerase chain reaction (PCR) artifacts! After this false-negative exclusion of the *COL11A1* gene, we turned our attention to other potential candidate genes. We screened more than a dozen genes, particularly those enriched in the cochlea, throughout the years, but did not detect any variant. After this exhaustive, lengthy, and unsuccessful candidate gene screening, as with other variant detection odysseys, DFNA37 needed a fresh perspective and a new approach. Next-generation sequencing (NGS) provided that opportunity.

The unresolved ADNSHL case later found its way into two laboratories, at the University of Nebraska Medical Center and the University of Iowa. Coincidentally, both labs independently restarted the investigation of the disease-causing gene in this family using exome sequencing. During an exchange of information, they both discovered that they had solved the mystery of DFNA37. A novel splice-site variant (c.652-2 A>C) resulting in *COL11A1* exon 5 skipping was confirmed by both teams. This finding provided a

¹Children's Mercy Hospital, Kansas City, MO, USA; ²University of Missouri–Kansas City School of Medicine, Kansas City, MO, USA. Correspondence: Zohreh Talebizadeh (ztalebi@cmh.edu)

Submitted 28 August 2018; accepted: 20 November 2018
Published online: 10 December 2018

flashback to what I observed in the proband's cDNA, a *COL11A1* transcript missing exon 5.

In addition to the splice-site variant, NGS also revealed a heterozygous single base pair polymorphism, a deletion among a string of 12 thymidine (T) at position c.652-6-17 in both the proband and the unaffected family member. This finding explains the out-of-phase electropherograms we obtained two decades ago and elucidates why we were unable to detect the splice-site variant in gDNA at that time.

Without a doubt, current advances in NGS have revolutionized detection of pathogenic variants in hereditary disorders. The discovery of a *COL11A1* splice-site variant in DFNA37 is one successful example of such application. However, delving into the behind-the-scenes history of the methods used for detecting this variant is also important and provides helpful insight about how a polymorphic variant (i.e., changes in a string of T near the splice-site variant) masked identification of the DFNA37 variant in the initial screening using Sanger sequencing.

The value of publishing negative results, as an important piece of a scientific puzzle, has recently been recognized by the scientific community. However, insufficient appreciation of their importance and the lack of a formal framework for reporting them has hindered their inclusion in the scientific literature. Discussing failed attempts in variant detection cases and how they could have been prevented may effectively demonstrate the educational and scientific value of reviewing negative results, especially when they enhance awareness about technical challenges and genome complexity.

A similar point was made in an exome sequencing study of an autosomal recessive retinitis pigmentosa (RP) by Tucker et al.³ that exemplified how an important finding could have escaped detection because of a technical challenge.⁴ Initially, the authors prioritized homozygous variants, but none were detected. The attention was then shifted to a lower priority gene list with plausible variants, including *MAK*, which presumably harbored two different variants in one exon. Surprisingly, PCR amplification and Sanger sequencing of this exon revealed a homozygous insertion of a 353-bp Alu repeat as the disease-causing variant. This homozygous insertion was not initially captured by the exome sequencing analysis because repeat sequences were trimmed off by the analysis algorithm, and would have remained undetected had the PCR not been performed. This unexpected variant discovery highlights that even advanced genomic methods are not impeccable; adopting them routinely to guide patients' care requires improving analytical methods and human polymorphism databases to enhance detection of pathogenic variants.³ It also highlights the need to use complementary

approaches to identify variants undetected by NGS, particularly those embedded in regions containing a highly repetitive sequence, as shown by Huang et al.⁵

In addition to technical problems, human errors may also contribute to cases of overlooked variants.⁶ In the case of DFNA37, a polymorphic deletion in a repeat sequence adjacent to the splice-site variant prevented variant detection using a forward PCR primer. This variant could have been easily uncovered if in addition to a forward primer we had also used a reverse primer.

While many scientific discoveries happen as a result of unexpected observations, it is alarming how often discovery of a variant may be delayed when genome complexity is intertwined with human errors. With this in mind, the DFNA37 story, as well as the RP example, highlight several key learning points: (1) never ignore ambiguous observations because they may be a clue to scientific discoveries; (2) do not easily dismiss unexpected PCR results as artifacts, without further investigation; and (3) do not underestimate the complexity of the genome or limitations of molecular techniques. Sometimes we need to put more faith in unexpected observations and use complementary approaches, such as visual inspections, to overcome the technical and knowledge-based limitations inherent in gene discoveries.

ACKNOWLEDGEMENTS

I would like to thank Olivia Veatch (University of Pennsylvania) for contributing to candidate gene screening, and Hela Azaiez (University of Iowa) and Shelley Smith (University of Nebraska Medical Center) for reviewing this commentary and providing feedback.

DISCLOSURE

Z.T. declares no conflicts of interest.

REFERENCES

- Booth KT, et al. Splice-altering variant in *COL11A1* as a cause of nonsyndromic hearing loss DFNA37. *Genet Med*. 2018 Sep 24; <https://doi.org/10.1038/s41436-018-0285-0> [Epub ahead of print].
- Griffith AJ, et al. Marshall syndrome associated with a splicing defect at the *COL11A1* locus. *Am J Hum Genet*. 1998;62:816–823.
- Tucker BA, et al. Exome sequencing and analysis of induced pluripotent stem cell identify the cilia-related gene *MAK* as a cause of retinitis pigmentosa. *PNAS*. 2011;108:E569–76.
- Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet*. 2012;13:36–46.
- Huang XF, et al. Identification of false-negative mutations missed by NGS in retinitis pigmentosa patients: a complementary approach to clinical genetic diagnostic testing. *Genet Med*. 2015;17:307–311.
- Djemie T, et al. Pitfalls in genetic testing: the story of missed *SCN1A* mutations. *Mol Genet Genomic Med*. 2016;4:457–464.