

A close-up photograph of a person's face, focusing on the eyes and nose. The image is overlaid with several semi-transparent rectangular blocks in various colors: blue, pink, red, and grey. The text is positioned on the left side of the image.

# The Genetic Horizon

Improving clinical sensitivity in difficult-to-sequence genes for rare hereditary disorders

Meeting clinical needs using custom solutions

# From Bench to Bedside: The Genetic Horizon

The long-awaited promise of tailored treatments for individual patients based on their genetic makeup is beginning to materialize. Next generation sequencing, along with artificial intelligence, have facilitated the rapid, accurate analysis and interpretation of genomic data for disease causing variants which may be treatable through emerging gene therapy technology.

The number of clinical trials and drug development pipelines for rare disease are quickly growing and, encouragingly, the first approved treatments have been successful. As the field of precision medicine is still in its infancy, increased awareness about the clinical utility of genetic testing is needed to drive the adoption of novel diagnostic and treatment modalities.

At Blueprint Genetics, we feel that being fully transparent about our successes, challenges and limitations is key to making effective, powerful diagnostic tools available to the rare disease community. By sharing our own troubleshooting strategies, we hope to move genetic diagnostics and the promise of precision medicine forward.

## Actionability for the rare

Rare diseases comprise a heterogeneous group of thousands of genetically and clinically distinct disorders. Molecular diagnosis of these disorders has the potential to allow a patient's treatment to be tailored and optimized. Recent advances in gene therapy hold tremendous promise as a cure for these devastating diseases.

As an example, retinal gene therapy has demonstrated the potential to cure different forms of inherited vision loss. The first FDA-approved gene therapy is a one-time treatment for individuals with retinal disease such as Leber congenital amaurosis and retinitis pigmentosa caused by homozygous *RPE65* gene variants. The treatment involves delivering a functional copy of the *RPE65* gene directly to retinal cells ([www.luxturna.com](http://www.luxturna.com)).

Similarly, targeted treatments for Spinal Muscular Atrophy (SMA) have recently entered clinical trials. SMA is a severe neuromuscular disease caused by variants in the *SMN1* gene leading to the loss of motor neurons and resulting in progressive muscle weakness and paralysis. Phase 1 clinical trials have begun investigating the efficacy of gene therapy in the treatment of SMA Type 2. Treatment

involves delivering a functional copy of the *SMN1* gene in a one-time intravenous infusion ([www.avexis.com](http://www.avexis.com)).

The power of gene therapies is in the specificity of the treatments: understanding the genetic mechanism of the disease and creating a therapy to counter it. Genetic testing therefore becomes an integral part of rare disease management as determining the precise gene and variant responsible for the disease is a prerequisite for targeted treatments and gene therapy.

## The importance of transparency in clinical genetic testing

Our ability to accurately detect disease causing variants in the human genome is continuously improving. Single nucleotide variants, small insertions and deletions and exon-level deletions and duplications can be reproducibly detected, almost throughout the entire genome, and their identification is becoming more important in clinical care. Absolute transparency in genetic diagnostics is essential so that clinicians understand the quality and limitations of any given test and the suitability of the test for their patient. Blueprint Genetics is committed to providing full disclosure of our analytic validation of testing, the description of our technologies and datasets used in the diagnostic process and to the critical assessment of the performance metrics.

However, not all patients with rare genetic disease receive a molecular diagnosis. While the industry's ability to accurately detect clinically actionable variants is increasing, there are still areas that current next generations sequencing (NGS) strategies and standard Sanger sequencing do not adequately cover which leads to poor sensitivity and decreased diagnostic yield.

## Ongoing challenges

### **Clinically relevant but highly homologous and repetitive regions within important genes are an ongoing challenge to analyze:**

- Certain genomic regions are difficult to analyze due to complex sequence variations therein
- Not only are there variants in these regions but there are types of variants that we cannot detect in the clinical diagnostic laboratory with current technologies (Table 1)
- We are limited in our ability to validate certain regions/ variants because approximately 12% of the human genome is masked in gold-standard reference data sets

- Masked regions in the reference data sets contain uncertain or ambiguous variant calls that are the source of false positives, false negatives and other genotype calling errors (e.g., calling homozygous variant at a heterozygous site)<sup>1</sup>.

Blueprint Genetics remains committed to resolving difficult-to-sequence regions that are hard to validate, interpret and confirm, by developing custom solutions. In this paper, we share our strategies and what is next in our R&D pipeline.

**Table 1** Percentage of challenging regions in the human genome.

|      |   |
|------|---|
| 1.2% | Low-coverage regions                      |
| 6.9% | Segmental duplications                    |
| 6.8% | Structural variants                       |
| 0.4% | Genes with long repeat expansions         |
| 1.9% | Sub-exon level deletions and duplications |

<sup>1</sup> Zook JM et al. Nat Biotechnol. 2014 Mar; 32(3): 246–51.

# Our Approach to Difficult-to-Sequence Genes — Opening the Black Box

Our first step in providing tailored diagnostic services was to establish a specialized team of dedicated members to troubleshoot and develop custom solutions for challenging genetic regions. Our Clinical Research & Development (R&D) team consists of genetics and molecular biology professionals. The team's core focus is on genomic regions too complex to resolve with standard NGS analysis and developing new methods to analyze technically challenging, but clinically important, genes.

## **Difficult-to-analyze genomic regions include genes that have:**

- pseudogenes
- other highly homologous regions in the genome
- longer stretches of repetitive sequences

Additionally, inversions, translocations, and other complex genetic rearrangements are not typically amenable to NGS methods and require custom methods for successful sequencing and variant confirmation.

Our Clinical R&D team's work is driven by the opportunity to tackle the most challenging parts of our genome and by the reward inherent in finding a genetic diagnosis for a patient.



# Clinical Research & Development: Going the Extra Mile

## Innovative, customized solutions for difficult-to-sequence genes

Developing new methods to analyze challenging genes requires innovations in both bioinformatic data analysis algorithms as well as laboratory protocols optimized for a particular genetic region.

To date, our key accomplishments with this approach include successful sequencing of the highly repetitive ORF15 region of the *RPGR* gene, associated with X-linked retinitis pigmentosa, accurate determination of the copy numbers of each of the highly homologous *SMN1* and *SMN2* genes, associated with spinal muscular atrophy and a method to identify *Alu* element insertions and variants in the polycystic kidney disease associated *PKD1* gene.

When a variant is identified in the NGS data that routine Sanger sequencing is not able to confirm, a customized confirmation assay is needed. Our Clinical R&D team develops custom solutions to confirm these variants for individual patients after the reporting geneticist refers the case to the Clinical R&D team for evaluation and work-up.

### **The need for custom variant confirmation solutions may arise when:**

- a low-quality variant is detected in a difficult-to-sequence or duplicated region,
- unusual SNP clusters are observed, or
- other quality control steps, such as coverage analysis, detects an event deviating from normal

### **The R&D team then develops custom laboratory methods that often involve:**

- Specialized primer and probe designs
- Long-range PCR and sequencing
- Quantitative PCR

Results of a custom confirmation analysis may take up to 4–6 weeks depending on the scale of development work required.



# A Clinician–Lab Partnership

## Detailed Clinical information can make or break the diagnosis, especially with difficult-to-sequence genes

Sometimes, when a likely complex disease-causing variant is identified, and a detailed clinical information is provided, the reporting geneticist will refer the case to Clinical R&D to work on customized solutions. The need for customization may arise when a low-quality variant is detected at difficult-to-sequence or duplicated region, when strange SNP clusters are observed or when other quality control steps such as coverage analysis detects any events deviating from normal.

### Did you know?

- At the exome level, every patient has approximately 20,000–35,000 variants. The challenge is to find the 1 or 2 that are related to the patient's phenotype
- Detailed clinical information may prompt additional scrutiny, beyond what is called by the bioinformatic pipeline, when the patient's phenotype is consistent with variants in a particular gene
- Too little clinical information may result in more VUS being reported or in clinically relevant variants not being reported at all as they are not consistent with the patient's phenotype as reported to the lab

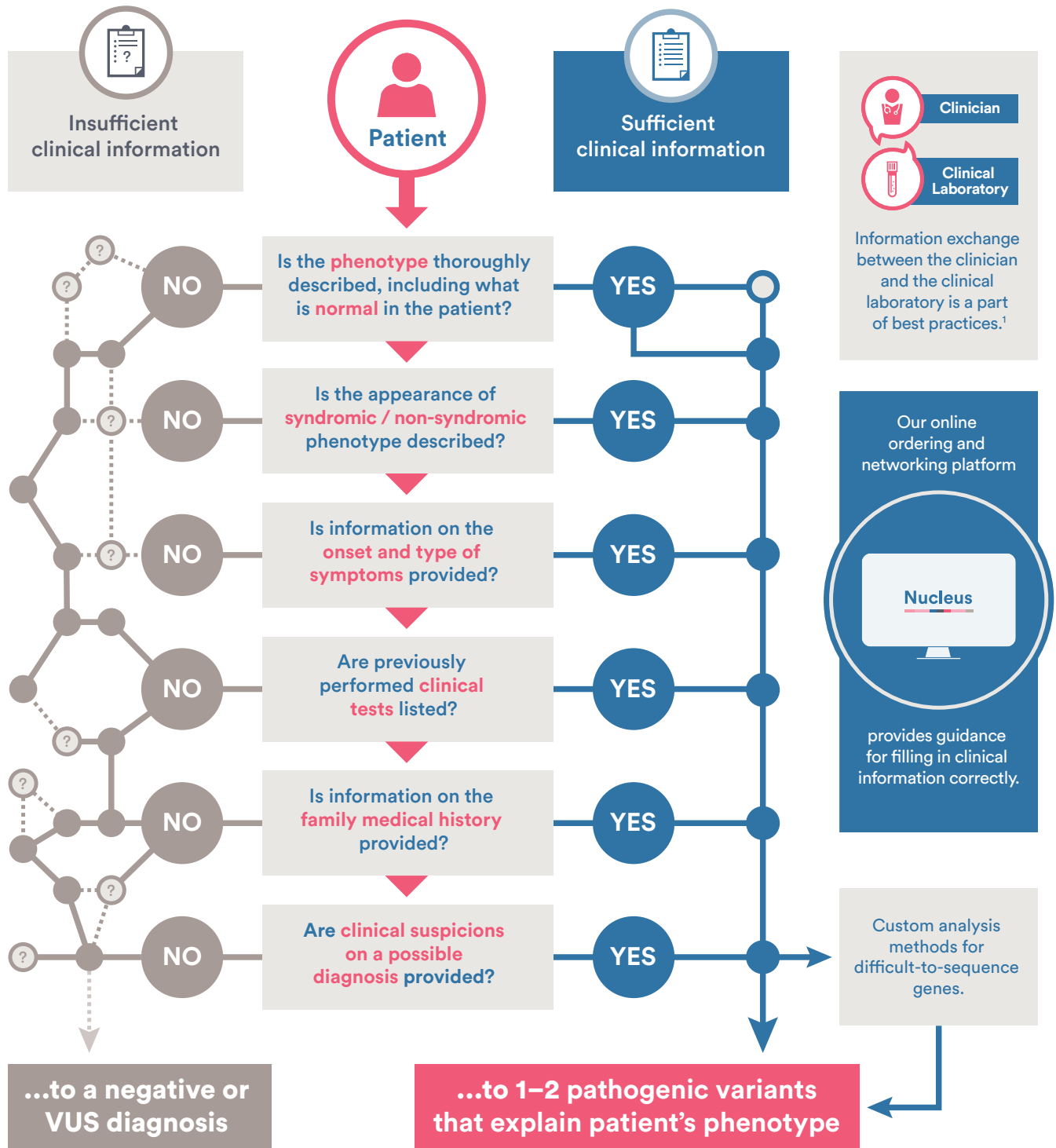
## A quick roadmap to providing sufficient clinical information about your patient

- 1 Provide a clear description of the phenotype—list your patient's signs and symptoms and also include what is normal about your patient. Be thorough. Sometimes the most atypical feature is the most valuable.
- 2 Tell us if the phenotype appears syndromic or non-syndromic in your patient.
- 3 Describe when and how your patient's symptoms began.
- 4 Provide results of all previous testing including those that have been normal.
- 5 Describe the relevant family history. A pedigree with phenotypic information and age of onset is helpful.
- 6 Share your thoughts and clinical suspicions about what the diagnosis could be.

List provided by Helena Kääriäinen (MD, PhD) and Eveliina Salminen (MD, PhD).

# Clinical Information Can Lead to Finding a Variant That Might Otherwise Be Missed

From 20,000–35,000 possible variants explaining the phenotype...



<sup>1</sup> Bush et al. on behalf of the ACMG SELI committee. Genet Med. 2018 Feb; 20(2): 169–171.

# Difficult-to-Sequence Genes at a Glance — Key Examples

## **RPGR and X-linked retinitis pigmentosa**

### A general overview

X-linked retinitis pigmentosa (XLRP) accounts for 10–20% of families with retinitis pigmentosa (RP) and is the most severe form of RP. Males with XLRP become symptomatic in early childhood and most have complete vision loss by the end of their third decade. Female carriers have a broad spectrum of fundus appearances, ranging from normal to extensive retinal degeneration. Typically, retinal disease in females with XLRP is less severe than that seen in males. In a study by *Rozet et al.*, age at disease onset in affected females was delayed compared to affected males with similar truncating variants (20–40 years vs. 10–20 years; PMID: 11950860).

Variants in the *RPGR* gene are associated mainly with X-linked retinitis pigmentosa (XLRP, OMIM #300029) and mutations in this gene account for over 70% of patients with XLRP. *RPGR* (OMIM \*312610) encodes a protein with a series of six RCC1-like domains (RLDs), characteristic of the highly conserved guanine nucleotide exchange factors. This protein localizes to the outer segment of rod photoreceptors and is essential for their viability. RP affects 1/5,000 worldwide (PMID: 30137803) and, until recently, treatments were symptomatic.

### Why is *RPGR* important and what are the challenges in its analysis?

The ORF15 exon of *RPGR* is a known hotspot for pathogenic variants. ORF15 encodes 567 amino acids and has a repetitive domain with high glutamic acid and glycine content (PMIDs: 10932196, 12657579) which make sequencing of this region technically challenging with traditional NGS technologies. The *RPGR* isoform which includes ORF15 is encoded by exons 1–15 and part of intron 15 (1152 amino acids, transcript id NM\_001034853). Pathogenic variants have been identified in exons 1–15 of ORF15 while no disease-causing variants have been reported in exons 16–19 (PMID: 17195164). Currently, HGMD lists 203 different *RPGR* mutations in NM\_000328.2 and 231 in NM\_001034853.1 (ORF15) (HGMD Professional 2018.1). The majority of the variants are nonsense and frameshift variants leading to protein loss of function.

### **There is notable inter- and intrafamilial phenotypic variability in XLRP caused by *RPGR* mutations:**

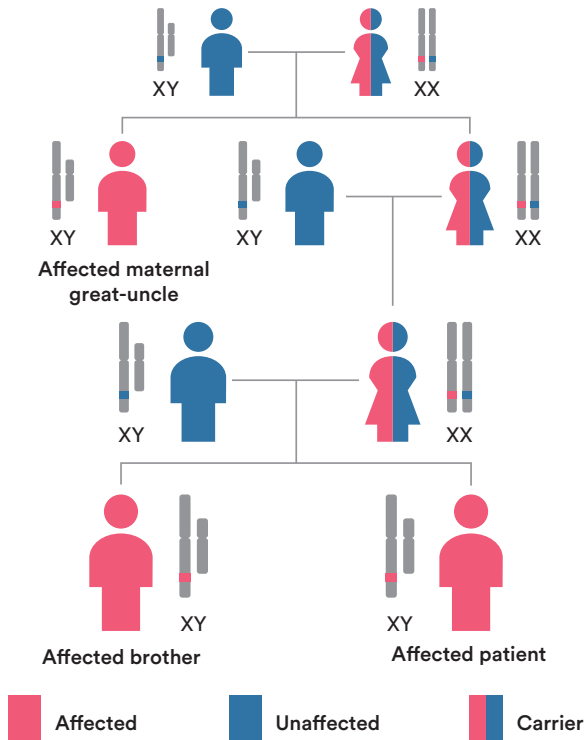
- Patients with mutations in exons 1–14 have been shown to demonstrate smaller visual fields than patients with mutations in the ORF15 region (PMID: 14564670).
- Truncating variants in the C-terminal part of ORF15 have been associated with XL cone-rod dystrophy (*i.e.*, c.2965G>T p.Glu989\*, c.3197\_3198delAG, c.3300\_3301delTA) (HGMD; PMID: 23150612).



# Retinitis pigmentosa (*RPGR*-ORF15) case report

## Patient information

Patient is a 15-year-old male with retinitis pigmentosa. His brother and maternal great-uncle are also affected.



**Figure 1**

Family history of a patient with retinitis pigmentosa.

## Genetic testing

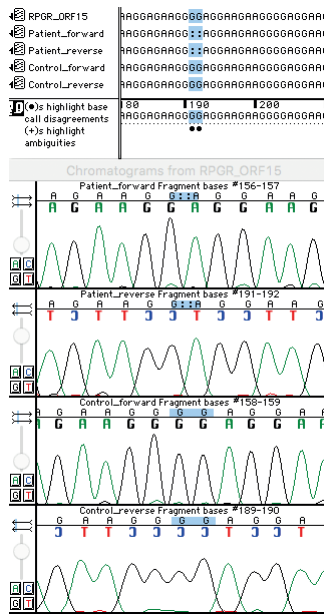
A comprehensive Blueprint Genetics Retinal Dystrophy Panel was requested which tests for 266 genes including copy number variants and assessment of known disease-causing, deep intronic variants. The Retinal Dystrophy panel is ideal for patients with a clinical suspicion or diagnosis of isolated retinitis pigmentosa or isolated or syndromic retinal dystrophy.

## Diagnostic summary

Sequence analysis identified a pathogenic hemizygous variant c.2601\_2602del, p.(Glu868Glyfs\*210) in the technically challenging *RPGR* ORF15 region leading to the diagnosis of X-linked retinitis pigmentosa (XLRP,

MIM#300029). The variant was confirmed with a custom Sanger sequencing method optimized for the ORF15 region consisting of GA-rich repetitive sequence.

This variant is rare, results in a frameshift transcript with a premature stop codon and is predicted to cause loss of normal protein function through protein truncation. The variant was initially reported in one family when Bader et al. conducted a comprehensive screening for *RP2* and *RPGR* gene mutations including *RPGR* exon ORF15 in 58 index patients with X-linked retinitis pigmentosa (described as g. ORF15 848\_849delGG, PMID: 12657579). Subsequently, the variant has been reported in additional male patient with X-linked RP (PMID: 28322733). In addition, we have detected the variant in two patients with X-linked retinitis pigmentosa (unpublished observations).



**Figure 2**

Sanger confirmation of patient's variant c.2601\_2602del, p.(Glu868Glyfs\*210). Confirmation of variants found in *RPGR* ORF15 is made possible with a customized sequencing method, optimized for highly repetitive GA-rich sequence.

## Diagnostic implications

- The diagnosis of *RPGR*-related X-linked retinitis pigmentosa was confirmed
- Genetic counseling and family member testing are now available to family members to clarify their risk to carry the disease causing variant or develop symptoms.
- Because the genetic etiology of this patient's retinitis pigmentosa has been identified, they may be eligible to participate in gene therapy clinical trials for patients with pathogenic *RPGR* variants.

## SMN1/SMN2 and spinal muscular atrophy

### A general overview

Spinal muscular atrophies (SMA) are autosomal recessive disorders characterized by degeneration of the anterior horn cells of the spinal cord, leading to symmetrical muscle weakness and atrophy (OMIM #253300). Five types of SMA are recognized based on the age of onset, the maximum muscular activity achieved, and survivorship (GeneReviews). Homozygous loss of the survival of motor neuron 1 gene (*SMN1*), caused by deletions or point mutations, causes SMA (OMIM #253300). Homozygous absence of exon 7 of *SMN1* has been identified in approximately 95% of patients with SMA (PMID: 9950358).

The incidence of SMA is estimated to be 4-10/100,000 live births in Europe and USA with a carrier frequency of 1/50-90 (GeneReviews). There have been no approved drug treatments for SMA until recently when a novel drug, Nusinersen, became available (PMID: 27939059).

### Why are *SMN1* and *SMN2* copy number important and why are they challenging to analyze?

*SMN1* produces a full-length transcript, whereas *SMN2* predominantly produces an alternatively spliced transcript lacking exon 7, which results in a less stable SMN protein (PMIDs: 10339583, 10607836). Absence of *SMN1* is partially compensated for by *SMN2*, which produces enough SMN protein to allow for relatively normal development in cell types other than motor neurons (MIM \*600354). However, *SMN2* cannot fully compensate for loss of *SMN1* because the majority of *SMN2* transcripts are truncated and less stable than *SMN1* transcripts (PMIDs: 7813012, 17307868). Milder SMA phenotypes are usually associated with a gene conversion of *SMN1* sequences into *SMN2* sequences, or more than 2 copies of *SMN2* (PMIDs: 14705979, 16508748, 11839954).

*SMN1* and *SMN2* are located close to each other at the complex SMN region on chromosome 5q12.2-q13.3 where repetitive sequences, pseudogenes, transposable elements, deletions and inverted duplications are not unusual (PMID: 9950358). The *SMN1* gene (MIM \*600354) shares more than 99% nucleotide identity with the *SMN2* gene (MIM \*601627); both genes encode a 294-amino acid RNA-binding protein, SMN, that is required for efficient assembly of small nuclear ribonucleoprotein (snRNP) complexes. These two genes, both containing nine exons, can be distinguished only by eight nucleotides (5 intronic and 3 exonic with 1 each located in exons 6, 7, and 8) (PMID: 9950358) making molecular differentiation of these genes extremely difficult.

## Spinal muscular atrophy (*SMN1*/*SMN2*) case report

### Patient information

Patient is a 60-year-old male with spinal muscular atrophy type III diagnosed at the age of 27 years based on electromyoneurography (ENMG) and muscle biopsy findings. The patient presented with symptoms throughout childhood and adolescence including clumsiness while running and difficulty climbing stairs. He is unable to transition independently, has generalized weakened control of body and cervical spine movements, and shortness of breath. There is no known family history of similar disease.

### Genetic testing

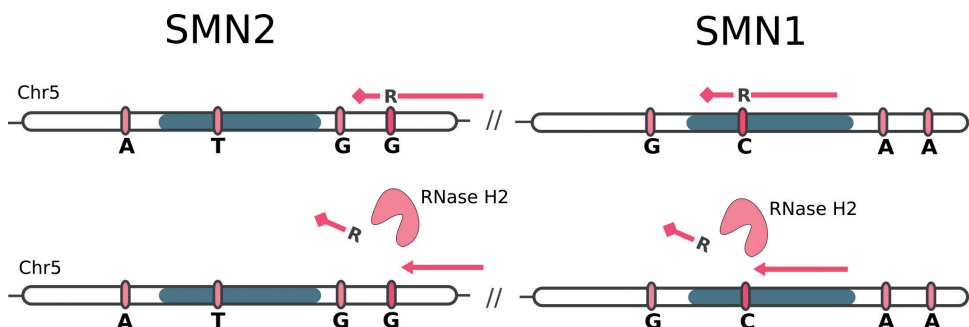
A Blueprint Genetics Spinal Muscular Atrophy Panel was requested which tests for 30 genes including copy number variants and assessment of known disease-causing, deep intronic variants. The Spinal Muscular Atrophy Panel is ideal for patients with a clinical suspicion of distal hereditary motor neuropathy or spinal muscular atrophy.

### Diagnostic summary

Bioinformatic analysis developed specifically for copy number determination of the *SMN1* and *SMN2* genes identified a homozygous deletion including at least exon 7 of the *SMN1* gene, which is considered as a marker for whole gene deletion. In addition,  $\geq 3$  copies of *SMN2* were detected. These findings were confirmed using a RNase H2-dependent PCR (rhPCR) developed specifically for *SMN* copy number analysis. Considering the current literature and well-established role of *SMN1* deletions as a disease-causing variant, it was classified as pathogenic.

### Diagnostic implications

- The diagnosis of SMA was confirmed.
- Genetic counseling and family member testing are now available to family members to clarify their risk to carry or develop the disease.
- Because the genetic etiology of this patient's SMA and the number copy number of *SMN2* has been identified, they may be eligible to participate in gene therapy clinical trials for patients with multiple copies of *SMN2*.



**Figure 3**

Our *SMN* copy number confirmation method takes advantage of RNA-blocked primers, which are activated by an RNase H2 enzyme only when stable primer annealing takes place. This allows us to differentiate the copy numbers for highly homologous *SMN1* and *SMN2* genes.

## **PKD1 and autosomal dominant polycystic kidney disease**

### **A general overview**

Autosomal dominant polycystic kidney disease (ADPKD) is an adult-onset multisystem disorder characterized by cysts in the kidneys, liver, seminal vesicles, pancreas, and arachnoid membrane. Vascular abnormalities include intracranial aneurysms, aortic dilatation and dissection and abdominal wall hernias. Renal complications include hypertension and renal insufficiency. Approximately 50% of individuals with ADPKD have end-stage renal disease (ESRD) by the age of 60 years. Substantial variability in severity of renal disease and extrarenal manifestations occurs even within the same family.

The prevalence of ADPKD is between 1:400 and 1:1,000. Approximately 10% of people who receive a renal transplant have ADPKD. The penetrance is very high; almost all adults with a *PKD1* or *PKD2* pathogenic variant have multiple bilateral cysts. The penetrance for ESRD is reduced although the majority of individuals with truncating *PKD1* variants develop ESRD during their lifetime. (ADPKD1 - GeneReviews).

### **Why is PKD1 important and why is it challenging to analyze?**

Mutations in *PKD1* cause autosomal dominant polycystic kidney disease type 1 (ADPKD1; OMIM \*601313). The *PKD1* gene encodes polycystin-1 protein, a glycoprotein containing a large N-terminal extracellular region, multiple transmembrane domains and a cytoplasmic C-tail. It is an integral membrane protein that functions as a regulator of calcium permeable cation channels and intracellular calcium homeostasis. Polycystin-1 forms a complex with polycystin-2 that regulates multiple signaling pathways to maintain normal renal tubular structure such as cilium length and function. It is involved in fluid-flow mechanosensation by the primary cilium in renal epithelium. There are a total of 1273 *PKD1* variants classified as definitely pathogenic, highly likely pathogenic or likely pathogenic in the PKD1 Mutation Database (Feb 2018); frameshifts account for 35.3%, missense for 24.6%, nonsense for 20.7%, splice site for 9.7%, in-frame deletions/insertions for 8.1% and gross deletions/duplications for 1.6%. There is no mutational hot spot for *PKD1* and *PKD2*, which means variants are usually private, highly variable and spread throughout the entire gene. It is estimated that *PKD1* explains 85% and *PKD2* 15% of ADPKD cases. *PKD2*-related ADPKD is widely acknowledged to be of milder severity than *PKD1*-related disease

The genomic region encompassing exons 1–33 of the *PKD1* (chr16:2147418-2187265) shares 97.7% sequence homology with six known pseudogenes on chromosome 16, making this region very difficult to resolve using traditional NGS methods and resulting in reduced sensitivity to detect disease-causing variants.

# Polycystic kidney disease (PKD1) case report

## Patient information

Patient is a 25-year-old female diagnosed with polycystic kidney disease at 4 years of age. There is a family history of ADPKD in multiple generations.

## Genetic testing

A Blueprint Genetics Polycystic Kidney Disease Panel was requested which tests for 10 genes including assessment of non-coding variants. The Polycystic Kidney Disease Panel is ideal for patients suspected to have autosomal dominant or autosomal recessive polycystic kidney disease.

## Diagnostic summary

The patient was identified to have a pathogenic heterozygous nonsense variant c.4957C>T, p.(Gln1653\*) in exon 15 of *PKD1*. This variant was confirmed with Sanger sequencing using primers that were manually designed to maximize their *PKD1* specificity over several highly homologous pseudogenes.

The variant is rare, causes a premature stop codon and is thus predicted to cause loss of normal protein function either through protein truncation or nonsense-mediated mRNA decay. The variant has been listed as pathogenic in the PKD1 Mutation Database, where it has been reported in four families with ADPKD.

## Diagnostic implications

- The diagnosis of ADPKD was confirmed.
- Genetic counseling and family member testing are now available to family members to clarify their risk to develop the disease.
- The molecular diagnosis allows for testing of asymptomatic family members to identify those at risk of developing disease in addition to identifying related kidney donors.

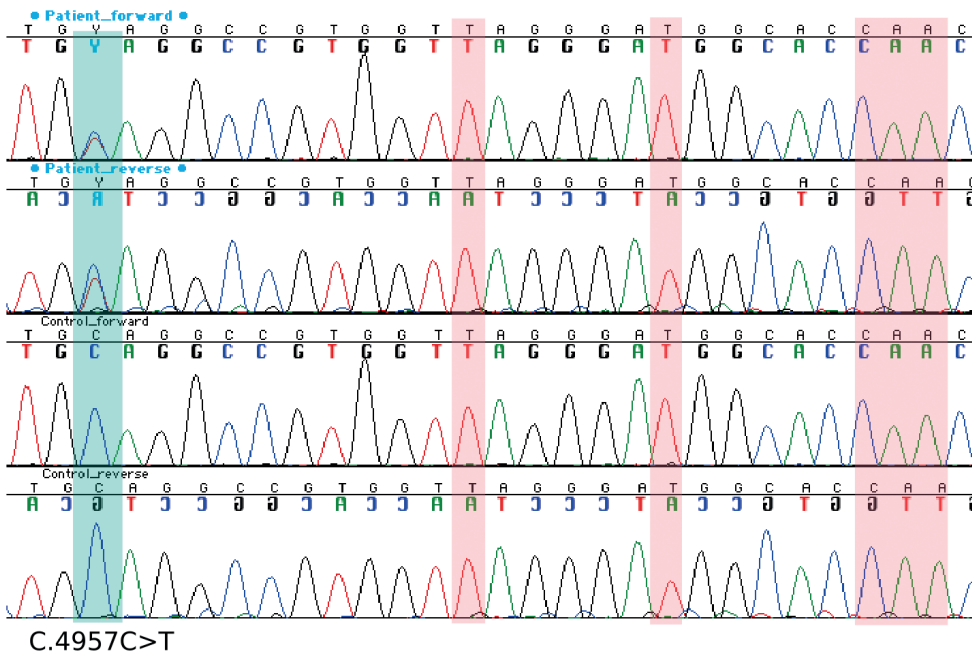


Figure 4

*PKD1* has multiple highly homologous pseudogenes in the genome. Patient's variant c.4957C>T, p.(Gln1653\*) is clearly visible in forward and reverse Sanger sequences, while no contaminating signal from the pseudogenes is observed in positions that are unique for *PKD1* (marked in red).

# On the R&D Horizon

We are constantly striving to expand our tests to cover more difficult-to-analyze genetic regions and variants. This means improving our NGS protocols and sequence data analysis algorithms as well as developing new ancillary assays when required.

## Upcoming targets

### Maximizing high-quality sequencing data:

- *OPN1MW*  
(Blue cone monochromacy, cone dystrophy)
- *OPN1LW* (Blue cone monochromacy, cone dystrophy)

### On-going validations for the detection of structural variants:

- *CYP21A2* (Congenital adrenal hyperplasia)
- *HBA1* (Hemoglobinopathy)
- *HBA2* (Hemoglobinopathy)
- *MSH2* (Lynch syndrome)



# Our commitment to the genetic community

We urge researchers, genetic testing laboratories, healthcare professionals, patients and other organizations to share their thoughts on what clinically relevant, challenging genes should be looked at next. Where we focus our R&D efforts arise from the feedback we receive from the healthcare community and unmet clinical needs in rare inherited disorders.

As part of our commitment to the clinical community, we will continue to share our advancements in the area of difficult-to-sequence genes.

Also available online at [blueprintgenetics.com](https://blueprintgenetics.com)

## Contact information

### Blueprint Genetics

Keilaranta 16 A-B,  
02150 Espoo  
Finland

### Customer support

Europe, Middle East, Asia:  
[support@blueprintgenetics.com](mailto:support@blueprintgenetics.com)  
+358 40 2511 372

North and South America:  
[support.us@blueprintgenetics.com](mailto:support.us@blueprintgenetics.com)  
+1 650 452 9340

Canada:  
[support.ca@blueprintgenetics.com](mailto:support.ca@blueprintgenetics.com)  
+1 833 697 4665

Australia:  
[support.au@blueprintgenetics.com](mailto:support.au@blueprintgenetics.com)  
+61 1800 952 540

Our website [blueprintgenetics.com](https://blueprintgenetics.com) also offers a wide range of FAQ's

Clinical Genetics Support  
[genetics.support@blueprintgenetics.com](mailto:genetics.support@blueprintgenetics.com)

**At Blueprint Genetics,  
we feel that **transparency**  
is the key to unlocking  
the mysteries of genetic  
testing and providing  
powerful and novel  
tools to the rare disease  
community.**

Join the conversation  
#GeneticKnowledge



[www.blueprintgenetics.com](http://www.blueprintgenetics.com)

*We are continuously developing our services and offering. We may amend service descriptions from time to time by posting new versions on our website. For up-to-date information, please visit [blueprintgenetics.com](http://blueprintgenetics.com).*

**Blueprint Genetics**

