Finding the Causes of Rare Diseases

By definition, a rare disease is one that is only observed in a small percentage of the population. Despite their individual scarcity, the total number of rare diseases is estimated to be 5,000–8,000 worldwide, affecting 6–8% of the population over their lifetime[1]. Most of these rare diseases have a genetic component such as an inherited or *de novo* mutation. Diagnosis of these diseases can be difficult due to the challenge of identifying a rare causal variant. If the variant is a *de novo* mutation, the lack of an informative family pedigree can complicate this challenge even further. Patients and their families often embark on an extensive diagnostic odyssey, and some never reach a definitive diagnosis.

But there is hope. Next-generation sequencing (NGS) technologies are proving to be remarkably successful in finding the causes of rare genetic diseases. NGS technology can be used to identify causative *de novo* or inherited variants in a single test[2,3], without prior knowledge of these variants[4].

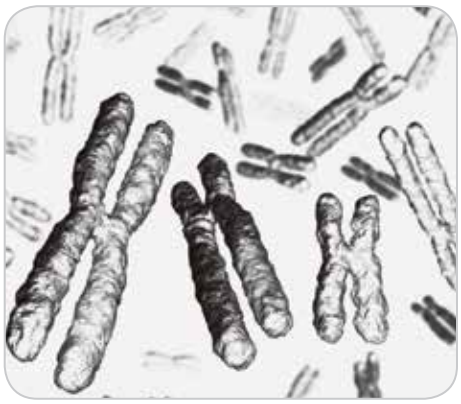
Introduction

Rare genetic diseases are often characterized by early-onset and severe symptoms. Although they are often genetic in nature, rare diseases are not always inherited and can appear spontaneously in a generation as a result of *de novo* mutations. In addition to aiding in diagnosis, understanding the genetic basis of a particular condition—whether the mutation is inherited or *de novo*—may also inform risk assessment for a patient's family members.

Rare genetic diseases can take a significant emotional and financial toll on the afflicted patients and their families. A rapid and accurate diagnosis may help alleviate some of this hardship and avoid potentially costly or invasive testing. By providing a base-by-base view of the entire human genome, NGS can offer insight into genomic changes and help accelerate reliable identification of causative variants and potential treatment options. The key is using a technology that produces high-quality data capable of detecting these rare mutations with a high degree of confidence. Alternate methods can also be limited in scope, targeting specific areas of the genome that may or may not be related to the observed disease. NGS provides a genome-wide view that can detect variants that are sometimes missed using other technologies.

Affected Populations

Depending on the number of patients afflicted with a disease and their relationship to one another (biologically related vs. unrelated), one of four general strategies can be employed to identify causative genetic variants (Table 1).



After the total set of variants has been found, various methods can be used to isolate a smaller set of potentially causative variants. As an initial step in this process, the variants in a patient can be compared to a well-characterized control group, e.g., a healthy set of individuals. A variant that occurs in a healthy control group is unlikely to cause the disease and can be eliminated as a candidate. In most cases, several approaches will be required to find the subset of variants that may be causative.

Table 1: Clinical Research Strategies for Identifying Disease Variants with NGS*

Situation	Assumption	Strategy	Approach	Consideration
Multiple patients affected within a single family	Fully penetrant mutation segregated with the disorder	Linkage	Sequence the patient and his/her biological family members, whether affected by the disease or not	Additional family members can be sequenced to limit search region
Single patient sporadically affected within a family	Fully penetrant causative *de novo* mutation	Trio	Sequence the patient and his/her biological mother and father	Does not rely on healthy control groups or other prioritization assumptions; identifies recessive Mendelian disease as well
Multiple unrelated patients showing the same phenotype	The disorder is completely (or mostly) monogenic and all patients suffer from the same disorder	Overlap	Sequence the unrelated patients for comparison	Only a limited set of patients are needed
Single patient affected by disease, parents unavailable	A single rare homozygous or two rare compound heterozygous mutations	Affected only	Sequence the patient only	Only a single patient is required; relies on reference databases

* Table adapted from Gilissen C, Hoischen A, Brunner HG, Veltman JA (2012) Disease gene identification strategies for exome sequencing. Eur J Hum Genet. **20:** 490–497.

Sequencing to Identify Disease Variants

Several sequencing strategies can be used to identify variants. The choice of strategy depends upon sample availability, current knowledge of the disease variant, and the extent of sensitivity and specificity required to answer the clinical question.

Sequencing Approaches

When studying a rare genetic disease, it may be necessary to sequence the entire genome to obtain sufficient information to find the causative variant. Until recently, this was a major endeavor. Today, NGS can produce a large volume of data in a short time, making it a powerful tool for whole-genome and whole-exome sequencing, as well as targeted resequencing.

Whole-Genome Sequencing (WGS)

WGS provides visibility to variants across the whole genome, including coding, regulatory, and intronic regions. It can reveal disease-relevant variants as well as those that are unrelated to a particular disease phenotype. Out of all the sequencing approaches, WGS provides the most comprehensive view of the genome, but current costs and analytical obstacles make widespread adoption challenging.

Saunders CJ, Miller NA, Soden SE, Dinwiddie DL, Noll A, et al. (2012) Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. Sci Transl Med 4:154ra135.

In newborns with genetic defects, disease progression is extremely rapid and molecular diagnoses must occur quickly to be relevant for clinical decision making. WGS and automated data analysis are used to deliver a differential diagnosis of genetic disorders of 7 children, 2 with known molecular diagnoses and 5 presenting symptoms suggesting genetic disorders, within 50 hours. This is substantially faster than the current norm of 4 to 6 weeks.

Illumina Technology: HiSeq® 2500 rapid WGS with 100 bp paired-end reads with a 34-41-fold aligned genome coverage

Whole-Exome Sequencing (WES)

WES targets the protein coding regions, or exons, and, in some cases, the intron/exon boundaries within the genome. To date, most disease-causing mutations have been found to reside in these protein-encoding regions of the genome. Sequencing just the exome directs efforts towards the genomic regions that contain variants with the most characterized disease impact. It produces less data, simplifying analysis. However, it is important to note that any disease-causing variant or variants that lie outside of the exonic region will not be found with WES.

Kim JC, Lee NC, Hwu PW, Chien YH, Fahiminiya S, et al. (2012) Late onset of symptoms in an atypical patient with the cblJ inborn error of vitamin B(12) metabolism: Diagnosis and novel mutation revealed by exome sequencing. Mol Genet Metab 107:664–8.

Exome sequencing was used to diagnose a rare inborn error of vitamin B12 metabolism. This is only the third known patient exhibiting the cblJ disorder. It appears that the observed mutation is less damaging than previously reported *ABCD4* mutations and the disease phenotype may be broader than previously recognized. The authors note that the unusual presentation precluded diagnosis using standard biochemical and genetic approaches.

Illumina Technology: HiSeq 2000 with 100 bp paired-end reads of exome captured with the TruSeq® Exome Enrichment Kit

Targeted Resequencing

Targeted resequencing focuses on a set of specific genes or regions. Expert-selected panels[3], such as the Illumina TruSight™ Inherited Disease content set, offer a cost-effective and focused way to interrogate known disease genes.

Kingsmore SF, Dinwiddie DL, Miller NA, Soden SE, Saunders CJ (2011) Adopting orphans: comprehensive genetic testing of Mendelian disease of childhood by next-generation sequencing. Expert Rev Mol Diagn. 11: 855–868.

The authors review the utility of NGS as a tool for scalable, economical, rapid, and multiplexed diagnosis of rare genetic disorders. They also describe a targeted panel for 592 childhood diseases in a multiplexed sequencing run using NGS.

Illumina Technology: HiSeq 2000 and the precursor to the TruSight Inherited Disease content set

Finding the Causative Variant

WES and WGS often detect large numbers of variants per sample, making further analysis necessary to isolate the key variant or set of variants responsible for a disease. Several analytical tools, such as Illumina VariantStudio software, are available to simplify this process. With VariantStudio software, clinical researchers and laboratory directors can easily isolate variants of interest. Sequencing data can be filtered according to read depth or quality scores to identify true variants, or by annotations such as known disease association, allele frequency, or functional impact. Family-based filters leverage available pedigrees, enabling the identification of variants consistent with a suspected inheritance mode. Variants can be classified into categories such as pathogenic, likely pathogenic, and benign, and can then be reported to physicians. To learn more, visit www.illumina.com/variantstudio.

Looking to the Future...

Traditional rare genetic disease diagnosis based on phenotypic interpretation can often be difficult and time-consuming. With more than 3,500 monogenic diseases characterized, and clinical tests only available for a limited number[5], the availability of a single test, such as WGS/WES, that covers all variants in one pass would offer a distinct advantage.

Illumina is dedicated to helping geneticists in this endeavor. The proven, widely used Illumina NGS technology offers reliable solutions for rapid, accurate variant identification through sequencing (Table 2), moving science forward to a better understanding of rare diseases.

To learn more, visit www.illumina.com/raredisease

Table 2: Illumina NGS Solutions*

Solution	Application
TruSight One Panel	Largest targeted sequencing panel covering genes that have an associated clinical phenotype, designed for trio sequencing with a single run on the MiSeq® system
TruSight Sequencing Panels	Expert-defined, disease-specific panels, enabling laboratories to realize the benefits of NGS
Illumina Clinical Services Laboratory	Whole-genome sequencing services in a CLIA-certified, CAP-accredited laboratory for diagnosis or preventive care
TruSeq DNA PCR-Free Library Preparation Kit	Faster library preparation providing unbiased data quality and unprecedented coverage for whole-genome sequencing
Nextera® Rapid Capture Exome Kit	All-in-one kit for library preparation and exome enrichment that allows identification of coding variants 70% faster than any other method
Nextera Rapid Capture Custom Enrichment Kit	Customizable assay enabling laboratories to add unique regions of interest to established panels, such as the Nextera Rapid Capture Exome Kit or TruSight panels
TruSeq Custom Amplicon	End-to-end solution for custom, targeted amplicon resequencing enables a wide range of applications for identifying genetic variants in a rapid and efficient manner
HiSeq 2500 System	Fast, highly accurate whole-genome and whole-exome sequencing
MiSeq Benchtop Sequencer	Fast, highly accurate sequencing technology for targeted sequencing
Illumina VariantStudio Data Analysis Software	Easy-to-use desktop application that provides rich annotation, interactive filtering tools, and customizable reporting to identify disease-relevant and actionable variants

* For a complete list of Illumina NGS solutions, visit www.illumina.com.

References

1. Ayme S, Rodwell C, eds (2012) 2012 Report on the State of the Art of Rare Disease Activities in Europe of the European Union Committee of Experts on Rare Diseases.
2. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, et al. (2010) Exome sequencing identifies *MLL2* mutations as a cause of Kabuki syndrome. Nat Genet 42: 790–793.
3. Kingsmore SF, Dinwiddie DL, Miller NA, Soden SE, Saunders CJ (2011) Adopting orphans: comprehensive genetic testing of Mendelian diseases of childhood by next-generation sequencing. Expert Rev Mol Diagn. 11: 855–868.
4. To learn how NGS works, review the Illumina Sequencing Primer (www.illumina.com/documents/products/Illumina_Sequencing_Introduction.pdf).
5. Saunders CJ, Miller NA, Soden SE, Dinwiddie DL, Noll A, et al. (2012) Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. Sci Transl Med 4: 154ra135.

FOR RESEARCH USE ONLY

Pub No. 770-2013-005 Current as of 19 November 2014

illumina®