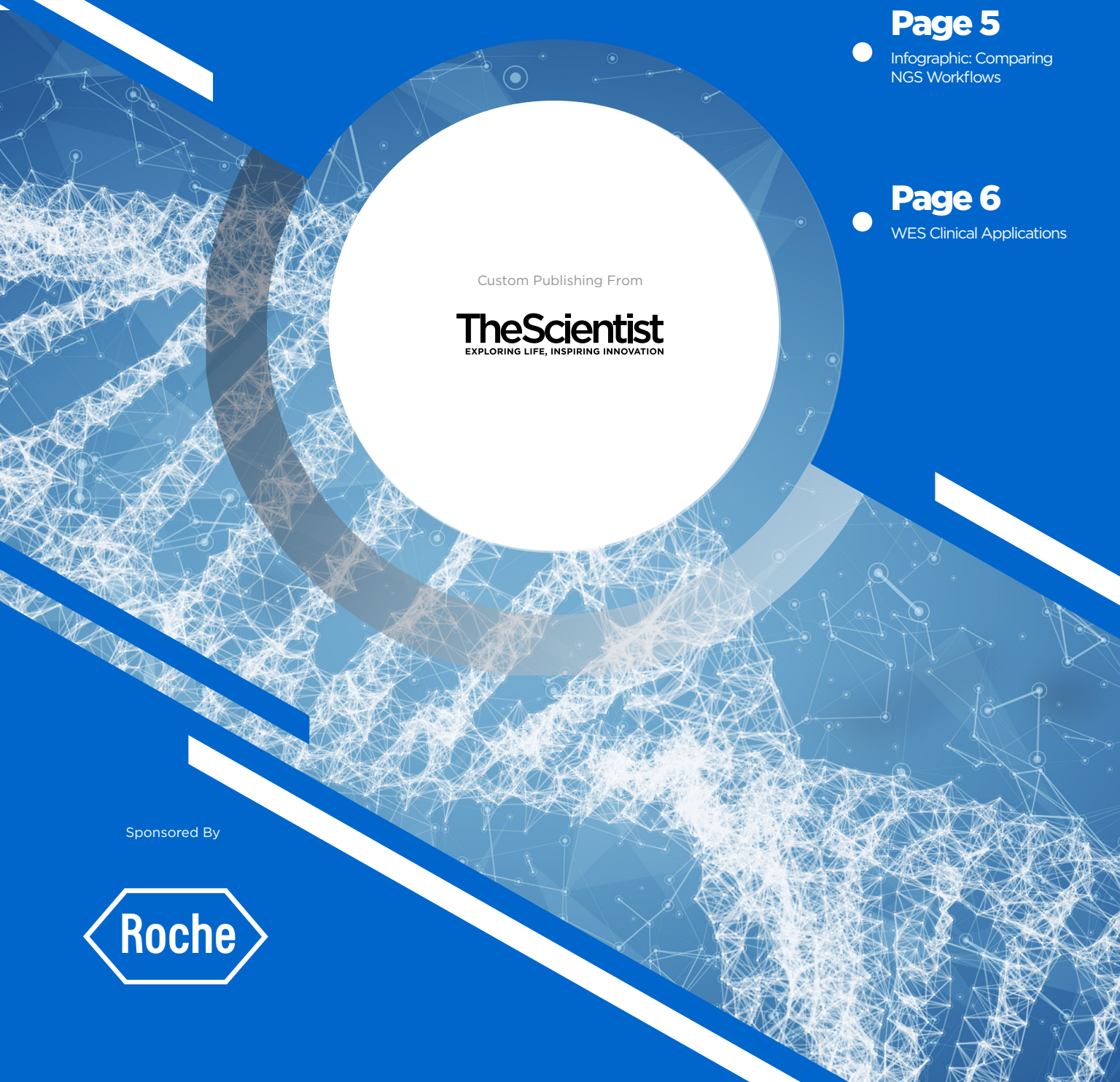
WHOLE-EXOME SEQUENCING: MINING THE CODING REGIONS



Custom Publishing From

TheScientist
EXPLORING LIFE, INSPIRING INNOVATION

Sponsored By

Roche

BETTER BY DESIGN

KAPA Target Enrichment Portfolio for Targeted NGS

Answer your most challenging research questions by leveraging more than a decade of probe-design experience and an improved manufacturing process.

Roche's **NEW KAPA Target Enrichment Portfolio** offers fully customizable target enrichment panels for hybridization-based capture before next-generation sequencing.

Design your custom target enrichment panels by either (1) using our online, user-friendly HyperDesign tool or (2) working directly with one of our expert designers.

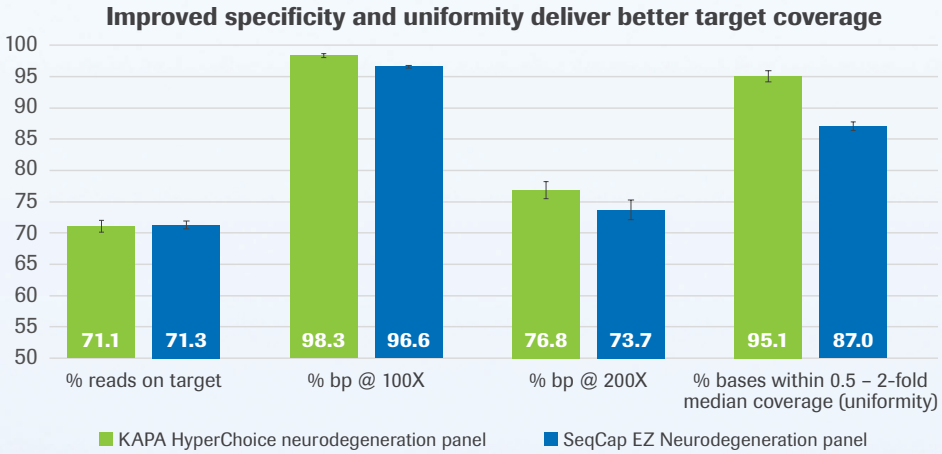
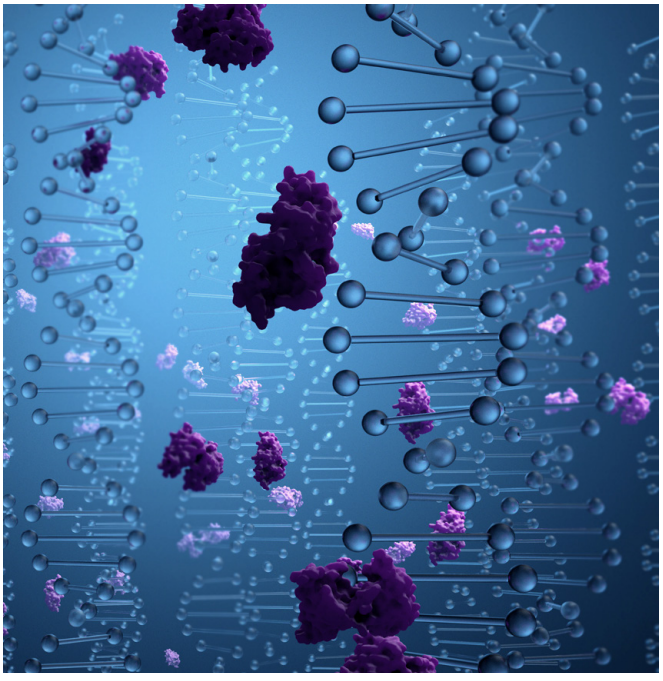


Figure 1. The new KAPA HyperChoice neurodegeneration probe panel outperforms the SeqCap EZ Neurodegeneration panel with higher uniformity and better target coverage, without the need for rebalancing. Performance was compared between the SeqCap EZ Neurodegeneration panel and a new KAPA HyperChoice panel covering similar regions. For each panel, 8 target-enriched DNA libraries were prepared from Coriell control DNA using the appropriate workflow: KAPA HyperCap v2 (SeqCap EZ) or KAPA HyperCap v3 (KAPA HyperChoice). Libraries were captured in 8-plex reactions and sequenced on an Illumina MiSeq instrument (2 x 100 bp). The KAPA HyperChoice panel was used out-of-the-box, while the SeqCap EZ panel was empirically rebalanced to improve performance.

Learn more and begin designing your custom targeted panel at: go.roche.com/HyperChoice

WES vs. WGS vs. Custom Panels

The human genome contains ~3 billion base pairs, approximately 1-5% of which are translated into functional proteins. Mutations in these proteins are the most likely to result in a direct phenotypic consequence. Although whole-genome sequencing (WGS) provides rich information about single nucleotide, structural, or copy number variants, whole-exome sequencing (WES) often makes more sense when time or resources are limited. For scientists looking at specific mutations or genes associated with particular diseases, custom-designed targeted panels offer even greater precision.



Whole-Genome Sequencing

Whole-genome sequencing determines the order of all nucleotides in an individual's DNA and can uncover variation in any part of the human genome, including coding, non-coding, and mitochondrial DNA (mtDNA) regions. In some instances, WGS is the better option because DNA variations outside protein-coding regions can affect gene activity and protein production, potentially leading to genetic disorders. However, WGS requires more sequencing reagents and produces very large datasets that require sophisticated bioinformatics expertise to decipher, increasing both the cost and time required for analysis.

Whole-Exome Sequencing

Whole-exome sequencing focuses on the genomic protein-coding regions (exons). Although WES requires additional reagents (probes) and some additional steps (hybridization), it is a cost-effective, widely used NGS method that requires fewer sequencing reagents and takes less time to perform bioinformatic analysis compared to WGS. Although the human exome represents only 1-5% of the genome, it contains approximately 85% of known disease-related variants.[1] As such, researchers performing WES achieve comprehensive coverage of coding variants such as single nucleotide variants (SNVs) and insertions/deletions (indels). Despite lengthier sample preparation due to the additional target enrichment step, scientists benefit from quicker sequencing and data analysis compared to WGS. WES provides greater sequencing depth for researchers interested in identifying genetic variants for numerous applications, including population genetics, genetic disease research, and cancer studies.

Custom Panels

Scientists sometimes require sequences from specific portions of the human genome, particularly when they are interested in a particular disease or collection of diseases. Custom panels provide this precision without driving up sequencing costs to achieve the required depth of target regions. For example, researchers may use a panel targeted to genes associated with hereditary eye disease as a first-tier test for patients with inherited retinal dystrophy,[2] or for mapping pathogenic variants in breast cancer research.[3] Custom panels are often very small (250 kb-5 Mb) thus bring down sequencing requirements drastically; answering distinct scientific questions becomes a lot easier. Although they are not suitable for broad discovery research, custom panels maximize sequencing economy for specific, suitable applications, especially when researchers use optimized panels for a particular disease or disorder.

References

1. E.L. van Dijk et al., "Ten years of next-generation sequencing technology," *Trends Genet*, 30:418-26, 2014.
2. L. Wang et al., "Application of whole exome and targeted panel sequencing in the clinical molecular diagnosis of 319 Chinese families with inherited retinal dystrophy and comparison study," *Genes (Basel)*, 9(7):360, 2018.
3. E.M. Miller et al., "Development and validation of a targeted next generation DNA sequencing panel outperforming whole exome sequencing for the identification of clinically relevant genetic variants," *Oncotarget*, 8(60):102033-45, 2017.

Overcoming Sequencing Challenges with WES and Custom Panels

As genomic DNA samples can often be collected only once, researchers cannot afford to repeat sequencing time and time again to achieve good coverage. For example, formalin-fixed, paraffin-embedded (FFPE) tissues—generally collected from cancer patients for histopathological diagnosis—are difficult to extract good quality DNA from due to specimen processing protocols.[1] Researchers may only have one chance at obtaining good NGS results from these samples. However, certain genomic regions are notoriously difficult to sequence, which is a cause for concern when samples are limited. These difficult regions can result in non-uniform coverage of target exome regions. New tools and techniques, such as probe and panel optimization, are now available to overcome challenges associated with regions that are problematic for sequencing.

Sequencing Difficulties

During NGS library preparation, DNA molecules are fragmented, ligated to adapters suitable for the particular sequencer used, size selected, and amplified using PCR. Many enzymatic steps within library construction protocols have the potential to introduce sample composition bias. A likely source of bias is the PCR amplification step because amplification is not uniform among fragments. GC-rich or AT-rich fragments are not amplified as efficiently as other fragments, leading to notable inaccuracies in sequencing results over several rounds of amplification.

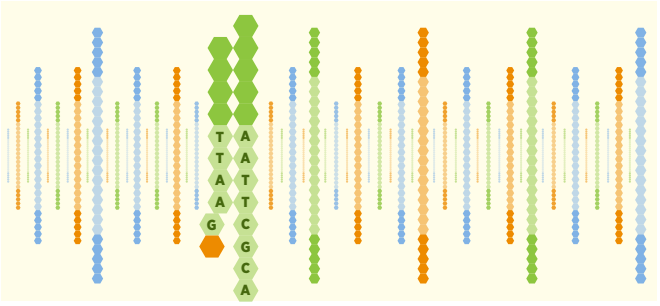
The human genome reference assembly (GRCh38/hg38 release) contains approximately 300,000 single nucleotide polymorphisms (SNPs) in regions with ≥75% GC content, leaving many potentially interesting variants difficult to access with NGS.[2,3]

Hybridization-based target enrichment of DNA sequencing libraries can enable sequencing of these difficult regions, especially when well-designed, high-quality probe panels are used.

Panel Selection

Whether you are performing WES, using some other predesigned panel, or using a custom probe panel, correct panel selection is critical for successfully accessing difficult genomic regions, for obtaining uniform capture, and for increasing fold-enrichment of target sequences.

Several specialized hybridization probe panels are available to help researchers overcome sequencing challenges. These include probe panels that have been extensively optimized for the latest release of the human genome assembly (GRCh38/hg38) as well as panels for model organisms such as mouse (GRCm39) and zebrafish (GRCz11). Optimized panels enable researchers to increase sequencing throughput, achieve better coverage compared to standard panels, and improve uniformity across GC-rich regions.



Probe Optimization

Tools that enable researchers to create custom target enrichment designs are also now available. These take advantage of online optimization algorithms. Scientists can enter such information as gene names, sequence identifiers, and genomic coordinates, which are then processed through various algorithms to output custom-designed probe sequences.

Custom-designed and optimized probes allow researchers to improve target enrichment panels to meet unique research needs, such as for specific cancers or rare disease research. Optimized probes and panels enable scientists to perform research more efficiently and to discover variants more easily.

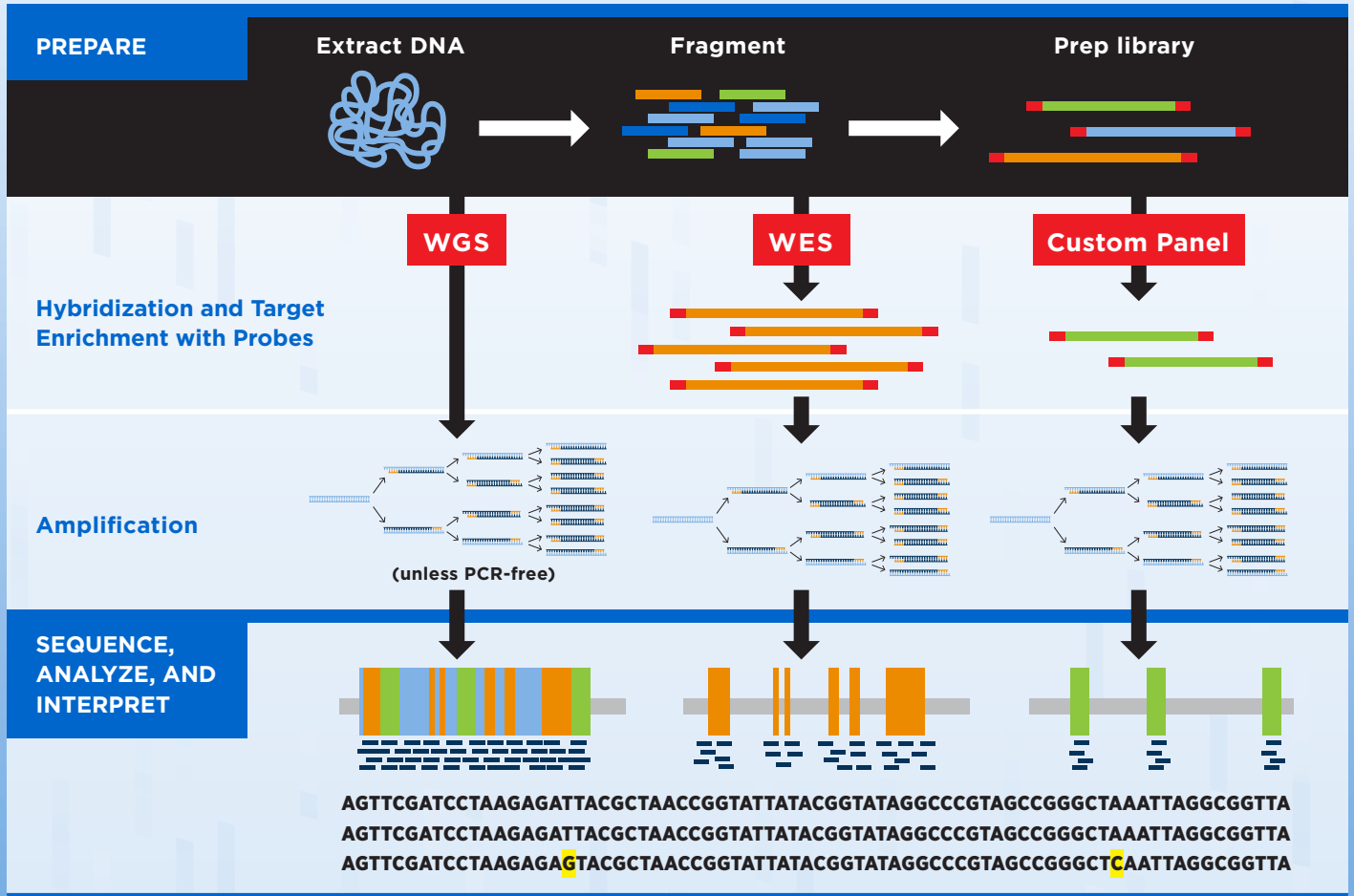
By using well-designed exome panels for WES, or custom panels designed to target their regions of interest, scientists can obtain proficient, robust, and automatable workflows, even for difficult-to-sequence areas of the genome. Better uniformity in sequencing means that scientists can increase sequencing depth and throughput, which also reduces the amount of resources required to uncover variants relevant to their research questions.

References

1. P. Robbe et al., "Clinical whole-genome sequencing from routine formalin-fixed, paraffin-embedded specimens: pilot study for the 100,000 Genomes Project," *Genet Med*, 20:1196–1205, 2018.
2. A. Mowjoodi et al., "Discrimination of SNPs in GC-rich regions using a modified hydrolysis probe chemistry protocol," *Biotechniques*, 57:313-16, 2014.
3. Y. Guo et al., "Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis," *Genomics*, 109(2):83-90, 2017.

WGS, WES, or A Custom Panel?

WGS, WES, and custom panels are all great options for mapping variants. However, each technique differs in cost, time, and sequencing capability, making each suitable for different research purposes.



	WGS	WES	CUSTOM PANEL
% total known disease-related mutations potentially identified	100%	>85%	Depends upon panel design.
Provides information about	All types of variation, known and unknown: SNPs, indels, structural variants. Includes coding regions and noncoding regions.	Known/unkown SNPs/indels within protein coding regions	Known/unkown SNPs/indels in a small focused subset
% genome sequenced	~100%	1-5%	Depends upon panel design
Sequencing resources required	Higher sequencing costs Greater analysis time and expense Greater requirements for data storage	Requires probes (exome panel) Additional steps (hybridization etc.) Reduced sequencing costs Less time and expense for analysis Less requirement for data storage	Requires probe design Requires probes Additional steps (hybridization etc.) Very-much-reduced sequencing costs (for small panels) for deeper coverage Less time and expense for analysis Less requirement for data storage
Best suited for	Novel variant discovery, genome-wide association studies, regulatory and intergenic regions, structural variants, sequencing of genomes that are not well-annotated	Identification of variants in protein-coding regions (and, with some exome panels, well-annotated regulatory regions)	Focus on specific areas of interest to achieve greater sequencing depth or sequence a large number of samples
Challenges	Human genome is 3 billion bases at 1x. Very expensive to sequence and achieves less depth	Misses noncoding mutations and any large structural rearrangements	Need to know what regions are most critical for your research Requires probe design expertise

WES Clinical Applications

Researchers who once relied on WGS for reliable variant information are now turning to WES for its faster turnaround time and cost-effectiveness. Considerable evidence is emerging that applying WES in clinical research settings will lead to improved diagnosis and, in some cases, treatment of genetic disease. WES may improve patient health outcomes and facilitate the more efficient use of healthcare resources.

Mendelian Disorders

Exome sequencing has revolutionized Mendelian disorder research. The first report of selectively sequencing a whole exome was published in 2009 by Sarah Ng and her colleagues at the University of Washington.[1] The research group reported the targeted capture and massively parallel sequencing of the exomes from 12 people, four of which were affected by a rare, dominantly-inherited disorder called Freeman Sheldon syndrome (FSS), which is caused by mutations in the *MYH3* gene. FSS is characterized by multiple contractures at birth, head and face abnormalities, hand and foot defects, and skeletal malformations. Although the genetic defect behind the disease was already known, the research defined the *MYH3* gene as disease-causing in FSS patients from more than 300 million bases of DNA noise, and demonstrated proof-of-concept for WES as a tool for studying Mendelian disorders using a small number of unrelated, affected individuals.[1]

Depending on the exome panel used, WES can provide coverage of more than 95% of exons and contain 85% of the known mutations resulting in Mendelian disorders and many of the disease-predisposing SNPs that occur throughout the genome.[2] Variants detected cumulatively from WES studies are used widely in clinical services, leading to more accurate genotype-phenotype correlations and new insights into the role of rare genomic variation in disease. Online Mendelian Inheritance in Man®, an online catalog of human genes and genetic disorders (OMIM.org), has recorded a steady increase in both the number of phenotypes with an identified genetic etiology, and the number of genes associated with a clinical phenotype. This research, along with other worldwide efforts, has elucidated the molecular and genomic architecture of Mendelian conditions. The broader availability of exome sequencing has supported these discoveries.[3]

Diagnostics

The first diagnosis established by WES was published by Murim Choi and his colleagues from the Yale University School of Medicine in 2009.[4] The team used WES to study a patient referred for Bartter syndrome, a rare inherited disorder characterized by a defect in the thick ascending limb of the loop of Henle. They showed that the patient carried a novel homozygous mutation in the *SLC26A3* gene, which is associated with congenital chloride-losing diarrhea (CLD). In this case, WES established the correct diagnosis as CLD.

Since then, WES has become common in diagnosing rare diseases. For example, a study by a research team led by Alexander Hoischen at Radboud University Medical Center showed that exome sequencing could be used as a genetic test to diagnose primary immunodeficiencies. The researchers tested 254 patients with primary immunodeficiencies and showed that the techniques granted a diagnosis for 28% of patients. They concluded that exome sequencing harbored an advantage over gene panels as a truly generic test for all genetic diseases, including in silico extension of existing gene lists and re-analysis of existing data.[5]

Other diseases diagnosed using WES include a new ocular variant associated with Leber congenital amaurosis (*PEX1* gene mutation),[6] oculocutaneous albinism (*SLC45A2* gene mutation),[7] and neutropenia (*G9PC3* gene mutation).[7]

Cancer Research

WES has potential to provide insight into cancer mechanisms because exome sequence variation may influence the predisposition for cancer development. Although smaller targeted panels are valuable for cancer research and diagnostics because of the depth of sequencing that they enable, WES is more suitable for discovering specific regions to further investigate.

WES is ideal for analyzing formalin-fixed paraffin-embedded (FFPE) samples, which are frequently used for storing tumor biopsies, where limited DNA yields are common.[8]

For example, a team led by Edward Generozov at the Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency used WES to sequence DNA extracted from FFPE tumor samples to better understand the mechanisms of prostate cancer pathogenesis, which could serve as a basis for developing new therapeutic approaches.[9]

Although exome sequencing cannot identify structural and noncoding variants across the entire genome as WGS can, it allows at least 20 times as many samples to be sequenced over the same time frame. In studies focused on identifying rare variants or somatic mutations with medical relevance, sample size and the interpretability of functional impact are critical to success.

References

1. S.B. Ng et al., "Targeted capture and massively parallel sequencing of 12 human exomes," *Nature*, 461(7261):272-6, 2009.
2. B. Rabbani et al., "The promise of whole-exome sequencing in medical genetics," *J Hum Genet*, 59:5–15, 2014.
3. Y.A. Barbitoff et al., "Systematic dissection of biases in whole-exome and whole-genome sequencing reveals major determinants of coding sequence coverage," *Sci Rep*,10:2057, 2020.
4. M. Choi et al., "Genetic diagnosis by whole exome capture and massively parallel DNA sequencing," *Proc Natl Acad Sci USA*, 106(45):19096-101, 2009.
5. P. Arts et al., "Exome sequencing in routine diagnostics: a generic test for 254 patients with primary immunodeficiencies," *Genome Med*, 11:38, 2019.
6. J. Majewski et al., "A new ocular phenotype associated with an unexpected but known systemic disorder and mutation: novel use of genomic diagnostics and exome sequencing," *J Med Genet*, 48:593–96, 2011.
7. B. Meder et al., "Targeted next-generation sequencing for the molecular genetic diagnostics of cardiomyopathies," *Circ Cardiovasc Genet*, 4:110–22, 2011.
8. S.J. McDonough et al., "Use of FFPE-derived DNA in next generation sequencing: DNA extraction methods," *PLoS ONE*,14(4): e0211400, 2019.
9. A.S. Nikitina et al., "Data on somatic mutations obtained by whole exome sequencing of FFPE tissue samples from Russian patients with prostate cancer," *Data Brief*, 25:104022, 2019.

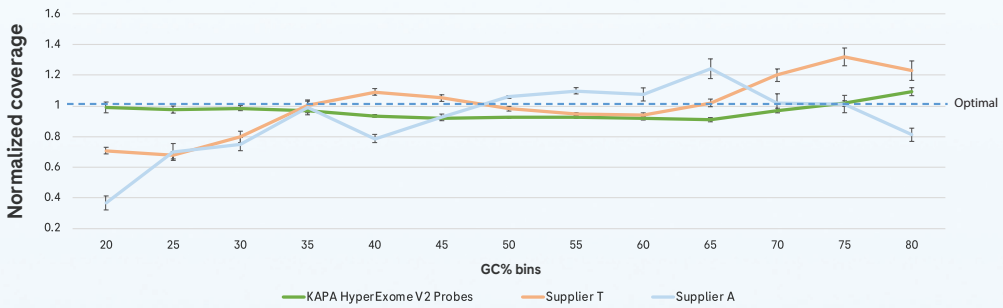
BETTER BY DESIGN

KAPA HyperExome V2 Probes for Whole-Exome Sequencing

Efficient whole-exome sequencing (WES) begins with expertly designed probes that effectively capture challenging genomic regions.

- Reliably enrich challenging, previously inaccessible exonic regions
- Unlock unique insights in translational research by capturing more content from key genomic databases
- Achieve higher uniformity across the entire range of %GC content

Exceptional uniformity of normalized capture coverage with KAPA HyperExome V2 Probes



The blue dashed line represents the optimal uniformity in the ideal state in which all regions—regardless of their GC content—would be equally covered. The normalized coverage per GC% bins of KAPA HyperExome V2 Probes were compared to those of suppliers T and A using each supplier's workflow.

Request a sample or evaluation at: go.roche.com/HyperExomeV2

Project: NGHC, Roche Sequencing Solutions, Pleasanton, CA 2019
For Research Use Only. Not for use in diagnostic procedures.

MC-US-08247 A826 08/24