



MILESTONE 9

Exploring whole exomes

Historically, pinpointing the genetic basis of a monogenic disease relied on the ability to first delineate a specific chromosomal region within the vast genomic search space where the causal variation resides. The reason for this bottleneck was the lack of a suitable technology to allow researchers to efficiently interrogate the DNA sequences of all protein-coding genes simultaneously. Instead, gene hunters applied tools such as linkage analysis in large pedigrees to identify genetic markers that co-segregated with the disease phenotype, thereby reducing their search space to a tractable number of candidate genes within a defined genomic interval. The development of whole-exome sequencing was a key breakthrough that removed this bottleneck, allowing researchers to identify disease-causing mutations without any prior knowledge of the chromosomal location or biological role of the causal gene.

The technological advance that laid the essential groundwork for whole-exome sequencing was the adaptation of microarrays to perform targeted capture of exon

sequences from genomic DNA before high-throughput sequencing. This capture step yields an enrichment of exons relative to other regions of the genome, allowing for high-depth sequence coverage of protein-coding segments at a fraction of the cost of sequencing entire genomes.

In 2009, Sarah Ng and colleagues at the University of Washington reported the first successful application of whole-exome sequencing to monogenic disease. In this landmark proof-of-principle study, the authors performed targeted capture and massively parallel sequencing of whole exomes from eight control samples and four individuals with Freeman–Sheldon syndrome, a rare autosomal dominant disorder known to be caused by mutations in *MYH3*. After filtering out known common variants as well as variants observed among the eight control samples, *MYH3* emerged from their analysis as the only gene with at least one non-synonymous coding variant or splice-site disruption in all four affected individuals. Thus, this study established both a working technology and an analytical

“ The ability to interrogate all protein-coding regions at high sequencing depth in a cost-efficient way has dramatically accelerated the pace of human disease gene discovery

”

framework to examine whole exomes for disease-causing mutations in a robust and cost-effective way.

To explore the potential of this strategy to elucidate the genetic basis of a monogenic disorder of unknown cause, Ng and colleagues next applied their whole-exome sequencing workflow to Miller syndrome, a recessive disorder that had been refractory to standard genetic approaches. By performing whole-exome sequencing of four affected individuals from three independent families and filtering out variants observed in population samples, they identified a single gene, *DHODH*, that was disrupted by rare recessive variants in all affected individuals. To validate this discovery, they analysed *DHODH* using Sanger sequencing in three additional individuals with Miller syndrome and again found biallelic *DHODH* mutations in all three cases, illustrating the power of whole-exome sequencing for monogenic disease gene discovery. In a third study published later that year, Ng and colleagues successfully used whole-exome sequencing to identify mutations in *MLL2* (also known as *KMT2D*) as a major cause of Kabuki syndrome, an autosomal dominant developmental disorder.

Although these early applications of whole-exome sequencing focused on monogenic diseases, this technology has had a transformative impact on many areas of human disease genetics, including cancer genomics and common diseases with complex genetic inheritance. The ability to interrogate all protein-coding regions at high sequencing depth in a cost-efficient way has dramatically accelerated the pace of human disease gene discovery, particularly for rare diseases, and is poised to yield important insights into the genetics of more common diseases as whole-exome sequencing is applied at scale to very large population samples.

Kyle Vogan, *Nature Genetics*

ORIGINAL ARTICLE Ng, S. B. et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–276 (2009)

FURTHER READING Albert, T. J. et al. Direct selection of human genomic loci by microarray hybridization. *Nat. Methods* **4**, 903–905 (2007) | Okou, D. T. et al. Microarray-based genomic selection for high-throughput resequencing. *Nat. Methods* **4**, 907–909 (2007) | Porreca, G. J. et al. Multiplex amplification of large sets of human exons. *Nat. Methods* **4**, 931–936 (2007) | Hodges, E. et al. Genome-wide in situ exon capture for selective resequencing. *Nat. Genet.* **39**, 1522–1527 (2007) | Ng, S. B. et al. Exome sequencing identifies the cause of a Mendelian disorder. *Nat. Genet.* **42**, 30–35 (2010) | Ng, S. B. et al. Exome sequencing identifies *MLL2* mutations as a cause of Kabuki syndrome. *Nat. Genet.* **42**, 790–793 (2010)