

Performance and Applications of Helix's Exome+ Assay

Helix's Exome+[®] assay is run on all samples sequenced at Helix. This single assay combines a clinical exome with uniform panel-grade tests, a genome-wide backbone, and full mitochondrial coverage. Post sequencing, rigorous quality checks are applied to ensure that the data generated are appropriate for a broad application of clinical genetic tests. These tests range from population health screens, to diagnostic or pharmacogenomic panels to whole exome sequencing. Meanwhile, the Exome+ backbone of whole genome single nucleotide polymorphisms (SNPs), augmented with imputation, enables the generation of ancestry, traits, and polygenic risk scores. By ensuring that this single assay is appropriate for answering a wide list of clinical questions, the Exome+ assay establishes a foundation for Sequence Once, Query Often[®]. Sequence Once, Query Often is a process that allows various genetic test results to be delivered for an individual over the course of their life based on when each might be relevant or informative.

In building a test menu relying on the Exome+ assay, we have focused on three important pillars: quality, breadth of testing, and turnaround time, as discussed in the following sections.

Quality of Testing

Coverage and Callability

Over the past decade, Helix has optimized the Exome+ assay for uniformity of coverage, and target region completeness in order to maximize scalability and utility of genetic testing. Our assay focuses on callability (percent of targeted bases that produce a reliable genotype call) and CNV sensitivity by balancing depth of coverage & mapping/alignment quality metrics down to the individual base level. For all diagnostic panels, we guarantee a callability rate exceeding 99.8%. Performance is delivered efficiently through a tiered coverage approach, resulting in ~140x coverage across cancer genes, ~85x coverage across cardio genes, and > 500x coverage across the mitochondrial genome.

Small Variants and Structural Variants

Small variants and copy number variants (CNVs) are generated across the full exome, with performance of these variants captured in **Table 1**. Small variants are called using a customized

version of Sentieon while all other variant types are called using proprietary algorithms, including CNVs which are detected by both read depth and split reads. For variants that are more challenging—for example, due to interference by pseudogenes (e.g. PMS2 terminal exons) or special variant types (e.g. the MSH2 inversion also known as the Boland Inversion)—specialized informatics are applied and a special validation is performed.

Small variant results are based on whole-exome data from five National Institute of Standards and Technology (NIST) cell lines. CNV results are based on 36 known single-exon CNVs and 73 known multi-exon CNVs. The Boland inversion detection rate is measured against 3 samples run in duplicate. The BRCA2 Alu insertion detection rate is measured against 3 samples across 11 replicates. Detection rates for PMS2 small variants were measured against 13 samples across 45 replicates. Detection rates for PMS2 CNVs were measured against 24 samples across 40 replicates.

Table 1: Summary performance metrics for variants on the Helix Exome+ assay based on repeated analytical validations.

Variant Type	Sensitivity	Specificity
SNV	99.98%	All variants are confirmed prior to reporting
Deletion ≤ 20 bp	99.94%	
Insertion ≤ 20 bp	99.75%	
MNV	99.74%	
Complex	99.17%	
Substitution	~ 100%	
Single-exon CNVs	≥ 90%	
Multi-exon CNVs	100%	
MSH2:c.942+3A>T	100%	
MSH2 Boland inversion	100%	
BRCA2 Alu insertion	100%	
PMS2 small variants, exons 11-15*	100%	
PMS2 CNVs, exons 11-15*	100%	

SNV = Single nucleotide variants. MNV = multi-nucleotide variant, or phased variants. Substitutions = an insertion and deletion with different lengths sharing the same location and strand. Complex = two different variant types sharing the same location but mapping to a different allele.

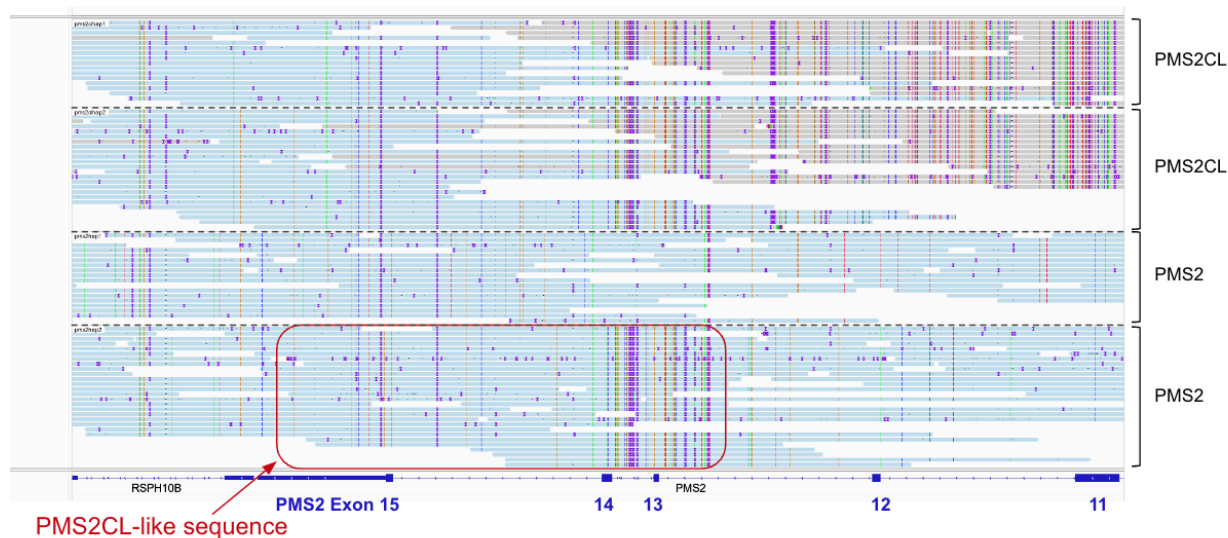
** PMS2 terminal exons are included in all diagnostic tests but not included in the Tier One Population Screen and other screening tests.*

Additionally, we conducted validation studies to compare Exome+ small variants against Sanger sequencing. Analysis included 1,236 samples (1,141 sourced from saliva and 95 from cell lines) with 1,251 variants and 172,711 reference sites, and demonstrated > 99.9% concordance between Exome+ and Sanger.

Variant Confirmation

All variants are confirmed before reporting. Confirmation methods include visual inspection of the reads in simple cases such as SNVs with clear allele fractions and no concerning genomic context, short read whole genome sequencing for more complicated small variants and CNVs, and long read sequencing when appropriate. All PMS2 variants in terminal exons are confirmed by PacBio long reads. This allows us to confidently determine which variants are in PMS2 and which are in PMS2CL, and also to identify gene conversion events such as the one seen in **Figure 1**.

Figure 1: Screenshot of a PacBio run in which a saliva sample has been flagged for confirmation using the Exome+ assay. A gene conversion event was identified in a saliva sample using PacBio long-read sequencing. In the second PMS2 allele, exons 1-12 were present as usual, but were combined with PMS2CL exons 4-6 in place of PMS2 exons 13-15.



Pharmacogenomics

Whole exome sequencing is performed on every sample received for clinical genetic testing. Therefore, pharmacogenomic testing is based on whole gene results and not limited to targeted star alleles. For all genes with star alleles, all known star alleles are determined regardless of how common or rare they may be, and novel alleles are flagged. This ensures that the most accurate metabolizer status can be reported. Additionally, special informatics have been applied to CYP2D6 to ensure the reliable detection of star alleles representing gene fusions and copy numbers up to 5 copies. Both accuracy and range of results for pharmacogenomic genes is captured in **Table 2**, with results based on > 300 data sets from Genetic Testing Reference Material Coordination Program (GeT-RM) samples.

Table 2: Summary of reportable and analytical range and accuracy for Pharmacogenomic genes with star alleles on the Exome+ assay. Not listed, but included in our tests, are individual SNPs such as for VKORC1.

Gene	Star Alleles	Accuracy	Copy Number	Novel Alleles
CYP2C9	*1-*61	100%	✓	✓
CYP2C19	*1-*19, *22-*26, *28-*39	100%	✓	✓
CYP2D6	*1-*15, *4N, *17-*65, *68-*75, *81, *83-*114	99.45%^	✓	✓
CYP3A4	*1-*24, *26, *28-*38	100%	✓	✓
CYP3A5	*1, *3, *6-*9	100%	✓	✓
CYP4F2	*1, *3	100%	✓	✓
DPYD	Reference, c.46C>G, c.61C>T, c.62G>A, c.295_298delTCAT (*7), c.313G>A, c.343A>G, c.451A>G, c.496A>G, c.498G>A, c.525G>A, c.557A>G, c.601A>C, c.632A>G, c.703C>T (*8), c.775A>G, c.868A>G, c.929T>C, c.934C>T, c.967G>A, c.1003G>T (*11), c.1024G>A, c.1057C>T, c.1108A>G, c.1129-5923C>G, c.1129-5923C>G, c.1236G>A (HapB3), c.1156G>T (*12), c.1180C>T, c.1181G>T, c.1218G>A, c.1260T>A, c.1278G>T, c.1294G>A, c.1314T>G, c.1349C>T, c.1358C>G, c.1371C>T, c.1403C>A, c.1475C>T, c.1484A>G, c.1519G>A, c.1543G>A, c.1577C>G, c.1601G>A (*4), c.1615G>A, c.1679T>G (*13), c.1682G>T, c.1774C>T, c.1775G>A, c.1777G>A, c.1796T>C, c.1896T>C, c.1898delC (*3), c.1905C>G, c.1905+1G>A (*2A), c.1906A>C, c.1990G>T, c.2021G>A, c.2161G>A, c.2186C>T, c.2194G>A (*6), c.2195T>G, c.2279C>T, c.2303C>A, c.2336C>A, c.2482G>A, c.2582A>G, c.2623A>C, c.2639G>T, c.2656C>T, c.2657G>A (*9B), c.2846A>T, c.2872A>G, c.2915A>G, c.2921A>T, c.2933A>G, c.2977C>T, c.2978T>G, c.2983G>T (*10), c.3049G>A, c.3061G>C, c.3067C>A	100%	✓	✓
HLA-B	*15:02, *57:01	100%	—	—
SLCO1B1	*1-*16, *19, *20, *23-*34, *36-*44, *47-*49	100%	✓	✓

NUDT15	*1-*20	100%	✓	✓
TPMT	*1, *2, *3A, *3B, *3C, *4-*44	100%	✓	✓
UGT1A1	*1, *6, *27, *80 (UGT1A1*80 is used as a proxy for UGT1A1*28)	100%	✓	✓

^ Two incorrect calls were identified in samples with a complex CYP2D6 rearrangement, as follows:

1. Truth = *1/*10/*36/*36; Call = *1/*10/*10/*36, both are Normal Metabolizers
2. Truth = *10/*36/*36; Call = *10/*10/*36, both are Intermediate Metabolizers

Comprehensive PGx tests are necessary to avoid misclassifications

Helix's approach to a comprehensive reportable range for pharmacogenomics is critical to supporting diverse populations. When star alleles are excluded from the reportable range, approximately 16.8% individuals are often mis-classified as *1, resulting in a normal metabolizer classification when the metabolizer status is in fact unknown (**Table 3**)¹. Additionally, gene copy number is important to understanding true metabolizer status, as duplications can quickly increase metabolism. Across a set of 16k individuals, ~ 16% of samples were identified to have whole gene CYP2D6 duplications (with another 5% harboring whole gene deletions), highlighting the importance of analyzing CYP2D6 copy number in conjunction with CYP2D6 star alleles.

Table 3: Impact of using an incomplete CYP2D6 panel. This table presents the outcome of using CYP2D6 tests with an incomplete star allele list as compared to the Helix PGx Pipeline, based on 30,000 individuals. For comparison, a representative panel was used that includes the common alleles: *1, *2, *3, *4, *4N, *5, *6, *9, *10, *17, *29, *35, *36, *41, and duplications.

Consequence of using an incomplete panel	Samples affected
Mischaracterization of ultrarapid metabolizers	7.7%
Mischaracterization of poor metabolizers	4.4%

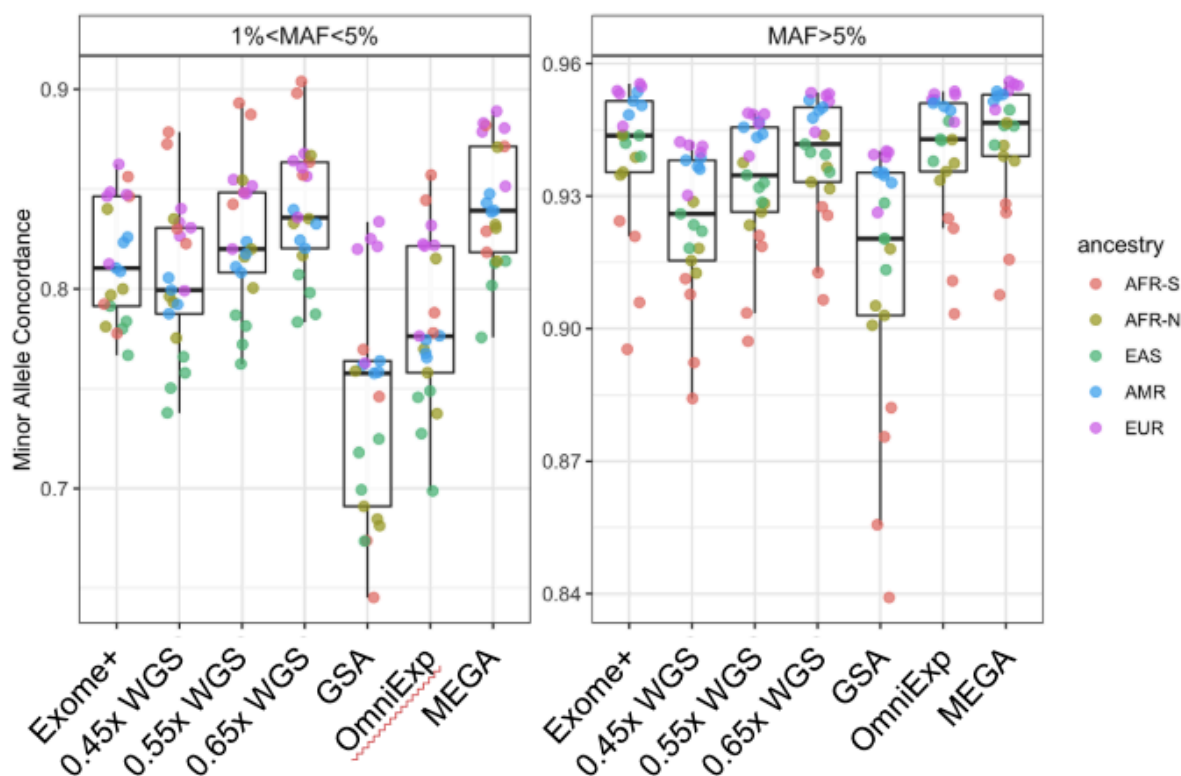
Helix reports all star alleles within the reportable range and includes assessment and reporting of gene copy number and novel star alleles to ensure the most accurate results for all populations.

Imputation

Both ancestry and polygenic risk scores can be easily and automatically generated from Exome+ results, leveraging a backbone of SNPs across the full genome that complements the exome itself. As a result, Helix is able to offer robust genome-wide imputation with tens of millions of high-confidence SNPs imputed with technical equivalence to ~0.6x WGS, Infinium OmniExpress, and Infinium MEGA for common alleles (**Figure 2**). For rare alleles, the Exome+ results were approximately equivalent to 0.5x

WGS, with improved performance over GSA and OmniExp (**Figure 2**). This strong performance is possible because imputation is not limited to the hundreds of thousands of SNPs targeted in the Exome+ assay, but instead leverages all directly sequenced data, including the flanking regions of those SNPs and the full exome.

Figure 2: Demonstration of minor allele concordance between imputed variants and directly sequenced variants across five ancestral populations. Results are presented for typical Exome+ runs, WGS at different coverage levels (0.45x, 0.55x, 0.65x), and three microarrays (GSA, OmniExp, and MEGA). AFR-S: Sub-saharan African. AFR-N: North African. EAS: East Asian. AMR: Indigenous American. EUR: European.



Breadth of Testing

Helix's Exome+ assay has been designed to be used for a wide range of testing categories and additional informatic tools or procedures to ensure fast turnaround times are implemented according to the tests' specific needs (**Table 4**). This enables Helix to offer a variety of high-quality efficient testing options based on a single sample and sequencing.

Table 4: Comparison of Helix Test Types and Characteristics

	Dx	PGx	Carrier	Screening
Examples	Helix Multi-Cancer Panel	Helix Clopidogrel Test	Helix Core Carrier Screen	Helix Tier One Population Screen
VUS reported?	Yes	n/a	No	No
Completeness	<ul style="list-style-type: none"> • ≥ 99.8% callability • Includes full PMS2 • Includes complicated variants as described in the test definitions 	<ul style="list-style-type: none"> • Based on full star alleles in PharmVar • Includes gene-level copy number • Includes novel alleles 	<ul style="list-style-type: none"> • > 98% sensitivity including small variants and CNVs • Special solutions for challenging genes 	≥ 99.4% callability
First Use TAT	Typically 6 to 21 days (6 to 11 days for STAT)	Typically 6 to 9 days	Typically 7 to 14 days	Typically 1 to 4 weeks
Requery TAT	Typically ≤ 5 days (≤ 2 days for STAT)	Typically ≤ 5 min	Not supported yet	Typically ≤ 21 days

Polygenic Risk Scores

In addition to developing panel-grade products, we also have the capability to impute tens of millions of common genotypes in support of polygenic risk score applications, which are intended to be added as complements to our diagnostic tests in the near future. Polygenic Risk Scores (PRS)

are emerging as a powerful tool to predict an individual's risk of disease. PRS can be used to predict disease risk as effectively as variants within single genes ².

Because PRS are based on many genomic loci, often focused in non-coding regions, their coverage requirements are significantly different from those for standard genetic testing of monogenic disease. The Exome+ assay supports analysis of PRS with direct coverage of hundreds of thousands of non-coding variants plus tens of millions of genome-wide imputed variants. These millions of variant calls ensure that PRS of all sizes are supported.

Ancestry

Ancestry can help provide context for the relevance of clinical genetic testing, for example when certain populations have an increased concentration of carriers of specific risk alleles. The Exome+ assay results in ancestry assignments spanning 26 ancestries across eight regions (**Table 5**).

Table 5: List of 26 populations, grouped into eight regions, calculated by Helix.

Region	Populations	Region	Populations
European	Ashkenazi Jewish Finnish Northeast European Northwest European Southeast European Southwest European Sardinian	East Asian	Filipino Japanese Korean Northern Han Chinese Southeast Asian
African	East Bantu Nigerian Senegalese and Gambian Sierra Leonean and Liberian	Indigenous American	Indigenous North American Indigenous South American
South Asian	Gujarati Punjabi and Pathan South Indian	Middle Eastern & North African	Bedouin North African Persian
Arctic	Beringian	Oceanian	Melanesian

Turnaround time

The Exome+ assay offers the basis for Sequence Once, Query Often processes, where the data generated upon initial testing can be used during additional queries. The turnaround time for Helix tests will be significantly reduced for patients who have already been sequenced at Helix. **Table 4** summarizes typical turnaround time based on different test types.

Diagnostic and pharmacogenomic tests are prioritized within the Helix lab in order to meet quick turnaround times. After a provider orders a genetic test, Helix performs whole exome sequencing and then interprets only the specific genes within the test ordered. Helix then continues to store the patient's genetic information for future clinical use. With the patient's permission, their health care provider can order future medically necessary genetic tests from Helix without the need to submit another sample (in most cases). Tests that can be fully automated through digital analysis of a patient's genetic information, such as PGx tests, can typically be returned in under five minutes.

References

1. [A comprehensive landscape of CYP2D6 variation across 30,000 individuals](#)
2. Cirulli and Washington, [Researchers have access to new data on thousands of exomes](#). Here's what we found. March 27, 2019.