Applications for the Helix Exome+™ Assay

Helix's proprietary Exome+ assay combines a highly uniform panel-grade* exome with over 300,000 highly informative non-coding regions and is designed specifically to enable both current and future clinical and research applications, with many such examples described below. For example, the Exome+ assay provides ≥ 99.5% base by base coverage at ≥ 20x for the ~ 600 genes most relevant to clinical applications (hereditary cancer, cardiovascular disease, pharmacogenomics, and carrier screening). The same Exome+ assay also supports imputation of tens of millions of common genotypes for clinical polygenic risk score applications.

The Helix Exome+ assay is run exclusively at the Helix's CLIA and CAP accredited laboratory facility in San Diego, CA (CLIA #05D2117342, CAP #9382893). The Helix Laboratory is a highly automated facility with the ability to process millions of Exome+ assays annually. This paper provides a quick data reference for common Exome+ clinical applications. For custom or more detailed queries, please contact the Helix team directly. The data itself is of high quality, as demonstrated in Table 1[1, 2].

Table 1: Summary performance metrics for Helix exome+ assay data outputs.

Gene	Sensitivity	Specificity
Single Nucleotide Variants	99.5%	99.9999995%
Copy Number Variants	100% for > 2 exons	99.998%
CYP2D6 star alleles	99.8%	100%

The applications of Helix Exome+ assay allow for high quality panel grade tests for medically actionable conditions like hereditary cancer, cardiovascular, or carrier screening conditions. Our partners can also design their own panels based on the condition of interest. This assay also supports analysis of difficult-to-sequence genes such as CYP2D6 for pharmacogenomics and robust data for developing and implementing polygenic risk scores in multiple-ethnicities. We provide coverage checks and share information on the quality of the data to help you choose the best test for your patients.

Below are some examples on performance data for some of our panels. Please reach out to the Helix team if you have questions about specific panels.

Medically Actionable Conditions

The American College of Medical Genetics and Genomics (ACMG) has identified a set of 59 genes that are medically actionable for conditions such as certain cancers, specific cardiovascular diseases, and metabolic disorders including malignant hyperthermia susceptibility [11]. Genetic testing of these genes can identify predisposition to cancer, metabolic, or cardiovascular disease. Helix's *Exome+* assay offers a 99.911% variant call rate (the percentage of bases with a high-quality variant call across 4,000 production datasets) across these 59 medically actionable genes, with gene-level call rates described in Table 5. Additionally, CNV detection is 100% for events spanning > 2 exons.

Table 5: Helix Exome+ assay performance metrics for 59 medically actionable genes.

Variant Call Rate Per Gene							
ACTA2	100.00%	KCNH2	99.98%	OTC	100.00%	SMAD3	100.00%
ACTC1	100.00%	KCNQ1	100.00%	PCSK9	100.00%	SMAD4	100.00%
APC	100.00%	LDLR	99.96%	PKP2	100.00%	STK11	100.00%
APOB	96.94%	LMNA	100.00%	PMS2*	99.98%	TGFBR1	100.00%
ATP7B	100.00%	MEN1	99.93%	PRKAG2	100.00%	TGFBR2	100.00%
BMPR1A	100.00%	MLH1	100.00%	PTEN	99.89%	TMEM43	100.00%
BRCA1	100.00%	MSH2	100.00%	RB1	100.00%	TNNI3	100.00%
BRCA2	100.00%	MSH6	99.98%	RET	100.00%	TNNT2	100.00%
CACNA1S	100.00%	MUTYH	100.00%	RYR1	100.00%	TP53	100.00%
COL3A1	100.00%	MYBPC3	100.00%	RYR2	100.00%	TPM1	100.00%
DSC2	100.00%	MYH11	100.00%	SCN5A	99.99%	TSC1	100.00%
DSG2	100.00%	MYH7	100.00%	SDHAF2	100.00%	TSC2	100.00%
DSP	100.00%	MYL2	100.00%	SDHB	100.00%	VHL	100.00%
FBN1	100.00%	MYL3	100.00%	SDHC	99.82%	WT1	99.98%
GLA	100.00%	NF2	100.00%	SDHD	100.00%		

* PMS2 exons 12-15 are not included due to the inability to detect gene recombination events, which impact the relevance of PMS2 variants in these exons.

CDC Tier 1

The CDC Office of Public Health Genomics (OPHG) has defined Tier 1 (T1) genomics applications as those with significant potential for making positive improvements in public health according to evidence-based guidelines and recommendations [3]. The conditions chosen as CDCT1 have proven clinical utility and validity. These include Familial Hypercholesterolemia (FH), Hereditary Breast and

Ovarian Cancer Syndrome (HBOC), and Lynch Syndrome. The prevalence of these conditions in the general population is estimated to be ~ 1 to 2% [4]. Genetic testing for these disease areas in healthy individuals can identify those that are at high risk for disease despite the lack of obvious symptoms and indicate that these individuals are candidates for increased preventive treatment.

Familial Hypercholesterolemia

Familial Hypercholesterolemia (FH) is one of the most common forms of hereditary cardiovascular conditions with a prevalence of 1:200-1:250 in the United States [5,6]. FH confers an increased risk of developing heart disease in individuals that may appear otherwise completely healthy. Studies have shown that individuals with FH have a higher chance of developing coronary artery disease than individuals without a diagnosis of FH. Proactive treatment and management can significantly reduce the risk of developing the disease [7]. The four genes tested for cases of FH are *APOB, LDLR, LDLRAP1*, and *PCSK9*. Helix's Exome+ assay offers high performance across an FH panel, as shown in Table 2. Variant call rates and CNV call rates represent the percentage of bases or exons across the gene assigned a clinical-quality genotype or copy number value across 4,000 Helix Exome+ assay datasets generated in production.

Table 2: Helix Exome+ assay performance metrics for FH.

Gene	Variant Call Rate*	CNV Call Rate**
APOB	99.98%	100%
LDLR	100%	100%
LDLRAP1	99.99%	100%
PCSK9	99.99%	100%

Hereditary Breast and Ovarian Cancer Syndrome

BRCA1 and *BRCA2* variants can result in Hereditary Breast and Ovarian Cancer Syndrome (HBOC). This syndrome is present in both men and women. Individuals with this condition can have a high risk of developing cancers such as breast (~38 to 87%), ovarian (16 to 63%), melanoma, and pancreatic cancers [8]. In addition to melanoma and pancreatic cancers, men also are at an increased risk of developing male breast cancer and prostate cancer (up to 20%). Pathogenic (disease-causing) variants in additional genes such as *ATM, CDH1, CHEK2, PALB2, PTEN, STK11*, and *TP53* can also convey increased risk for breast and/or ovarian cancer. Helix's Exome+ assay coverage for these genes is described in Table 3.

Table 3: Helix Exome+ assay performance metrics for genes related to breast and ovarian cancer, including but not limited to HBOC.

Gene	Variant Call Rate	CNV Call Rate
BRCA1	100%	99.98%
BRCA2	100%	99.95%
ATM	100%	99.89%
CDH1	99.99%	100%
CHEK2	99.20%	100%
PALB2	100%	100%
PTEN*	99.96%	100%
STK11	100%	100%
TP53	100%	100%

* CNV not offered for *PTEN* exon 9. Pathogenic/likely-pathogenic variants have not been found isolated to *PTEN* exon 9 [9].

Lynch Syndrome

Lynch Syndrome, also known as hereditary nonpolyposis colorectal cancer (HNPCC), confers increased risk to many different cancer types, particularly to colon, endometrial, and gastric cancers. Regular screening and intervention, such as colonoscopies, can significantly reduce the risk of developing colon cancer in individuals with Lynch Syndrome [10]. Helix's Exome+ assay coverage for the genes underlying this disease are described in Table 4.

Table 4: Helix Exome+ assay performance metrics for Lynch Syndrome genes.

Gene	Variant Call Rate	CNV Call Rate
EPCAM	100%	100%
MLH1	99.81%	100%
MSH2	99.81%	99.86%
MSH6	100%	100%
PMS2*	99.70%	100%

* PMS2 exons 12-15 are not included due to the inability to detect gene recombination events, which impact the relevance of PMS2 variants in these exons.

Carrier Screening

Carrier screening can range from single sites or single genes to expanded carrier testing across hundreds of genes. The Helix Exome+ assay supports all approaches for Carrier Testing, offering high coverage for individual genes of interest, as well as a full breadth of covered genes. Table 8 provides an example of a common core carrier test.

Table 8: Helix Exome+ assay performance metrics for genes commonly tested during Ashkenazi Jewish carrier testing.

Gene	Disease	Variant Call Rate	CNV Call Rate
ASPA	Canavan Disease	100%	99.89%
BCKDHB	Maple Syrup Urine Disease (Type 1B)	100%	99.92%
BLM	Bloom Disease	100%	100%
CFTR	Cystic Fibrosis	100%	100%
FANCC	Fanconi Anemia (Group C)	100%	99.92%
G6PC	Glucose-6-phosphate Dehydrogenase Deficiency	100%	100%
HEXA	Tay-Sachs Disease	100%	100%
IKBKAP	Familial Dysautonomia	100%	100%
MCOLN1	Mucolipidosis IV	100%	100%
SMPD1	Niemann-Pick Disease (Types A and B)	99.71%	100%

Pharmacogenomics

Pharmacogenomics (PGx), including comprehensive CYP2D6 calling, is fully supported by the Helix Exome+ assay. CYP2D6 accounts for 25% of all prescription drugs; including the anti-arrhythmics, antidepressants, antipsychotics, beta-blockers, antiemetics, analgesics, and antitussives [12]. For most genes, SNPs and indels are easily converted to star alleles (Table 6). This includes Tier 1 and Tier 2 *CYP2C9* and *CYP2C19* star alleles recommended for clinical testing by the Association for Molecular Pathologists (AMP) [13,14]. For *CYP2D6*, Helix has built a proprietary probabilistic caller that identifies all star alleles in PharmVar version 3.4[15] (excluding *82), as listed in Table 7, with ~100% accuracy.

Table 6: Performance for a subset of PGx-relevant targets.

Gene	Star Allele	Variant	Call Rate
CYP1A2	***1F**	rs762551	100%
	***1K**	rs2069526	100%
		rs12720461	100%
CYP2B6	***4**	rs2279343	68.88%
	***6**	rs3745274	100%
		rs2279343	68.88%
CYP2C9	***2**	rs1799853	100%
	***3**	rs1057910	100%
	***4**	rs56165452	100%
	***5**	rs28371686	100%
	***6**	rs9332131	100%
	***8**	rs7900194	100%
	***11**	rs28371685	100%
CYP3A4	***13**	rs4986909	100%
	***22**	rs35599367	100%
CYP3A5	***3**	rs776746	100%
CYP4F2	***3**	rs2108622	100%
DPYD	***2A**	rs3918290	100%
	***13**	rs55886062	100%
CYP2C19	***2**	rs12769205	99.98%
		rs4244285	99.50%
	***3**	rs4986893	100%
	***4A**	rs28399504	100%
	***4B**	rs28399504	100%
		rs12248560	100%
	***5**	rs56337013	100%
	***6**	rs72552267	100%
	***7**	rs72558186	100%
	***8**	rs41291556	100%
	***9**	rs17884712	100%
	***10**	rs6413438	99.53%
	***17**	rs12248560	99.93%
	***35**	rs12769205	99.98%
SLCO1B1	***5**	rs4149056	100%
TPMT	***2**	rs1800462	96.95%
	***3A**	rs1800460	99.75%
	***3C**	rs1142345	100%
VKORC1	**G3673A**	rs9923231	99.30%

Call Rate = Rate at which this site was assigned a high-quality genotype call across 4,000 samples.

Helix's proprietary pharmacogenomics product reports CYP2D6 star alleles based on the combination of defining SNPs, indels, and exon-level copy number across both CYP2D6 and CYP2D7. In addition to reporting on simple and hybrid alleles, Helix reports whole gene duplications, which can quickly increase activity scores from normal metabolizers to ultrarapid metabolizers. Across a set of 16k individuals, ~ 16% were identified to have whole gene duplications (with another 5% harboring whole gene deletions), highlighting the importance of analyzing CYP2D6 copy number in conjunction with CYP2D6 star alleles.

Helix offers a comprehensive reportable range for CYP2D6 (Table 7) by including all star alleles listed in PharmVar (excluding *82). While some of these star alleles are rare and/or lack interpretation, the importance of identifying these remains significant. When these star alleles are excluded from the reportable range, individuals are often mis-classified as *1 with high-confidence as a normal metabolizer

when the classification for this individual, in truth, remains undetermined and interpretation should not be provided with any confidence.

Table 7: Analytical Range of the CYP2D6 Pipeline, which currently focuses on *CYP2D6* specifically.

CYP2D6 Star Alleles					
*1	*19	*37	*55	*75	*99
*2	*20	*38	*56	*81	*100
*3	*21	*39	*57	*83	*101
*4	*22	*40	*58	*84	*102
*4N	*23	*41	*59	*85	*103
*5	*24	*42	*60	*86	*104
*6	*25	*43	*61	*87	*105
*7	*26	*44	*62	*88	*106
*8	*27	*45	*63	*89	*107
*9	*28	*46	*64	*90	*108
*10	*29	*47	*65	*91	*109
*11	*30	*48	*68	*92	*110
*12	*31	*49	*69	*93	*111
*13	*32	*50	*70	*94	*112
*14	*33	*51	*71	*95	*113
*15	*34	*52	*72	*96	*114
*17	*35	*53	*73	*97	
*18	*36	*54	*74	*98	

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. 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 [16]. For example, a PRS for coronary artery disease (CAD) has been shown to identify 8% of the population at ≥ 3-fold increased risk, revealing a significantly higher frequency of affected individuals than those found by search of monogenic pathogenic variants [17]. Additionally, a PRS associated with a 3.7-fold increased risk of early-onset myocardial infarction has been shown to identify ten times the number of individuals at risk as compared to monogenic testing for this disease (which is associated with a 3.8-fold increased risk). While the risk of disease is similar

between PRS and monogenic testing in this case, PRS is a more effective tool at identifying those individuals carrying such risk [16].

Because PRS are based on many genomic loci, often focused in non-coding regions, their coverage requirements are significantly different than those for standard genetic testing of monogenic disease. The Helix 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.

Research

The breadth of the Exome+ assay allows researchers to not only conduct comprehensive common-variant GWAS analysis, but also discover and analyze novel low-frequency and rare variant contributions to phenotypes and disease. The Helix Research Team is experienced with a variety of clinical, pharmacological, and statistical methods, and is available to collaborate with partners with IRB-approved research protocols to: facilitate panel design; perform sample stratification and selection, algorithm selection, and common and rare-variant GWAS; develop, test, and replicate polygenic risk scores; and assist with data integration and normalization. Our high performance and scalable infrastructure enables screening of thousands of traits against hundreds of thousands of samples in less than a day. This research platform was employed to analyze the 50,000 exomes on > 4,000 traits from the UKBiobank, with results described and published on our blog[18] in under two weeks.

Helix's specialized research database is regularly updated with newly-sequenced samples, allowing allele frequencies to be calculated and updated on whole or cohort-specific subsets and further subdivided against selected sample metadata, all in just seconds. Because every sample has already been fully sequenced, updates or additions to a study design can be supported immediately, without having to re-sequence participants. The Helix Research Team encourages submissions to scientific conferences and peer-reviewed publications based on Helix Exome+ data, such as this upcoming paper in Nature Medicine [4] describing population health genetic screening for CDCT1 conditions, where it was found that 90% of at-risk carriers were missed by standard protocols.

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 Helix Exome+ assay results in ancestry assignments for each patient tested, spanning five regions plus the identification of individuals of Ashkenazi Jewish descent (Table 9).

Table 9: List of high-level populations reported by Helix.

Ancestry Populations
African
Ashkenazi Jewish
East Asian
European
Indigenous American
South Asian

Learn More

Analytic performance for our variant calls are described in various other white papers. Of interest, small variant performance is described in our Performance White Paper [1] and CNV performance is described in our CNV White Paper [2].

References

1. Performance White Paper: https://cdn.shopify.com/s/files/1/2718/3202/files/Helix_Performance_White_Paper_v4.pdf?7992
2. CNV White Paper: https://cdn.shopify.com/s/files/1/2718/3202/files/EV2__CNV_White_Paper_2.pdf?7992
3. https://www.cdc.gov/genomics/implementation/toolkit/tier1.htm
4. Grzymski et al., Population Health Genetic Screening for Tier 1 Inherited Diseases in Northern Nevada: 90% of At-Risk Carriers are Missed. bioRxiv 650549; doi: https://doi.org/10.1101/650549.
5. Akioyamen et al., Estimating the prevalence of heterozygous familial hypercholesterolaemia: a systematic review and meta-analysis. BMJ Open. 2017; 7(9).
6. Singh and Bittner, Familial hypercholesterolemia -- epidemiology, diagnosis, and screening. Cur Atheroscler Rep. 2015; 17(2)482.
7. Safarova, M. S., & Kullo, I. J. (2016). My Approach to the Patient With Familial Hypercholesterolemia. *Mayo Clinic proceedings*, *91*(6), 770–786. doi:10.1016/j.mayocp.2016.04.013
8. Petrucelli et al., BRCA1- and BRCA2-Associated Hereditary Breast and Ovarian Cancer. Gene Reviews. 1998 Sep 4
9. Adam et al., PTEN Harmartoma Tumor Syndrome. Gene Reviews. Updated 2016.
10. Jarvinen HJ, Mecklin JP, Sistonen P. Screening reduces colorectal cancer rate in families with hereditary nonpolyposis colorectal cancer. Gastroenterology 1995;108:1405–11
11. Kalia et al., Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genetics in Medicine 19(2) Feb 2017.
12. Monte, Andrew A et al. “CYP2D6 Genotype Phenotype Discordance Due to Drug-Drug Interaction.” *Clinical pharmacology and therapeutics* vol. 104,5 (2018): 933-939. doi:10.1002/cpt.1135

13. Pratt et al., Recommendations for Clinical CYP2C9 Genotyping Allele Selection: A Joint Recommendation of the Association for Molecular Pathology and College of American Pathologists. The Journal of Molecular Diagnostics (accepted 22April2019).
14. Pratt et al., Recommendations for Clinical CYP2C19 Genotyping Allele Selection. The Journal of Molecular Diagnostics, Volume 20, Issue 3, 269-276.
15. https://www.pharmvar.org/gene/CYP2D6, with v3.4 star alleles found in the excel spreadsheet dated 'Nov 5, 2018', downloaded here: https://api.pharmgkb.org/v1/download/file/attachment/CYP2D6_allele_definition_table.xlsx
16. Khera et al., Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet 2018 Sep; 50(9):1219-1224.
17. Khera and Chaffin et al., Whole-Genome Sequencing to Characterize Monogenic and Polygenic Contributions in Patients Hospitalized With Early-Onset Myocardial Infarction. Circulation 2019 Mar; 139(13):1593-1602.
18. Cirulli and Washington, Researchers have access to new data on thousands of exomes. Here's what we found. March 27,2019. https://blog.helix.com/uk-biobank-helix-research/