

Clinical Validation of Sema4 Signal Hereditary Cancer Test on Traversa™

OVERVIEW

Sema4 has designed and validated a Universal Hereditary Cancer Panel, which consists of 112 genes associated with HC conditions such as Hereditary Breast and Ovarian Cancer (HBOC) syndrome, Lynch syndrome, Li-Fraumeni syndrome, Familial Adenomatous Polyposis, and others. Sema4 Signal Hereditary Cancer Universal Panel consists of 112 genes. Fourteen additional, nested panels covering both comprehensive and tissue-specific options will be offered. The complete list of genes and panels can be found in **Supplemental Table 1**. The list of covered regions is comprised of the coding sequence (CDS -13/+6) and functional regions of the 112 genes represented in the Human Gene Mutation Database (HGMD), ClinVar, and other similar resources, carefully reviewed by Sema4's in-house clinical geneticists. Included in the target regions are any areas flanking known genomic variants of clinical significance. The validation study demonstrated that the overall accuracy exceeded 99%.

The platform for the Sema4 hereditary cancer (HC) offering is our NGS-based medical exome platform, called Traversa™, which includes >6,000 total genes. Running the medical exome in the background gives Sema4 the ability to easily add genes to these HC panels as additional evidence becomes available.

BACKGROUND

Hereditary cancer accounts for approximately 5-10% of all cancers. Multigene panels for hereditary cancer risk assessment have multiple advantages, including time and cost-effectiveness, compared to traditional testing focused on select genes (e.g., *BRCA1* and *BRCA2* genes) (1 & 2). Sema4 offers a comprehensive menu for hereditary cancer comprised of 17 panels including 2 traditional *BRCA1/2* offerings (testing for *BRCA1* plus *BRCA2* and the 3 *BRCA* founder mutations in the Ashkenazi Jewish population), in addition to custom gene sequencing and targeted analysis for known familial variants.

Four of these panels cover multiple cancers: 1) the Universal Cancer Panel (112 genes); 2) the Comprehensive Cancer Panel (72); 3) the High Prevalence Cancer Panel (36); and 4) the Pediatric Cancer Panel (49). A description of the three general cancer panels can be found in **Table 1**. Sema4 also offers cancer panels by specific organ systems, including breast, colorectal, pancreas, prostate, melanoma, endocrine, brain, renal, and urinary tract. These panels include genes with definitive, strong, or moderate evidence association with hereditary cancer (5).

The goal of this validation study was to establish the sensitivity, specificity, and precision of our test compared to industry standards.

METHODS

The HC assay is comprised of an NGS panel with full coverage of exons, exon/intron boundaries, and other known variant regions. Additionally, full gene coverage for 13 common genes (*BRCA1/2*, *PALB2*, Lynch and polyposis genes, *TP53*, *PTEN* & *VHL*) is built into the assay design in order to assist in detection of deletions and duplications. Multiplex Ligation-dependent Probe Assay (MLPA) for 7 genes (*BRCA1*, *BRCA2*, *PMS2*, *MSH6*, *EPCAM*, *MLH1*, and *MSH2*) is performed in parallel to provide increased sensitivity of copy number variants (CNVs). Reportable variants are confirmed using orthogonal methods, including MLPA, qPCR, and/or Sanger sequencing. Variants identified in the high homology region of *PMS2* will be confirmed using long-range polymerase chain reaction (LRPCR) followed by exon-specific nested PCR and Sanger sequencing.

A total of 114 unique positive samples were used during the validation of this assay. This sample set consisted of previously characterized HC positive patient samples and well-categorized reference standards. An additional 401

Table 1: General Cancer Panels

Universal Panel (112 genes)	Includes 72 genes from Comprehensive Panel plus 40 genes with emerging evidence of cancer predisposition
Comprehensive Panel (72 genes)	All genes on this panel are either guideline-driven or have well-established evidence of increased cancer risk but not yet included in guidelines
High Prevalence Panel (36 genes)	All genes on this panel are either guideline-driven or have well-established evidence of increased cancer risk for the most common cancer types (i.e. high prevalence cancer diagnoses)

uncharacterized samples of various specimen types, including saliva, whole blood, FFPE tissue, and buccal swabs, were used to assess specimen type performance, assay input, and to establish a baseline for CNV calling. A breakdown of the specimen types used in this validation can be found in **Table 2**.

The underlying technology for the Traversa™ platform includes genomic library preparation with targeted capture utilizing 120-base oligonucleotide probes. For genomic DNA fragmentation and barcoded library preparation, we utilize Kapa Biosciences' Hyper Plus kit together with customized barcoded adapters from IDT Technologies. Capture is performed with the use of custom probes synthesized by Twist Biosciences.

Resulting libraries are sequenced on Illumina NovaSeq™ 6000 Sequencing System. The data generated was processed through the in-house developed pipeline with the following key components: GATK4 Haplotype Caller for single nucleotide polymorphism/Indel (insertion/deletion) detection of the 112 genes, GATK4 Haplotype Caller for CNV calling of the 112 genes, and Manta Breakpoint Discovery Calling for breakpoint analysis of the *MSH2* inversion, also known as the Boland inversion (3).

Variant assessment was based on standards and guidelines for the interpretation of sequence variants recommended by the American College of Medical Genetics and Genomics with the Association for Molecular Pathology in 2015 (4). Sema4's variant curation team uses a variety of databases (i.e. Gnomad, ExAC, ClinVar, HGMD, and gene-specific databases) along with exhaustive publication review.

Table 2: HC Sample Breakdown

Sample Type	Number
HC Positive Samples	105
Whole Blood	96
Saliva	8
Chronic Villi*	1
HC positive Controls	9
Boland inversion Spike-In	1
NIBSC	7
Seraseq®	1
Normal Specimens	401
Total Samples	515

*Sample was used as a CNV positive, but will not be a specimen type in production

RESULTS

- Sequencing performance during this validation showed an average of 99.2% of regions had >20x coverage, and the average HC-specific mean target coverage was 195x.
- Inter- and intra-assay reproducibility were assessed using eight samples encompassing 17 variants. Results showed 99% reproducibility (**Table 3**).
- Sensitivity was assessed using 105 previously characterized patient specimens and eight reference standards covering 147 variants. Results showed 100% concordance with expected results (**Table 4**).

Table 3: Precision Assessment

Sample	# of Variants Expected	Variants Run 1/ Rep 1	Variants Run1 /Rep 2	Variants Run 1/ Rep 3	Variants Run 2	Variants Run 3
1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	1	1	1	1	1	1
4	3	3	3	3	3	3
5	5	5	5	5	5	5
6	1	1	1	1	1	1
7	3	3	3	3	3	3
8	1	1	1	0*	1	1

*This was a mosaic variant that was observed at 10-13% in four of the replicates of this sample and was seen at a lower % in the gVCF of the fifth replicate.

Table 4: Sensitivity Assessment

Category	# Expected Variants	# Observed Variants	Concordance
Previously characterized patient cases N=106	139	139	100%
Reference Standards with known HC germline variants N=7	16	16	100%

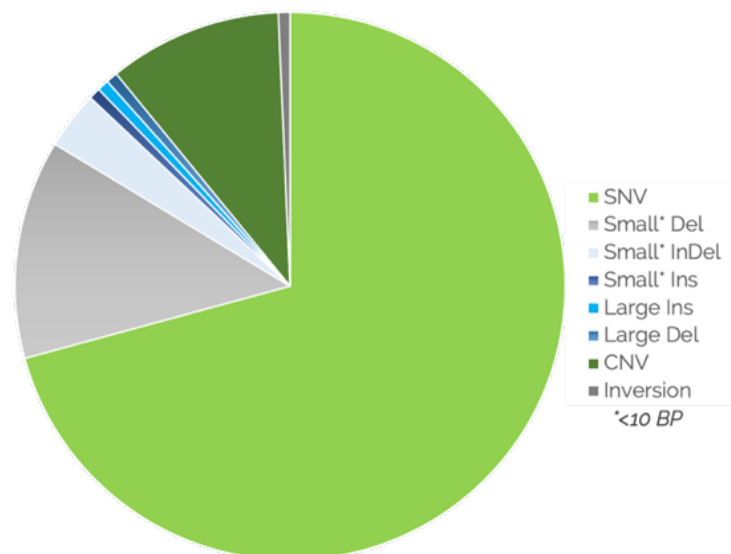
- Specificity was assessed by reviewing the false positive calls of 18 previously characterized samples. Results showed 100% specificity for single-nucleotide variants (SNVs) and Indels while only showing 72% specificity for CNVs (Table 4). All CNVs will be confirmed using MLPA, qPCR, or UHRMEA during clinical testing, which will effectively bring the specificity to 100%, and as the same false positives were seen across multiple specimens, it is likely that these will be deemed artifacts.
- Eight replicates of the Seraseq® Inherited Cancer DNA Mix v1 were processed as part of the validation. Of the 22 expected variants detectable by our pipeline, 21 were seen across all eight replicates. One variant (*MSH2* c.942+3A>T) was detected in seven replicates and missed in one.

- A breakdown of the number of variants detected during the validation by gene and type can be found in **Figure 1** and **Table 5**.

Table 5: Variant Breakdown by Gene

Variant Type	AIP	APC	ATM	BARD1	BLM
CNVs	0	0	1	0	1
SNVs/Indels	1	14	8	2	1
Variant Type	BRCA1	BRCA2	BRIP1	CDH1	CHEK2
CNVs	0	1	0	0	2
SNVs/Indels	8	15	3	1	7
Variant Type	DICER1	FANCC	FH	FLCN	MET
CNVs	0	0	0	0	0
SNVs/Indels	6	2	2	1	1
Variant Type	MITF	MLH1	MRE11A	MSH2	MSH6
CNVs	0	1	0	6	1
SNVs/Indels	1	6	6	1	4
Variant Type	MUTYH	NBN	NF1	PALB2	PMS2
CNVs	0	0	0	1	1
SNVs/Indels	2	2	3	5	6
Variant Type	POLE	POT1	RAD50	RAD51C	RAD51D
CNVs	0	1	0	0	0
SNVs/Indels	2	0	3	2	1
Variant Type	RB1	RET	SDHA	SDHC	SMAD4
CNVs	0	0	0	0	0
SNVs/Indels	3	2	1	1	1
Variant Type	SMARCA4	TP53	TSC2		
CNVs	0	0	0		
SNVs/Indels	2	1	4		

Figure 1: Variant Breakdown by Type



CONCLUSION

Our validation studies of the Sema4 HC panels showed that the inter- and intra-assay reproducibility was as high as 99%. The validation study yielded 100% sensitivity with expected results and 100% specificity for SNVs and Indels. The most relevant variants validated as part of this study can be found in **Table 6**.

A detailed study to determine the variant of uncertain significance (VUS) rate of these panels will be conducted and reported separately.

REFERENCES

1. LaDuca H, *et al.* Utilization of multigene panels in hereditary cancer predisposition testing: analysis of more than 2,000 patients. *Genet Med.* 2014 Nov;16(11):830-7. doi: 10.1038/gim.2014.40. Epub 2014 Apr 24. PubMed PMID: 24763289.
2. Lincoln SE, *et al.* A Systematic Comparison of Traditional and Multigene Panel Testing for Hereditary Breast and Ovarian Cancer Genes in More Than 1000 Patients. *J Mol Diagn.* 2015 Sep;17(5):533-44. doi:10.1016/j.jmoldx.2015.04.009. Epub 2015 Jul 22. PubMed PMID: 26207792.
3. Rhees J, *et al.* Inversion of exons 1-7 of the MSH2 gene is a frequent cause of unexplained Lynch syndrome in one local population. *Fam Cancer.* 2014 Jun;13(2):219-25. doi: 10.1007/s10689-013-g688-x. PubMed PMID: 24114314.
4. Richards S, *et al.*; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015 May;17(5):405-24. doi: 10.1038/gim.2015.30. Epub 2015 Mar 5. PubMed PMID:25741868.
5. Strande NT, *et al.* Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. *Am J Hum Genet.* 2017 Jun 1;100(6):895-906. doi: .1016/j.ajhg.2017.04.015. Epub 2017 May 25. PubMed PMID: 28552198.

Table 6: Most Common and Difficult HC Variants Included in Validation

Gene	Variant	# Unique Specimens	Total # of Libraries	Detection Rate
MSH2	Boland inversion	1	1	100%
MSH2	c.942+3A>T	2	9	89%*
BRCA1	c.5266dupC	1	8	100%
BRCA1	c.68_69delAG	4	11	100%
BRCA2	c.5946delT	4	4	100%
APC	c.3920T>A	5	9	100%
CHEK2	c.1100delC	3	3	100%
CHEK2	c.1283C>T	2	2	100%
BRCA2	c.3395_3396delAAinsGG	2	2	100%
PMS2	c.2404C>T	1	1	100%
PMS2	c.1733G>A	1	9	100%

*The final detection rate in production will be close to 100% when coupled with visual review

Supplemental Table 1

Sema4 Signal Multigene Panels

*Universal Panel	Comprehensive Panel	High Prevalence Panel	Ashkenazi Jewish BRCA Panel	BRCA1 & BRCA2 Panel	Breast Guidelines Panel	Breast & Gynecological Panel	Lynch Syndrome Panel	Colorectal Panel	Pancreatic Panel	Pancreatic Plus Panel	Prostate Panel	Melanoma Panel	Endocrine Tumor Panel	Brain/Neurological Panel	Renal/Urinary Panel	Pediatric Panel
112 genes	72 genes	36 genes	‡3 variants	2 genes	11 genes	27 genes	5 genes	20 genes	21 genes	27 genes	15 genes	9 genes	21 genes	32 genes	25 genes	49 genes
AIP	AIP												AIP	AIP		
ALK	ALK													ALK		ALK
APC	APC	APC						APC	APC	APC			APC	APC		APC
ATM	ATM	ATM			ATM	ATM		ATM	ATM	ATM	ATM					ATM
AXIN2	AXIN2	AXIN2						AXIN2								
BAP1	BAP1											BAP1		BAP1	BAP1	
BARD1	BARD1	BARD1				BARD1								BARD1		
BLM	BLM															BLM
BMPR1A	BMPR1A	BMPR1A						BMPR1A	BMPR1A	BMPR1A						BMPR1A
BRCA1	BRCA1	BRCA1	BRCA1	BRCA1	BRCA1	BRCA1			BRCA1	BRCA1	BRCA1					
BRCA2	BRCA2	BRCA2	BRCA2	BRCA2	BRCA2	BRCA2			BRCA2	BRCA2	BRCA2	BRCA2				
BRIP1	BRIP1	BRIP1				BRIP1										
CASR										CASR						
CDC73													CDC73		CDC73	CDC73
CDH1	CDH1	CDH1			CDH1	CDH1		CDH1								
CDK4	CDK4	CDK4							CDK4	CDK4		CDK4				
CDKN1B	CDKN1B												CDKN1B	CDKN1B		
CDKN1C																CDKN1C
CDKN2A	CDKN2A	CDKN2A							CDKN2A	CDKN2A		CDKN2A		CDKN2A		
CEBPA	CEBPA															CEBPA
CFTR										CFTR						
CHEK2	CHEK2	CHEK2			CHEK2	CHEK2		CHEK2			CHEK2		CHEK2			
CPA1										CPA1						
CTC1																
CTRC										CTRC						
DDB2																
DICER1	DICER1	DICER1				DICER1							DICER1	DICER1	DICER1	DICER1
DIS3L2															DICER1	DIS3L2
DKC1																DIS3L2
EGFR	EGFR															
EPCAM	EPCAM	EPCAM				EPCAM	EPCAM	EPCAM	EPCAM	EPCAM	EPCAM			EPCAM	EPCAM	EPCAM
ERCC2																
ERCC3																
ERCC4																
ERCC5																
FANCA																
FANCB																
FANCC						FANCC										
FANCD2																
FANCE																
FANCF																
FANCG																
FANCI																
FANCL																
FANCM																
FH	FH												FH		FH	FH
FLCN	FLCN														FH	FLCN
GATA2	GATA2															GATA2
GPC3														GPC3		GPC3
GREM1	GREM1	GREM1						GREM1								
HOXB13	HOXB13	HOXB13									HOXB13					
HRAS	HRAS													HRAS		HRAS
KIT	KIT															
MAX	MAX												MAX			MAX
MEN1	MEN1								MEN1	MEN1			MEN1	MEN1		MEN1
MET	MET														MET	

*Universal Panel encompasses all the 112 genes available for Sema4 Signal Hereditary Cancer

‡Variants: BRCA1 (c.5266dupC), BRCA1 (c.68_69delAG), BRCA2 (c.5946delT)

Supplemental Table 1

Continued

*Universal Panel	Comprehensive Panel	High Prevalence Panel	Ashkenazi Jewish BRCA Panel	BRCA1 & BRCA2 Panel	Breast Guidelines Panel	Breast & Gynecological Panel	Lynch Syndrome Panel	Colorectal Panel	Pancreatic Panel	Pancreatic Plus Panel	Prostate Panel	Melanoma Panel	Endocrine Tumor Panel	Brain/Neurological Panel	Renal/Urinary Panel	Pediatric Panel
112 genes	72 genes	36 genes	†3 variants	2 genes	11 genes	27 genes	5 genes	20 genes	21 genes	27 genes	15 genes	9 genes	21 genes	32 genes	25 genes	49 genes
MITF	MITF											MITF			MITF	
MLH1	MLH1	MLH1				MLH1	MLH1	MLH1	MLH1	MLH1	MLH1			MLH1	MLH1	MLH1
MRE11A						MRE11A										
MSH2	MSH2	MSH2				MSH2	MSH2	MSH2	MSH2	MSH2	MSH2			MSH2	MSH2	MSH2
MSH6	MSH6	MSH6				MSH6	MSH6	MSH6	MSH6	MSH6	MSH6			MSH6	MSH6	MSH6
MUTYH	MUTYH	MUTYH				MUTYH	MUTYH									
NBN	NBN	NBN			NBN	NBN					NBN			NBN		NBN
NF1	NF1	NF1			NF1	NF1			NF1	NF1			NF1	NF1	NF1	NF1
NF2	NF2														NF2	NF2
NHP2																
NOP10																
NTHL1	NTHL1							NTHL1								
PALB2	PALB2	PALB2			PALB2	PALB2			PALB2	PALB2	PALB2					
PDGFRA	PDGFRA															
PHOX2B	PHOX2B													PHOX2B		PHOX2B
PMS2	PMS2	PMS2				PMS2	PMS2	PMS2	PMS2	PMS2	PMS2			PMS2	PMS2	PMS2
POLD1	POLD1	POLD1						POLD1								
POLE	POLE	POLE						POLE								
POLH																
POT1	POT1											POT1		POT1		
PRKAR1A	PRKAR1A												PRKAR1A	PRKAR1A		PRKAR1A
PRSS1										PRSS1						
PTCH1	PTCH1													PTCH1		PTCH1
PTEN	PTEN	PTEN			PTEN	PTEN		PTEN				PTEN	PTEN	PTEN	PTEN	PTEN
RAD50						RAD50										
RAD51C	RAD51C	RAD51C				RAD51C						RAD51C				
RAD51D	RAD51D	RAD51D				RAD51D						RAD51D				
RB1	RB1													RB1		RB1
RECQL4																RECQL4
RET	RET												RET			RET
RTEL1																
RUNX1	RUNX1															RUNX1
SDHA	SDHA												SDHA	SDHA	SDHA	SDHA
SDHAF2	SDHAF2												SDHAF2			SDHAF2
SDHB	SDHB												SDHB	SDHB	SDHB	SDHB
SDHC	SDHC												SDHC	SDHC	SDHC	SDHC
SDHD	SDHD												SDHD	SDHD	SDHD	SDHD
SLX4																
SMAD4	SMAD4	SMAD4						SMAD4	SMAD4	SMAD4						
SMARCA4	SMARCA4	SMARCA4				SMARCA4								SMARCA4	SMARCA4	SMARCA4
SMARCB1	SMARCB1													SMARCB1	SMARCB1	SMARCB1
SPINK1										SPINK1						
STK11	STK11	STK11			STK11	STK11		STK11	STK11	STK11						STK11
SUFU	SUFU													SUFU		SUFU
TERC																TERC
TERT																TERT
TINF2																
TMEM127	TMEM127												TMEM127			TMEM127
TP53	TP53	TP53			TP53	TP53		TP53	TP53	TP53	TP53	TP53	TP53	TP53	TP53	TP53
TSC1	TSC1	TSC1							TSC1	TSC1				TSC1	TSC1	TSC1
TSC2	TSC2	TSC2							TSC2	TSC2				TSC2	TSC2	TSC2
VHL	VHL	VHL							VHL	VHL			VHL	VHL	VHL	VHL
WT1	WT1														WT1	WT1
XPA																
XPC																
XRCC2	XRCC2					XRCC2										

*Universal Panel encompasses all the 112 genes available for Sema4 Signal Hereditary Cancer

†Variants: BRCA1 (c.5266dupC), BRCA1 (c.68_69delAG), BRCA2 (c.5946delT)