SOPHiA™

EXOME SOLUTIONS

Focusing on what matters for rare and inherited diseases

HIGHLIGHTS

- Customizable gene content associated with rare and inherited diseases
- Advanced analytical performance
- Robust solutions designed to detect multiple types of variants (SNVs, Indels and CNVs) in one single experiment
- Dedicated variant filtering options, including fully automated Familial Variant Analysis (trio-analysis)

The SOPHiA Exome Solutions include two genomic applications, Clinical and Whole Exome, that both bundle a smart capture-based target enrichment kit with the analytical power of SOPHiA™ AI and full access to the SOPHiA DDM™ platform.

Expertly designed, these solutions provide a comprehensive coverage of multiple types of genomic variants in more than 19,000 genes, enabling data-informed decision making.

SMART KIT DESIGN

- High affinity probe design ensuring extremely uniform coverage of the target region
- Comprehensive applications targeting:
  - 4,490 genes for Clinical Exome
  - >19,000 genes for Whole Exome
- Automated workflow available on leading liquid handling robots for high-throughput library preparation

ANALYTICAL POWER

- Advanced analytical performance (i.e. >99% sensitivity and reproducibility)
- High-confidence calling of SNVs and Indels
- Efficient CNV detection available for:
  - 98.1% of Clinical Exome genes
  - 93.3% of Whole Exome genes

UNIVERSAL PLATFORM

- Intuitive and user-friendly interface
- Secure storage of anonymized data
- Dedicated features for simplified data visualization and interpretation
- Fully customizable report

Democratizing Data-Driven Medicine

SOPHiA GENETICS supports clinical researchers to better analyze and interpret genomic data.
Experts who use our solutions benefit from:

SOPHiA AI
Over 200 genomic applications supported

Set Up Program
Rapid adoption of genomic applications

Data security policy
Compliance with national privacy laws, GDPR, HIPAA guidelines and applicable legislation

SOPHiA's community
Anonymized and safe knowledge sharing among experts worldwide

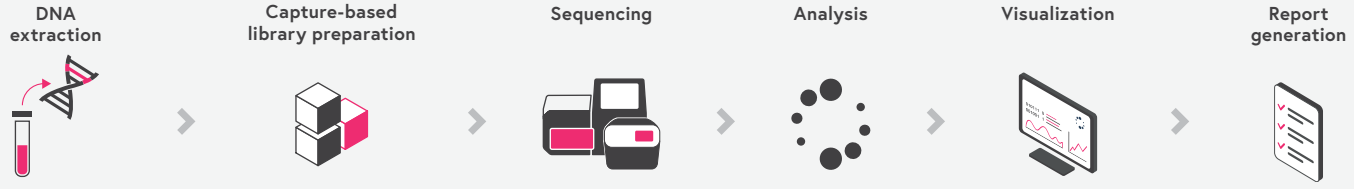
© SOPHiA GENETICS 2020 - PM_RUO_B.1.3.2.4_r3en

Exome Solutions

Streamlined workflow from DNA extraction to variant report generation

SOPHiA Exome Solutions provide a straightforward library preparation workflow. Ready-to-sequence target-enriched libraries are generated in just 2 working days, starting from 200 ng of DNA. For high-throughput needs, DNA extraction and library preparation can be fully automated, using pre-optimized protocols for a variety of liquid handling robots.

Library preparation of both applications is compatible with Illumina and Thermo Fisher Scientific platforms. Sequencing output files are then analyzed by SOPHiA, that adapts to the specifics of each sequencer, always ensuring advanced performance. Finally, results are displayed on the SOPHiA DDM platform where clinical researchers can easily interpret them and generate a complete variant report.



Relevant gene content

SOPHiA Clinical Exome Solution and Whole Exome Solution cover the coding regions (±5bp of exon-flanking regions) of 4,490 genes* associated with the vast majority of inherited diseases, and more than 19,000 genes enabling an exome-wide investigation, respectively.

Probe design is optimized to provide high coverage uniformity throughout the entire target region, resulting in valuable data quality. For specific needs, the gene content can be customized.

*Complete list of genes available upon request.

Smart kit specifications

Parameter	Details
Sample source	Blood
DNA input requirement	200 ng
Target region	12 Mb (Clinical Exome)
	39 Mb (Whole Exome)
Library preparation time	2 days

Sequencing and multiplexing recommendations

Sequencers	Flow Cell / Ion Chip Kit	Recommended samples per run (for >50x coverage depth)	
		Clinical Exome	Whole Exome
Illumina NovaSeq® 6000	SP	48 (per lane)	12
	S1	96 (per lane)	24 (per lane)
	S2	NA	56 (per lane)
Illumina HiSeq® 3000/4000	High Output (2x100bp)	24 (per lane)	NA
	High Output (2x150bp)	32 (per lane)	NA
Illumina HiSeq® 2500	High Output (2x125bp)	24 (per lane)	6 (per lane)
	Rapid Run Mode (2x150bp)	16 (per lane)	3 (per lane)
Illumina MiSeq®	v3 (2x300bp)	4	NA
Illumina NextSeq® 500/550	Mid Output Kit v2 (2x150bp)	16	3
	High Output Kit v2 (2x150bp)	48	12

Extremely uniform coverage

SOPHiA Exome Solutions achieve very high on-target rates, which ensure reliable coverage uniformity values across all the target regions, even in GC-rich regions (Fig. 1).
Equal read coverage is of crucial importance for the precise identification of multiple types of variations, including CNVs.

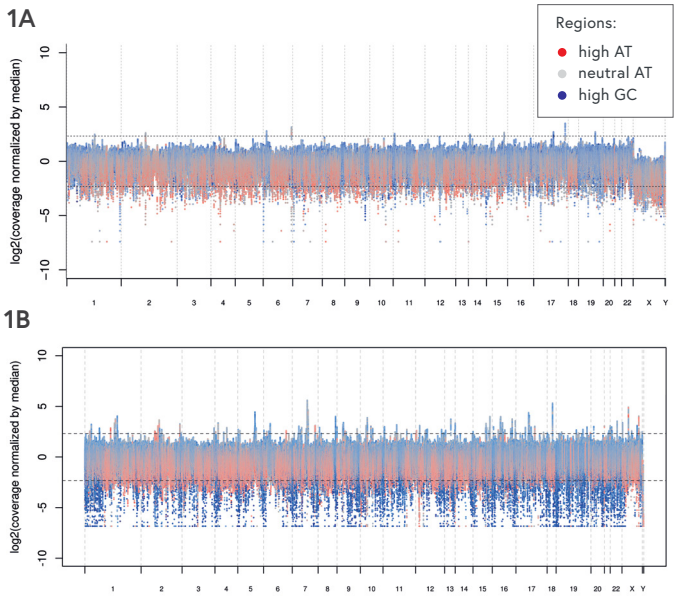


Figure 1: Coverage uniformity profile of a typical sample analyzed with SOPHiA Clinical (1A) and Whole Exome Solution (1B). The X-axis represents the chromosomic positions targeted by each solution and the Y-axis the log_2 coverage normalized by the median. The closer the dots are to the 0 line, the more homogenous the reads are covering each target. Dashed lines represent 20% (lower line) and 500% (upper line) of the median coverage.

For Research Use Only.
Not for use in diagnostic procedures

Exome Solutions

Advanced analytical performance

SOPHiA analyzes complex NGS data by detecting, annotating and pre-classifying SNVs, Indels and CNVs in all the genes* covered by the solutions, in a single experiment.

**CNV detection is available for 98.1% and 93.3% of the genes covered by SOPHiA Clinical and Whole Exome Solution, respectively.*

Analysis time from FASTQ files: Overnight
Analysis time may vary depending on the number of samples multiplexed and server load.

**Performance metrics are based on high confidence regions in a reference sample. Values have been calculated on a reference sample and 80 M reads per sample on Illumina HiSeq® (300bp read length)*
†Depth is >50x for SOPHiA Clinical Exome Solution and >20x for SOPHiA Whole Exome Solution

	Clinical Exome Solution		Whole Exome Solution
	Observed	Lower 95% CI	Observed
Sensitivity	99.45%	99.02%	> 99%*
Specificity	99.99%	99.99%	
Accuracy	99.99%	99.99%	
Precision	99.45%	99.02%	> 99%*
Reproducibility	99.99%	99.98%	> 99%
Average on-target rate	> 90%		> 90%
Coverage uniformity	> 98%		> 98%
Average % of target region†	> 96%		> 99%

High-confidence calling of CNVs

Copy Number Variations (CNVs) play an important role in a broad range of genetic disorders[1]. Accurate CNVs detection via exome-based profiling can result in increased analytical yield. However, classical extended exome application settings render the detection of CNVs very difficult due to the extended target regions and the increased depth of sequencing needed to reliably identify CNVs.

SOPHiA detects CNVs* at a resolution of 2-5 exons (Fig. 2) in both applications. This analysis is performed by evaluating the coverage levels of the target regions across all samples within the same sequencing run. For each sample, SOPHiA automatically selects a set of reference samples from the same run, based on the similarity of coverage patterns. Subsequently, the coverage is normalized by sample and target region using the reference samples, enabling CNV calling.

**Accurate CNV calling requires at least 8 co-captured samples*

Thanks to its accuracy, the use of both applications reduces the need for additional assays by allowing the simultaneous detection of SNVs, Indels and CNVs. The result is a fast, nimble and cost-effective workflow.

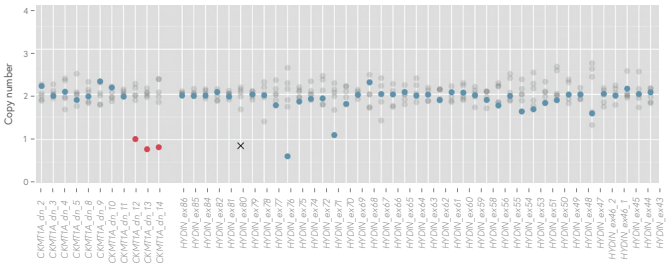
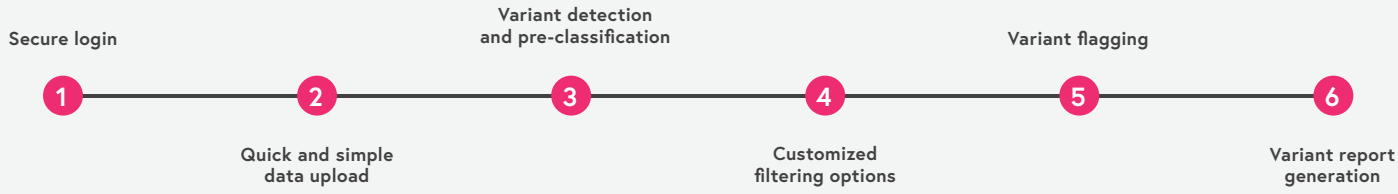


Figure 2: Normalized coverage levels of Copy Number status for SOPHiA Clinical Exome Solution.
Plot shows the normalized coverage levels in a given sample (blue and red dots) compared to the reference coverage levels (grey dots). Blue dots correspond to target regions without CNVs, red dots to deletions. Solid dots represent high-confidence CNV predictions.

Enhanced variant visualization and interpretation

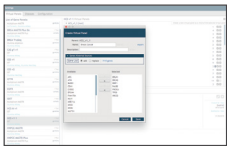
The SOPHiA DDM platform features intuitive variant filters, dual variant pre-classification and reporting functionalities to simplify data visualization and interpretation. The platform enables clinical researchers to explore and interpret genomic variants and report significant findings.



For Research Use Only.
Not for use in diagnostic procedures

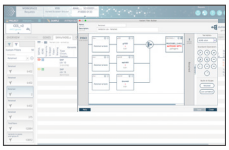
Exome Solutions

Integrated features for efficient variant prioritization and interpretation



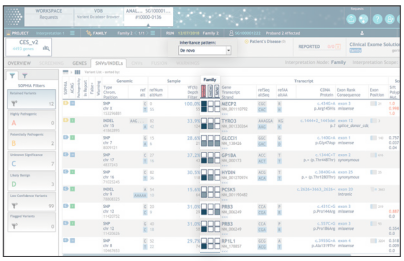
Virtual Panel

Limit the interpretation to a subset of genes for quicker screening of relevant variants. SOPHiA DDM offers direct access to the Human Phenotype Ontology (HPO) database to filter genes associated with specific diseases. This feature enables a quick view for the genes of interest with the related total number and types of variants, pathogenicity and prediction levels.



Variant Filter Builder

Define, design, and edit complex variant queries. Custom filters can be generated by dragging and dropping operators, selecting corresponding filters like population frequency, variant fraction, coverage and/or ACMG scoring. Additionally, the selected prioritization strategy can be saved and used for other analyses.



Familial Variant Analysis (trio-analysis)

Facilitate variant interpretation through a duo- or trio-based approach to increase the yield of NGS-based analysis. This feature facilitates the identification of the causative variant(s) responsible for a proband's phenotype through the analysis of parental samples. The number of variants under investigation is organized through the multi-layered statistical modelling of SOPHiA DDM, that streamlines the analysis based on specific segregation pattern. Users can filter the variants based on different modes of inheritance such as autosomal recessive, compound heterozygosity, autosomal dominant, X-linked, and also identify de novo variations.

Dual Variant Pre-Classification

Improve assessment of variants pathogenicity with both the ACMG score and SOPHiA's prediction. Concerning the ACMG score, SOPHiA automatically gathers and collates information from various sources (frequencies in the population, *in silico* scores, disease-specific data, splicing predictors, databases on protein domains, loss of function and repetitive regions) to evaluate the sets of criteria: one to classify pathogenic or likely pathogenic variants; one to classify benign or likely benign variants; and one to determine variants of unknown significance. Each pathogenic criterion is weighted as very strong, strong, moderate, or supporting, and each benign criterion is weighted as stand-alone, strong, or supporting. SOPHiA then combines the values of these criteria (true or false), according to the rules recommended by the ACMG, and calculates the corresponding pre-classification score, linked to the level of pathogenicity. SOPHiA's predictions for pathogenicity (A = highly pathogenic, B = likely pathogenic, C = variant of unknown or uncertain significance, D = most likely benign) are based on rules defined using machine learning technologies, enabling improved assessment of variants significance.

Access to SOPHiA's community

In SOPHiA DDM, experts from hundreds of healthcare institutions interpret the results and flag the pathogenicity level of variants in accordance to their knowledge. This highly valuable information feeds the variant knowledge base and is anonymously and safely shared among the members of the community.

Respect data privacy

SOPHiA DDM encrypts all data to the highest industry standards before storing it redundantly in secured and private data centers. The platform ensures data protection and respects national privacy laws, GDPR, HIPAA guidelines and applicable legislation regarding the data privacy.

Summary

SOPHiA Exome Solutions are comprehensive genomic applications enabling the detection of multiple types of variants associated with rare and inherited disorders. They enable the assessment of more than 19,000 genes in a single assay by leveraging the advanced analytical power of SOPHiA. As a result, these solutions globally offer a streamlined and standardized workflow that can be easily implemented by any healthcare institution.

References:
[1] *Pfundt R1., et al. Detection of clinically relevant copy-number variants by exome sequencing in a large cohort of genetic disorders. Genet Med. 2017 Jun;19(6):667-675. doi: 10.1038/gim.2016.163. Epub 2016 Oct 27.*

All product and company names are trademarks™ or registered® trademarks of their respective holders.
Use of them does not imply any affiliation with or endorsement by them.