



Instructions for Use

Refers to VarSome Clinical v11.9



© 2019, Saphetor SA



Table of Contents

Table of Contents	3
Explanation of symbols	8
Introduction	9
Result Validation	9
Regulatory Information	9
Device Description	10
Principle of operations	10
Intended Use	13
Indication for Use	13
Warning, caution and precautions	14
Introduction to VarSome Clinical	16
1. Getting help from support and citation	16
Get help via VarSome Clinical	16
Other methods	16
How to cite	16
2. Adding an assay to VarSome Clinical	17
VarSome Clinical customers	17
3. Installation	20
Registration/ Join VarSome	20
Login to VarSome Clinical with 2 Factor Authentication	21
Other 2FA methods	25
Disable 2FA Authenticator APP	28
Log out	28
Management	29
1. Group supervisor	29
2. Storage Management	29
Monitoring the storage space	29
Managing the storage space	30
2. Setting up VarSome Clinical storage preferences	31
3. Deleting / archiving data	32
4. VarSome Clinical Token management	33
SSO platform and Account Administrator	33
Token Definition	33
How to get a Token	33
Issue a token	34
Print a Token and see all past token orders	35
Claim a Token	35
Using the token to run an analysis in VarSome Clinical	37
Token Usage Tracking	37

Version: 11.9.1 - 19th December 2023

Varsomeclinical

Token expiry dates (claim and sample analyses)	38
Token-supported pip	pelines	38
Combined token and	alyses	39
VarSome Premium f	or token-based customers	39
Instructions		40
1. Upload files to VarSor	ne Clinical	40
1.1 Accepted input f	iles	43
1.1.1 Accepted f	ile names for FASTQ	43
1.1.2 Requireme	ents for submitted VCF files	43
1.2 Upload FASTQ fil	es from Illumina BaseSpace.	46
2. Manage		51
2.1 Samples		51
2.1.1 Create and	l define your samples	51
2.1.2 View/Edit	your samples	55
2.2 Workflows		56
2.2.1 Workflow	table	56
2.2.2 Create a ne	ew workflow	56
Main analysis pa	arameters	57
Filters and sub-a	analysis options	58
Lock workflow v	alues	59
3. Launch analysis		60
3.1 New analysis		60
3.1.1 Use an exis	sting workflow	60
3.1.2 Launch an	analysis without a workflow	62
3.1.3 Select sam	ples for analysis	63
3.1.4 Preview ar	nd launch analysis	67
3.2 New cohort anal	ysis (multi-sample VCF)	67
4. Launch sub-analysis		68
4.1 Gene list analysi	S	68
4.2 Algorithmic filter	rs	69
4.2.1 ACMG Act	ionable Genes	72
4.2.2 Carrier risk	for couple	72
4.2.3 Compound	Heterozygous Candidates	73
4.2.4 Compound	Heterozygous for Trios (n=3)	73
4.2.5 De novo (s	trict)	74
4.2.6 De novo ca	andidates (naive)	74
4.2.7 Exonic and	splicing	74
4.2.8 Family trio	recessive (coding and rare)	76
4.2.9 Fisher exac	ct test	77
4.2.10 Genes in	common	77
4.2.11 GWAS Ca	talog	77
4.2.12 Max othe	er samples	77
VarSome Clinical User Manual	Version: 11.9.1 - 19th December 2023	Page 4 of 254



4.2.13 PharmGKB		78
4.2.14 Segregating v	ariants	78
4.2.15 Trio Recessive		80
4.2.16 Variants in Co	mmon	80
4.2.17 VarSome Pick	S	81
4.3 CNV/SV analysis		83
4.3.1 CNV/SV calling	(from FASTQ)	84
4.3.2 CNV annotatio	n (from VCF)	87
4.4 Repeat expansion an	notation from VCF file	89
5. Merging analyses		91
5.1 Family trio analysis		92
5.2 Carrier analysis for a	couple	92
5.3 Merge analyses		93
6. VarSome Clinical Dashboa	rd and Analyses pages	93
6.1 "Analysis actions" op	tions	102
SNVs & Indels		102
Sample phenotype(s	·)	102
Update sample affect	ted samples	105
Sample/Analysis info	ormation	106
Reuse sample files		108
Gene Coverage		108
View Quality Contro	l report	110
Quality Control repo	rt for CNV analyses	116
View FASTQC report		117
Coding coverage rep	ort	125
Coverage report for	targeted regions	126
Region list coverage	report	127
Downloads		129
Download PGX repo	rt	130
Re-annotate Analysis	5	131
New Gene-List Analy	vsis	131
Add BAM file(s)		131
New Algorithmic Filt	er Analysis	131
New CNV Sub-analys	sis	132
New Repeat Expansi	ons Sub-analysis	132
Archive sample data		132
Delete FASTQ sample	e data	132
Delete BAM sample	data	132
6.2 User messages		132
6.3 Tabs		135
7. Variant table		136
7.1 Description of results	s page functionalities	137
e Clinical User Manual	Version: 11.9.1 - 19th December 2023	Page 5 of 254



Custom Germline & Somatic Variant Classifications	139
Potential artifacts	140
Comment on variants	140
Import list of custom classifications and variant comments	143
Classification Nomenclature	144
Clinical Cards	149
Green menu	157
Blue menu	158
7.2 Description of CNV analyses' functionalities	164
CNV variant table	164
Repeat Expansion variant table	171
8. Filter Sets	172
8 1 Create a new filter set	173
8.2 Applying an existing filter to a variant table	176
8 3 Shared Filters	179
9 Gene List	181
9.1 Conv and modify an existing gene list	181
9.2 Create a new gene list	182
9.3 Create a gene list from phenotypes	183
9.4 Browsing through gene lists	184
10 Final Report Generation	186
11 Visualization	198
11.1 Browsers	190
11.1 1 Browser interface settings	190
11.2 CNV Visualizations	202
11.2 CNV VISUAIZATIONS	202
11.2.2 CNV Blot	202
11.2.2 Known CNV/s	205
11.2.3 Known CNVS	210
11.2.4 Read alignment visualization in Jbrowse	211
11.2.5 Browsing through the samples of a Civy analysis	211
11.5 Sample View	211
12. Valsome clinical requently Asked Questions (rAQs)	220
	220
12.2 CIIIVal	224
12.2.1 How call I submit to Cillivar?	224
12.2.2 What is clinvar class?	220
12.3 Explanation of VCF attributes	227
12.4 How can I filter out artifacts from my samples?	230
12.5 why are variants identified in a single strand considered as artifacts, in particu using amplicon kits?	iar when 221
	201
12.0 Foculogenes	202
12.7 Valiant Juling	255

Varsomeclinical

	12.8 Tumor Mutational Burden (TMB)	233
	12.9 Variant not found	235
	12.10 Investigating exon coverage	236
	12.11 Publications Timeline Visualization	238
	12.12 Joint calling	241
	12.13 Variant calling and quality filters	241
	12.13.1 Variant calling local reassembly	245
	12.14 CNVs	246
	12.14.1 Length of CNVs	246
	12.14.2 Gene list analysis for CNV calling	246
	12.14.3 Can I reuse a standard set of control samples to call CNVs?	247
	12.14.4 How is frequency calculated for CNV calls?	248
	12.14.5 CNV calling with WES or targeted panel data	248
	12.15 Why does the position of some indels reported not match the position of the inde	ls
	reported by ExAC/Gnomad?	249
	12.16 Filter by Frequency or Pathogenicity	250
	12.17 Hemizygous Variants	251
13.	Troubleshooting	253
	13.1 Clear VarSome Cookies from the browser	253



Explanation of symbols

REF	Catalog Number		
	Manufacturer		
[]]	Consult Instruction for Use		
	The device complies with European Directive 98/79/EEC		
\triangle	When followed by " CAUTION ", this symbol means "Caution! Failure to observe could result in forfeited time or effort, or the need to abort use of the device."		



Introduction

This document provides general information and procedure to use VarSome Clinical, available by using the following url:

https://clinical.varsome.com/

CAUTION This Instruction For Use (IFU) contains important information for the safe use of this product. Please read the entire IFU before using our product.

Manufacturer:

Saphetor SA EPFL Innovation Park – C 1015 Lausanne +41 21 552 09 77

www.saphetor.com

Authorized Representative

Saphetor SA Greek Branch Leocharous 3 - Athens 105 60 Greece

Result Validation

We can provide detailed validation results for known individuals, whose cells/DNA can be ordered by post:

https://catalog.coriell.org/0/Sections/Search/Sample_Detail.aspx?Ref=GM12878.

The samples should be run through normal laboratory procedures. Then send us the FASTQ files and we will validate the variants against the Genome in a Bottle (GIAB) dataset giving detailed specificity and sensitivity.

Regulatory Information

According to the Annex III of the European Directive 98/79/EEC, VarSome Clinical is considered as an In-Vitro Medical device.

The use of VarSome Clinical is subject to the terms and conditions provided on our website. Please consult them.



Device Description

VarSome Clinical is an online software application that can process the results of Next Generation Sequencing (NGS), generating genetic variation information based on standard guidelines and databases with clinical evidence.

VarSome Clinical is composed of two main modules, with module 2 being subdivided in two submodules:

1) Module 1: Variant Calling

VarSome Clinical utilizes variant calling pipelines for analysis of Next Generation Sequencing data (Illumina or MGI technologies), from whole genome assays, targeted DNA capture-kit assays (e.g. exome, clinical exome) and amplicon kits for germline and somatic DNA.

Variant calling is a computational method used in genetics to detect differences (variants) in DNA sequences and involves comparing an individual's DNA to a reference genome.

2) Module 2.1: Annotation

VarSome Clinical will annotate all variants identified by bringing in known information about them from external databases. This will include, but is not limited to the gene(s) the variants fall in, the known functions of those genes, any publications mentioning the genes or the variant in their title or abstract, any reported clinical information about the variant (e.g. from ClinVar).

3) Module 2.2: Variant Classification

VarSome Clinical will then use the information from the Annotation step above to identify which germline or somatic variant classification rules are applicable to a variant and so reach a variant pathogenicity verdict. VarSome Clinical uses SVC v3 and AMP rules, which are guidelines considered standard in the field for determining pathogenicity for germline or somatic variants, respectively, issued by the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (<u>Richards et al., 2015</u>, <u>Riggs et al., 2020</u>, and <u>Li et al., 2017</u>).

VarSome Clinical provides the user with tools that allow filtering the results according to user-defined criteria and can also propose algorithmic filters for more sophisticated filtering such as identifying de-novo variants in family trio analyses or finding cases of compound heterozygosity.

Principle of operations

VarSome Clinical is accessible using internet browsers including Chrome, Safari or Edge and it is installed either on Saphetor's servers, or in Google Cloud or on the client's premises. The system consists of a number of software tools and databases installed on the servers which are



accessed by the users via a Graphical User Interface (GUI) provided by the VarSome Clinical Web Portal. The software tools are either developed in-house or obtained from third parties.

The scheme reported in Figure 1 is a summary of the analysis pipelines used in VarSome Clinical. VarSome Clinical is intended to be integrated within a routine NGS genetic testing workflow, which begins when a patient initially visits a specialized doctor, and the doctor suspects a genetic disease based on the clinical examination of the patient, or the patient suffers from cancer. The genetic testing workflow is carried out according to the following steps, which include the primary, secondary and tertiary analysis of NGS data (only the secondary and tertiary analysis are performed by VarSome Clinical, therefore the primary analysis is out of scope of this Performance Evaluation):

--- Previous to VarSome Clinical and out of scope of the current Performance Evaluation---

Step 1 – Once the patient is prescribed with a genetic test for the identification of DNA variants in one or multiple candidate genes, the laboratory takes a biological sample from the patient (blood is the most common).

Step 2 – DNA extraction is performed.

Step 3 – The extracted DNA is fragmented and prepared in a sequencing library.

Step 4 - During the **primary analysis**, millions of DNA fragments are sequenced using the NGS machines. The output of this step will be one or more text files ("fastq" files) with the sequence of DNA bases (A, C, T or G) as detected in the patient's sample along with the quality score for each sequenced DNA base. NGS sequencing results in millions of such small sequences ("reads"), many of which will correspond to the same areas of the sample's genome.

Of note, steps 1 to 4 are independent from VarSome Clinical. However, to guarantee a good quality of sequencing data for a clinical setting, steps 1 to 4 shall be performed with CE-marked medical devices.

---- VarSome Clinical ----

Step 5 - VarSome Clinical platform use starts. A specific interface which is available for uploading FASTQ or VCF files (see Section 2.8.3. for the description of the input files) is selected. The user simply needs to select the files to be uploaded. Once the files are uploaded, the user can select the files to be analyzed and launch the test.

Step 6 - During the **secondary analysis**, VarSome Clinical performs an analysis of the DNA reads. This analysis step involves aligning the reads to the reference genome and calling variants, in the case of starting from FASTQ files. Alternatively, the physician may choose to use a different service to perform the variant calling, in which case they would upload a VCF file containing the identified variants to VarSome Clinical and start using VarSome Clinical at Step 7.

VarSome Clinical User Manual

Version: 11.9.1 - 19th December 2023



Step 7 - During the **tertiary analysis**, VarSome Clinical annotates and classifies the identified or provided variants, that can easily be in the order of tens of thousands. First, it annotates the variants based on their possible effects on genes (e.g. truncation of protein, amino acid change, etc) and information gathered from external databases, such as relevant publications, evidence on their effect of a specific phenotype, clinical trials, etc. VarSome Clinical uses information from over 140 databases to provide the most comprehensive set of annotations. These data from external databases are first gathered and stored in Saphetor's server to create an integrated database (Molecular DB). You may find the list of databases we have integrated in VarSome on the following link https://varsome.com/datasources/.

Secondly, based on the annotated information on the variants, VarSome Clinical follows a set of rules and guidelines to classify the variants with respect to their pathogenicity (i.e., their likelihood of affecting the physiological state of an individual, which can translate to disease in some cases). For germline variants, the SVC v3 guidelines are followed, and variants are classified as either "Benign", "Likely Benign", "Likely Pathogenic", "Pathogenic" or "VUS" (Variant of Unknown Significance). For somatic variants, the AMP guidelines are followed, and variants are classified as Tier 1, Tier 2, Tier 3 and Tier 4 according to their implication in cancer (from clear to little evidence, respectively). Of note, CNVs are classified based on specific CNV guidelines (Riggs *et al.*, 2020).

Step 8 (optional) - Once the variant calling and/or the variant annotation/classification has finished, an email is sent to the user, who can then access the results through the user interface. Once the sample analysis is finished, the user can access the detected genetic variants and the quality reports associated with the sample and the analysis. At this point, the user can manually adjust each of the classification criteria based on prior information (such as a variant's de novo status, patient's and their family's health background and phenotypes, etc.) or their own domain knowledge to reach a final interpretation. To filter out the irrelevant variants, the user needs to design the filters that will be applied to the variant table. The system provides a built-in editor which can be used to generate a report .

Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8
Ē.	- ACA			Varsomecinica	•		
Biological sample collection Blood Saliva Tumour biopsy	DNA extraction	Library preparation -Partial genomes (Capture or amplicon kits) -Whole genomes	High-throughput sequencing (NGS) Primary analysis	VarSome Clinical platform use	Read alignment & Variant calling (Varsome Clinical) Secondary analysis	Variant annotation & classification (VarSome Clinical) Tertiary analysis	Report

Figure 1. Genetic test laboratory workflow. Steps 5, 6, 7 and 8 are supported by VarSome Clinical, which include the secondary and tertiary analyses, and report generation (optional).



Intended Use

The VarSome Clinical platform aids in making an informed decision and diagnosis, and in the selection of appropriate treatments for cancer, Mendelian and other genetic disorders by processing next generation DNA sequencing data.

VarSome Clinical utilizes variant calling pipelines for analysis of Next Generation Sequencing data (Illumina technologies), from whole genome libraries, DNA capture-kit libraries (exome, clinical exome) and amplicon kits for germline and somatic DNA. The Clinical platform identifies and reports SNVs, CNVs, indels and substitutions, as well as variant annotation to facilitate the variant curation and classification. Multi-sample analyses are also available, to perform joint-genotyping to generate a multi-sample VCF file with the merged genotypes for two or more samples. VarSome Clinical can also propose algorithmic filters for inheritance hypothesis testing.

Indication for Use

VarSome Clinical offers clinicians and researchers a solution to assist in making an informed decision and diagnosis on the selection of the best therapies for the treatment of patients of all ages for the following conditions:

- Patients suspected to be affected by heterogeneous mendelian diseases or undiagnosed (rare) diseases.
- Patients with common diseases that are in fact a group of rare diseases with eclectic phenotypic features, being caused by mutations in specific genes, such as developmental delay or intellectual disability, degenerative neurological diseases, immunodeficiency diseases, cardiomyopathies, etc.
- Patients with hereditary cancer syndromes.



Warning, caution and precautions

CAUTION When using NGS to investigate the cause of a disease, clinicians and patients must be aware that the test may cover many genes, most of them unrelated to the patient disease. This would lead to discovery of secondary or incidental findings on genes causing other diseases, e.g. cancer predisposition status (Kuhlen and Borkhardt 2015). Recently, policy statements from two major regulatory agencies (AMP and ACMG) have provided a list of recommendations (Green et al. 2013; Allyse and Michie 2013; ACMG Board of Directors 2015; Hegde et al. 2015). However, each country has its own directives about incidental finding reporting. Therefore, it is extremely important that the patient, or their legal representative, is informed of the incurred risks and that the patient's decision is recorded in an "informed consent" document (Anderson et al. 2016). It is the responsibility of the treating clinician to collect the informed consent form before ordering any NGS test, and ensure that the patient understands the potential outcomes of the test.

Users should be warned that the result reported shall not be used alone to modify the patient care.

Because certain chemotherapeutic agents can cause DNA damage, patients undergoing chemotherapy treatment may have variants created by the chemotherapy process itself. Such variants can be included in the results of VarSome Clinical.

There is a risk of misclassification of variants recently discovered and not documented enough in the scientific literature.

The DNA sequencer used to generate these data shall be CE-marked to guarantee the minimum standards of data quality.

The following warnings are sub-divided according to the VarSome Clinical modules: Variant calling and Variant annotation/classification.

Variant-Calling

- Large-scale genomic rearrangement analysis (structural variants, including CNVs, whole-chromosome deletions etc.) may not be guaranteed because such genomic events cannot be reliably detected from short read next generation sequencing data. See, for example, Tan R. *et al., Hum Mutat.* 2014 Jul;35(7):899-907. doi: 10.1002/humu.22537 and references therein). This is a limitation of NGS sequencing itself, caused by the short (~150-200bp) length of the reads, and not of the VarSome Clinical device.
- Performance was not validated specifically for hard to detect variants (those falling in regions of segmental duplication or other hard to sequence and/or process regions of the genome) because the system does not provide special tools for such variants.
- The performance of VarSome Clinical cannot be guaranteed for samples with very low concentrations of tumor cells.

Variant Annotation & Classification

VarSome Clinical implements three automated classifiers according to the ACMG & AMP



guidelines.

The classification results obtained with VarSome Clinical aim to reflect common expert consensus. The user shall be aware that the results generated after changing the classification settings are not part of the in vitro medical device scope and are therefore not validated.

The classifiers implement an automated evaluation of the **generic** guidelines for ACMG, AMP and CNVs. These automated classifications cannot substitute for the knowledge, experience and additional case information available to a human curator and are limited by the availability of machine-legible data.

Only the "generic" ACMG classification rules have been implemented, specifically, **disease-specific guidelines have not been implemented**. These typically require further adjustments to the various classifiers, rules, thresholds, exception lists etc. However, all the ClinGen expert panel reviewed variants are available in VarSome, visible in the ClinGen tabs (these entries are also replicated by ClinVar).

Please see the online documentation for the full detail of how the classifiers have been implemented in VarSome:

- ACMG germline: <u>https://varsome.com/about/resources/germline-implementation/</u>
- **AMP somatic:** <u>https://varsome.com/about/resources/somatic-implementation/</u>
- CNVs: <u>https://varsome.com/about/resources/sv-implementation/</u>

Please note that a visible discrepancy between the two platforms may be encountered since VarSome is updated prior to VarSome Clinical. There are two reasons why a variant may have a different pathogenicity annotation on VarSome Clinical and VarSome: updates in the databases utilized to infer the pathogenicity and improvements in VarSome's Germline Variant Classification..



Introduction to VarSome Clinical

1. Getting help from support and citation

Get help via VarSome Clinical

In case any help is needed either to understand how VarSome Clinical works or if you are facing a problem, it is possible to contact our support directly from VarSome Clinical from the Help menu option.



A new window will be displayed with the subject and main text field where you can describe your problem or request help. It also allows you to attach a document. Our support team will reply to you via email.

Support Having studule? Contact our support	Dashboard / Contact support
Support	
Fill in the contact form below and inform us of any issue or problem you experience with our applications. Ideas also for improvement are more than welcome.	User manual
When asking about a specific analysis, please mention the sample name or the URL of the results when contacting us.	Download manual
Subject	
Subject	
Your message	Help docs
Your message	Go to the varSome Help Center
	Make your own assay request
	new Assay Request
optionally select up to '5' files (max 10240.0 Kbytes each) to attach. Allowed extensions: ['gzlp', 'zp', 'tar', 'gz', 'tar', 'grg', 'bz', 'tar', 'grg', 'jpg', 'lif', 'tif', 'webp', 'bmp', 'ppm', 'pdf', 'doc', 'd	
Emulya apyziewi atv. erinegrafiek kareva apyzio.	
Submit	

Other methods

It is possible to reach out the support team directly by email (<u>support@varsome.com</u>) or by filling out the help center form (<u>https://docs.varsome.com/en/kb-tickets/new</u>). Please provide detailed information on the question or the problem you are facing as well as screenshots that may help us identify and reproduce the issue.

How to cite

If you use VarSome or VarSome Clinical for your work, please cite our paper in your publications and other communications:

VarSome: the human genomic variant search engine, Christos Kopanos, Vasilis Tsiolkas,



Alexandros Kouris, Charles E Chapple, Monica Albarca Aguilera, Richard Meyer, Andreas Massouras, Bioinformatics, Volume 35, Issue 11, 1 June 2019, Pages 1978–1980, <u>https://doi.org/10.1093/bioinformatics/bty897</u>

You can find VarSome citations on the following link: <u>https://varsome.com/about/general/varsome-citations/</u>.

2. Adding an assay to VarSome Clinical

VarSome Clinical customers

If you are a VarSome Clinical customer and your assay is not yet available in VarSome Clinical, we can add your assay if it is an assay we can support.

If you need us to add a new assay, the first question is whether this is a standard assay or if it requires special handling. Examples of special handling include but are not limited to assays with Unique Molecular Identifiers (UMIs) (of those, VarSome Clinical currently supports only Paragon CleanPlex UMI Lung Cancer Panel), assays that are designed to produce mixed RNA and DNA sequencing data, assays targeting specific classes of variants such as certain large deletions in genes like CFTR and various other special cases. If you do not know if an assay requires any special treatment, please talk to your assay providers and ask them if any special bioinformatics approaches are required to handle the data the assay will be used to generate.

You can now request to add your assay of preference by filling the corresponding form which is available in the following links depending on the installation you are working from: https://ch.clinical.varsome.com/new-assay-request/ https://eu.clinical.varsome.com/new-assay-request/ https://us.clinical.varsome.com/new-assay-request/.



ew	Assay Request
	Please fill in the form below to request the inclusion of a new assay.
	Note: This option is not available for free trial users.
Ass	ay Manufacturer
A	ssay Manufacturer
Ass	ay type
0	Capture kit
0	Amplicon kit
ls y	our assay commercially available or custom made for you?
0	Commercially available
0	Custom made
Ref	erence genome
0	hg19
0	hg38
Do	is the assay have UMIs?
0	Yes
0	No
Ass	ay name to be shown in VarSome
A	ssay name to be shown in VarSome
Wh	at File(s) are you uploading?
0	Manifest Bard
ŏ	Bed and Bedpe
-	
Set	eccupico 2 riles (max tuomo each). Autowed extensions: [gzip, zp, rz, rar, gz, ozz, tar, bed, bedpe, manirest, txt, zst]
	oose Files
wne	in requesting an amplicon kit, please make sure to either provide a manifest rule or, if providing a bed rule, make sure to also provide a bedge rule with the locations of the primers.
Are	there any special bioinformatics steps needed to handle data generated by this assay? If you do not know, please contact the assay manufacturer.
A	re there any special bioinformatics steps needed to handle data generated by this assay? If you do not know, please contact the assay manufacturer.

On this form you will need to fill the requested fields:

- The assay's manufacturer name
- Whether it is a capture or amplicon-based assay
- Whether it is a commercially available or custom made assay
- The version of reference genome (hg19 or hg38)
- If the assay includes Unique Molecular Identifiers (UMIs)
- The name of the assay to be shown on Varsome Clinical
- The type of files you are uploading (manifest, .bed or .bed and .bedpe)
- Choose the corresponding files
- Specify if any special bioinformatics approaches are required to handle the data the assay will be used to generate.

Please keep in mind that for assays targeting relatively small regions of the genome (i.e. assays that are neither whole-exome nor whole genome) variant calling will be performed in <u>targeted</u> <u>mode</u>, only looking at the regions mentioned in the assay.



If you ask us to add a new assay targeting mitochondria on hg19/GRCh37 reference genome, please use the rCRS "MT" coordinates in the accompanying BED file.

If your assay is not listed, you can choose the "Generic capture kit", for the case of capture-based libraries.

Launch Analysis			
	2	3	4
	Workflo	w Details	
	Analysis from:	VCF OFASTQ	
Analysis type 🛈			
Eg: Germline - Single Sample	~		
issay ()		Keep variants (j)	
Generic capture kit	~	All variants	× ~
Generic capture kit Illumina	× ~	Targeted Mode 🔅	
Genome (i) Inheritance from (i)		Ethnicity (i)	
▶ hg19 ◯ hg38 ● All ◯ OMIM		Not specified	~
pply filters (1)			
Search filters			~
Run a gene list sub-analysis 🛈			
Search gene list			~
Run a gene list sub-analysis based on sample phenotypes ①	Show only gene lis	st results ()	
Run algorithmic filter sub-analysis ()			
Search algorithmic filters			· · · · · · · · · · · · · · · · · · ·
Evill			Drevieus

As long as you choose the right chemistry type as described above, the results will be the same as if we had added your assay. The only differences will be that we will not give you statistics about the percentage of reads that fall on target and we will not run the variant calling in <u>targeted mode</u> (since we do not know the targets) so you may see variants that fall outside your regions of interest. However, the results will be correct and reliable.

The correct choice of the chemistry type is important because it affects:

- The way we handle possible PCR duplicates and
- The default thresholds set for minimum coverage needed to call a variant. For samples sequenced using an amplicon kit, since such kits tend to have very high coverage, the minimum threshold coverage is higher than for the capture kits.



▲ Please, note that the samples analyzed using one of the "Generic" kits can not be used for CNV sub-analyses.

The assay's information will be used to calculate alignment statistics and coverage of the targeted regions. The statistics are used to generate the <u>QC report</u>.

If you want to run a whole genome analysis you need to select the 'WGS' option when launching the analysis, as shown on the pic below. This option will guarantee the sample will be available for subsequent CNV analysis, if you wish to perform it.

Analysis type (i)	
Eg: Germline - Single Sample	~
Assay (i)	7
wgs	~
WGS pcr-free WGS -PCR	
WGS with pcr amplification WGS +PCR	
Genome (i) Inheritance from (i) hg19 hg38 All OMIM	

Please note that as VarSome Clinical can process any kind of NGS data, the list of Assay is very long. To eliminate the need to scroll through the long list of Assay whenever you are launching an analysis, you or your supervisor can simply mark certain assays as favorite. Remember to check your settings for favorite assays.

3. Installation

VarSome Clinical is a web based software that connects to our servers to run all the analyses. It can be used from different browsers such as Google Chrome, Firefox, Microsoft Edge and Safari. In order to have access to VarSome Clinical a subscription is required. If you want to get a trial or acquire a subscription please contact <u>sales@varsome.com</u>.

VarSome Clinical is available from 3 different servers, 2 from Google Cloud, one located in Germany (EU) and the other in the United States (US). The last one is on a server that is located in Switzerland (CH).

Registration/ Join VarSome

In order to be able to have access to VarSome Clinical either with a subscription or a trial you have to register to <u>https://varsome.com/</u> first.

On the top right click on Join and a new window will be displayed where you can provide your contact details, Email address, First name and Last name.





Email address*

Please enter your work email address if you work in life sciences or healthcare.

First name*

Last name*

Password*

Password should be at least 8 characters long, and contain at least 1 letter and 1 digit.

This site is protected by reCAPTCHA and the Google Privacy Policy and Terms of

Service apply.

We will deactivate your account and may block your IP address if you enter false information.



An email will be sent confirming the registration. Now you can request your VarSome trial / subscription to our business team.

Login to VarSome Clinical with 2 Factor Authentication

Once someone from our account team confirms that your trial / subscription has been activated then it is possible to log in to VarSome Clinical. The activation of your VarSome Clinical could be either on the CH, EU or US server.

You will only be able to log in to one of the servers:



https://ch.clinical.varsome.com/ https://eu.clinical.varsome.com/ https://us.clinical.varsome.com/

Once you open the VarSome Clinical web page you will see the following sign in information.

Varsomeclinical	
Welcome! Please log in to access our services	
This installation is on Google cloud servers (US)	
Sign in using VarSome SSO	

By clicking on *Sign in using VarSome SSO* you will be redirected to our SSO (Single sign-on) page and you can now provide your user name (email address) and password which you used to register to varsome.com.

	Varsome	
Login Sign In to your account		
Username Password Login		Forgot password?
Varsome	Varsomeapi	Varsomeclinical



The next step to complete the log in, is to set up the two-factor authentication, which is mandatory for new users, by providing a mobile phone number.

Please note that we do not accept numbers generated through a service that provides temporary telephone numbers. If such a number is used, the following error message will be displayed:

"This mobile phone number does not pass our system's security checks. Please use another number."

	Varsome		
Action requ	uired		
Activate two-factor au VarSome community. F	thentication. This improves account sec For more information click here.	urity and prote	cts the
Enable two-factor	authentication via SMS		

After clicking Submit, a text message with a code should have been sent to the provided mobile number. Once you receive the code this has to be given in the next step in order to complete the log in.

	•	Varsomeclinica	l i
		Varsome	2
Actio	n requi	ired	
Activate t VarSome	wo-factor authe community. For	entication. This improves account s more information click here.	ecurity and protects the
SMS			
Enter the coo	le received via SM	IS in the field below.	
Enter the c	ode received via	SMS on	Submit
Resend SMS	Provide anothe	r phone number	
Vars	ome	💙 varsomeapi	Varsomeclin

Everytime you log in to VarSome Clinical the Two-Factor Authentication method will be required. However, it is possible to remember the browser for 30 days and during that time you will not be asked for authentication.

	Varsomeclinica					
Varsome						
Two-factor a	authentication					
	da	Culture It				
Enter the authentication co	de	Submit				
Enter the authentication co Remember me for 30 days Resend SMS You need to provide the cod	e within 2:56	Submit				
Enter the authentication co Remember me for 30 days Resend SMS You need to provide the cod	e within 2:56	Submit				

Other 2FA methods

It is possible to enable 2 additional authentication methods, by using an application or saving backup codes. To enable the other methods there is a need to log in to <u>sso.varsome.com</u>, your VarSome account manager page. To land on the same page from inside VarSome Clinical, click on your name at the top right and then on Profile. This will redirect you to your VarSome account manager.

Tags About -						
💄 Profile 🛛 🚽						
··· Preferences						
ClinVar Submissions						
Illumina BaseSpace						
III Claim Assay Token						
Lock Session						
🔓 Logout						



Once logged in to the VarSome account manager, on the top right of the page there is a shield icon.



When clicking on it the user will be able to configure their privacy setting, such as change password, enable other authentication methods or see the API token.



After enabling the application method you can scan the QR from e.g. Google Authenticator and verify it by providing the code displayed.



Please note that when using the Authenticator App, no SMS will be sent automatically. Instead, you should enter the code provided by the Authenticator App when prompted by VarSome Clinical. You may still request an SMS code to be sent by clicking on the "Receive code through SMS" link:



Alternatively, users can generate backup codes that will be shown once and can be copied and pasted to a secure place. Only 3 codes will be displayed and once you use those you can generate new ones.





The backup codes should be used as the last option in case the user loses their phone or does not have access to the authenticator App.

Disable 2FA Authenticator APP

The Authenticator App can be disabled from the VarSome account manager by clicking on the "Disable two-factor authentication using application" button.

A Please note that it is not possible to disable 2FA entirely, so the the SMS option cannot be disabled.

Two-Factor Authentication						
SMS						
You have enabled two-factor authentication via SMS to +34647060618.						
Authenticator App						
You have enabled two-factor authentication using an application.						
Disable two-factor authentication using application						

Log out

Once you finish working with VarSome Clinical you can log out from the platform by clicking on your name on the top right corner and then Logout.

Tags About -							
💄 Profile							
••• Preferences							
ClinVar Submissions							
Illumina BaseSpace							
IIII Claim Assay Token							
Lock Session							
🔓 Logout 🛛 🚽 🛶							



For security reasons VarSome Clinical logs you out after 15 minutes of inactivity.

Management

1. Group supervisor

In VarSome Clinical it is possible to set up a group supervisor for your account. The group supervisor is a VarSome Clinical user with certain privileges that other users from the same group will not have.

This user will have the ability to create, change and track:

- Workflows: create workflows with fixed parameters which can be used by you or other colleagues to launch analyses using pre-defined parameters.
- Launch analysis without a workflow: the supervisor can limit this option so that users can only launch analyses defined by him or her.
- Assay preferences: select an assay of a list of assays so that those are shown in the first place when someone is selecting an assay for launching an analysis.
- Storage preferences: as a group supervisor, you can modify the storage preferences of your group and decide how much time you would like to keep FASTQ or BAM files for example. Find more information <u>here</u>.
- Analysis preferences: you can enable the sensitive mode for CNV calling and decide if VarSome Picks should run automatically.
- <u>Audit trail</u> access: this keeps a record of different actions performed by the users of your group.

If you do not have a group supervisor yet or would like to change the group supervisor role to a different user, please contact our <u>support team</u> to change this.

2. Storage Management

Storage fees are calculated at the end of each month, based on the volume of data that has been uploaded into VarSome Clinical in order to run analyses. They reflect the storage costs of Saphetor data centers or the storage costs charged by 3rd parties (like GCP - Google Cloud Platform) to Saphetor.

If a new sample has been uploaded during a given month, the storage costs for that month will be prorated based on the day of the upload.

Saphetor is not responsible for monitoring or managing the storage cost of any customer. This should be done by each end customer, possibly in coordination with the local distributor.

Monitoring the storage space

In order to check the volume of data that will be charged, each Account Administrator can visit <u>sso.varsome.com</u> (using his/her institutional user email address) and click on the Billing menu: VarSome Clinical User Manual Version: 11.9.1 - 19th December 2023 Page 29 of 254



This window contains the monthly storage reports, in pdf and xls format, as well as the most recent billing entries, invoices history and account statements. The account administrator can then either filter on specific files or choose from the list.



Managing the storage space

There are 5 different options to possibly limit the volume, and hence the related costs, of samples data storage:

- 1. Keep all the data
 - The Storage Fee applies to all the data available in the client's account. This is the default option and needs no user action to be applied.
- 2. Remove FASTQ; keep the BAM and results.
 - Space occupied is lower than option 1 by approximately 25%.
- 3. Remove BAM; keep the FASTQ and results.
 - Space occupied is lower than option 1 by approximately 50%.
- 4. Remove all the raw data (FASTQ and/or BAM); keep the results only.
 - Space occupied is approximately 75% lower than option 1, as the annotation data doesn't take up much space.
- 5. Archive the samples
 - No Storage Fee applies.
 - Samples results/annotations not available for browsing in VarSome Clinical.
 - The archived sample is still cross-referenced with other samples and sample links are still shown in the tab with cross-referenced samples.
 - Custom Variant Classifications set up initially for the archived sample will stay available when browsing other active analysis. The phenotypes assigned to the archived sample are still available in the sample information.
 - All sample data are deleted and not available anymore. Only the VCF file is stored for possible sample re-annotation when unarchiving the sample.



- The client may unarchive the sample. When doing so, the re-annotation fees will apply. For more details, please contact sales@varsome.com.
- When un-archiving the sample, VarSome Clinical will annotate the sample (VCF file) against the latest annotation data, which may cause annotation and classification differences between the original sample and the unarchived sample.

2. Setting up VarSome Clinical storage preferences

The storage preferences within VarSome Clinical can be modified in order to define for how long the user wants to store the FASTQ and BAM files and when they want their analyses to be archived.

In the VarSome Clinical platform, hover over your username on the right top of the window, then select "Preferences":

Tags About -						
Profile						
••• Preferences						
ClinVar Submissions						
Illumina BaseSpace						
Claim Assay Token						
Lock Session						
Logout						

<u>Only the group supervisor will have permission to modify these preferences</u>. If you do not have a group supervisor yet, please contact <u>support@varsome.com</u> and let them know who should be your group supervisor. Once the group supervisor has accessed to the Preferences menu, they will be able to set up the storage preferences for each type of file:



Keep FASTQ files for	
1	
unit	
Years	~
Keep BAM/BAI files for	
1	
unit	
Years	~
Archive sample files after	
1	
unit	
Years	~

3. Deleting / archiving data

Within VarSome Clinical it is also possible to immediately delete some FASTQ and BAM files, or also archive some analyses, without waiting for the time period defined previously.

For each single sample analysis that was launched from a FASTQ file, the user can click on the three horizontal bars on the right side of the sample name. They will then find the following options:



SNVs & Indels
Sample phenotype(s)
Sample analysis information
Reuse sample files
Gene coverage
View QC report
View FASTQC report
Coding coverage report
Region list coverage report
Downloads
Coverage report for targeted regions
Archive sample data
Delete FASTQ sample data
Delete BAM sample data
Re-annotate analysis
New Gene-List analysis
New algorithmic filter analysis 🥌
New CNV Sub-Analysis
New Repeat Expansion Sub-Analysis
E Description:

4. VarSome Clinical Token management

SSO platform and Account Administrator

Saphetor has developed an SSO platform on VarSome (<u>sso.varsome.com</u>) that can be used by the Account Administrator of the distributor to facilitate the usage of VarSome. To use this application, one unique user from the distributor must be defined as the Account Administrator. This user is set up by the Saphetor Billing Department.

Token Definition

A token is a form of prepayment, an activation code for VarSome Clinical, which upon claiming, permits the client to analyze a certain number of samples belonging to a certain NGS assay.

Some tokens are for somatic and some for germline analyses.

Tokens support only FASTQ files.

How to get a Token

VarSome Clinical User Manual



Tokens are issued by our partners, mainly distributors and assay manufacturers. <u>Get in touch to</u> <u>learn more</u>.

Issue a token

First, the token specifications for the user group must be set up by the VarSome Support Team (e.g. implementation of the token and setting up the group's account to access specific assays on the Clinical back-end). Also, 1 please note that for this to work, the Account Administrator of the distributor ordering the token with specific assays should have access to VarSome Clinical, because assays are fetched from the Clinical platform.

Once all of the above actions are completed, the Account Administrator of the distributor needs to follow the steps below to issue a token:

i. The Account Administrator must go to <u>sso.varsome.com</u> and click on the Billing menu:



iii. Click on the screen "issue tokens", type the assay and the type of sample (germline or somatic), the storage included in months, the number of samples and the number of tokens. Then, press order:



Assay	Type of sample		
\$	Number of tokens		
lumber of samples			
Number of samples per token	1		
Purchase number	Reference		
Purchase order number (Optior	Your reference (Optional)		
itorage included for			
0	Months 🗢		

Print a Token and see all past token orders

On the next screen, your token will appear, as shown below, as well as all the tokens you have ordered so far. In the future, you will be able to print your tokens by pressing the button "print".

On this screen, you will also be able to see all your token orders listed and filtered by assay.



Claim a Token

Claiming a token means activating it. A token must be claimed by the end customer in order for an analysis to be run.

1. Claim a Token in SSO

Any user from the end customer needs to log into <u>sso.varsome.com</u>, as it drives more accurate results than using the VarSome Clinical platform, and click on Labs. They will see all the tokens that have been purchased but not yet claimed and can claim them by entering the token code on the Claim token window (on the right of the page) and clicking on the "Claim" button:



Но	me / Labs										
	Tokens 10 rows Search token					Cl	Claim Token				
	Token	Assay	🔶 # Samples	\$ Samples Used	Reference	🔶 Issued at	Use until	Tol	ken Enter the token you received from y	Reference Your reference (optional)	
	There are no records matching your request						а	aim	Í,		
							1 > »				

2. Claim a Token in VarSome Clinical

The user needs to log into VarSome Clinical and click on their username on the top right, and then choose the Claim Assay Token option from the drop-down menu.

Tags About	÷
💄 Profile	
··· Preferences	
ClinVar Submissions	
Illumina BaseSpace	
III Claim Assay Token	-
Lock Session	
🔓 Logout	

Once the token has been received, the user will need to claim it to activate VarSome Clinical for the corresponding analysis.

Help					Tags	About +	
						Dashboard / Claim Token	
			Claim Token		Deference		
	Issued at	Use until	NGI1-BAZE-1PA9	~	test token	~	
ſ				Claim			


Using the token to run an analysis in VarSome Clinical

- Single Sample Analyses

Once a token is claimed, no further action is needed in order to run analyses on VarSome Clinical. The platform, however, will validate whether a user attempting to run an analysis possesses non-expired tokens for the specific assay and sample type (germline fastq, somatic fastq) to be run.

Once an analysis is initiated, the system will deduct the number of analyses initiated by the user from the number of analyses available in the token. For example if a token allows a user to run ten (10) germline analyses from fastq and the user runs six (6) analyses, four (4) more analyses will be available for future use of the token.

- Multi Sample Analyses

For multi sample analyses the same validation occurs as with single sample analyses.

Once a multi-sample analysis runs, the number of available analyses for that token will be reduced by the number of samples included in the multi-sample analysis. For example, if a token allows a user to run ten (10) germline analyses from fastq and the user initiates a multi sample analysis from fastq files composed of three (3) samples, seven (7) more analyses will be available for future use of the token.

Token Usage Tracking

Here you can find all tokens you have claimed, along with the number of analysis remaining for each of them and the number of analyses performed.

i. Token Tracking in SSO

The account administrator (from the distributor) or any end user must log into <u>sso.varsome.com</u>, as it drives more accurate results than using the VarSome Clinical platform, and click on the Partners tab.

All the tokens that have been bought will show on this page together with their actual status in terms of the number of analyses used and remaining, the price of the token per sample, the date of issue, the claiming date and the expiration date. There is also a price per sample list for all assays.



(Manufacts	rers / Suppliers)									
5 rows	Search toke	n		Search	by assay			OPin	Issue tokens	
n	Assay	¢ # samples	<pre></pre>	Reference	Claimed by	∲ ^{Issued}	¢ Claim until	o Use until	Assay	Type of sample
E7LW-	Repromeda MIP Carrier FastQ (Sermine)	200	€18.00	M022098672		09 May 2022	09 May 2023	10 May 2023	Number of samples	Number of tokens
-QAJR-	CLG_CarrierOnco-v3 FastQ(Sermine)	56	€ 36.00	- francision -	- Constitution of	04 May 2022	04 May 2023	05 May 2023	Purchase number Purchase order number (Optional)	Reference Your reference (Optio
V9DP-	Cardio v4.1 FastQ (Germine)	53	€ 36.00	NG20988501	-	29 Apr 2022			Storage included for	
N-FO4W-	Accel-Amplicon EGFR Pathway Panel FastQ(Senatic)	48	€38.00	10220000		06 Apr 2022	06 Apr 2023	06 Apr 2023	•	Months
PCNP	Accel-Amplicon EGFR Pathway Panel FastQ (Sematic)	48	€38.00	Management	inere la seconda	06 Apr 2022	06 Apr 2023	06 Apr 2023	Assay Token Pricing	
					E	e 1	2 3 4			Pricing / Samp
									Assay	FASTQ FASTQ (Germline) (Somati

ii. Token Tracking in VarSome Clinical

Claim an assay	token						Dashboard / Claim To	iken
Tokens 5 rows	Search token						Claim Token	
Token	Assay	# Samples	Samples Used	Reference	Issued at	Use until	Enter the token you received from your su; Your reference (optional)	
NDI1-BAZE-1PA9 FastQ (Germline)	Swift Biosciences Acel-Amplicon 56G Oncology Panel	48	5	test token 123	17 Apr 2020		Claim	
			ſ			1 > >		

Token expiry dates (claim and sample analyses)

The tokens have two (2) different expiry dates: one for the claim and one for the samples' analyses.

- 1. Token Claiming expiry date: twelve (12) months after the date of issue. This means that the end customer must claim the token within twelve (12) months of the token's issue date. If a token has not been claimed within twelve (12) months after it was purchased, it will be lost.
- 2. Analyses Expiry date: twelve (12) months after the claiming date. This means that the token is valid for twelve (12) months after it has been claimed. If some analyses linked to a token have not been used within twelve (12) months after the token was claimed, they will be lost. A billing entry is generated upon token issuing, regardless of whether the token will be claimed by the client or not.



Tokens only support analyses starting from FASTQ files. They can be issued for either somatic or germline analyses. By default, somatic tokens include AMP classification, which triggers extra charges. If a customer doesn't want to pay for this AMP classification, they must contact billing@saphetor.com to remove this default feature from the customer's account.

Tokens cannot be issued as combined pipelines (somatic and germline).

When a token is used, it cannot be converted to another assay or another version of the same assay than the one it was already issued for. For example, if a customer has ordered a germline token, has claimed it and started to use it, it could not be converted into a somatic one, in case they realized that it was ordered by error. The only way that this could be done is before the token is used, ask the Saphetor team to delete it and issue a new token instead. In such a case, the fact that it was claimed doesn't prevent the Saphetor team from deleting it.

Combined token analyses

Tumor normal analysis (paired analyses)

Tumor-normal analysis definition

The user has two different samples obtained from the same patient: a tumor sample and a "normal" (healthy) sample. The pipeline will report somatic variants observed only in the tumor sample.

For a tumor normal analysis, 2 tokens of the same assay must be used (a somatic token and a germline token) in conjunction, in order to obtain the desired results. The customer must buy both tokens (germline/somatic) of the same assay.

VarSome Premium for token-based customers

VarSome Premium is normally not available for commercial labs. If a token-based customer was exceptionally allowed to subscribe to both VarSome Premium and VarSome Clinical, they will not get any free monthly analyses on VarSome Clinical (up to the monthly value of the Premium subscription). The reason for this is that in our customer billing platform, we set the monthly credit limit to zero to avoid analyses that would go beyond the token amount and hence create extra charges. On the other hand, a VarSome Clinical account as a stand alone solution is set up with no monthly balance limitation and this is why it can be perfectly combined with the VarSome Premium product which allows limited free access to VarSome Clinical up to the monthly value of the Premium user.

▲ Please note that we developed the tokens solution mostly to answer distributors' specific needs to bundle assays and analyses. These tokens have proved to be quite complex to handle, especially when they are used outside of the most common analyses, so please contact Saphetor for any non-standard request from a token-based account.



Instructions

The sequence of actions to analyze a sample in VarSome Clinical are the following:

- Upload files to VarSome Clinical
- Create or define the samples
- <u>Create a workflow (optional)</u>
- Launch a new analysis: select a workflow/use blank and connect it to the samples to analyze

In the following sections you will find the detailed information on how to perform each of the steps.

1. Upload files to VarSome Clinical

- 1.1 Accepted input files
- 1.2 Upload FASTQ files from Illumina BaseSpace

To upload files to <u>VarSome Clinical</u>, click on **Upload/View files**. Then, click on "Select File(s)" and, in the file browser window that will appear, select all the files you want to upload. Files do not need to be from the same sample or in the same format. Once all files have been selected, the file names are displayed under the green "Select File(s)" button. To upload the files, click on "Start Upload".



Upload file(s)

Upload file(s) from one or more samples Supported file extensions: vcf, vcf.gz, vcf.bz2, fastq.gz, fq.gz, bam VCF files that include small variants and/or CNVs and/or STRs will be split to three separate files, one for small variants, one for CNVs and one for STRs. Lines with missing value on the alternate base(s) field (ALT), if any, will be filtered out from the uploaded VCF file(s). We can parse the paired-end reads according to the Illumina or MGI convention. Find below some examples: Illumina SampleName_S1_L001_R1_001.fastq.gz and SampleName_S1_L001_R2_001.fastq.gz or SampleName_S1_L001_1.fastq.gz and SampleName_S1_L001_2.fastq.gz MGI 12345_L02_48_1.fastq.gz and 12345_L02_48_2.fastq.gz or 6789_L02_56_R1.fastq.gz and 6789_L02_56_R2.fastq.gz E12345_L01_34_4321_1.fastq.gz and E12345_L01_34_4321_2.fastq.gz	
Please find more details here.	
+ Select File(s)	
Go to Sample Definition	
short_L001_R1_001.fastq.gz	
short_L001_R2_001.fastq.gz	3
Start upload	

Files are uploaded and checked, and the number of reads and bases in reads is calculated and displayed for each file. Files can be deleted before or after upload. Files with status set to dark green can be used for a subsequent analysis.

Upload your sequencer or vcf files					Dashboard / Upload	l sample files
Upload file(s)	Upl	oaded files				
Upload file(s) from one or more samples Supported file extensions: vcf, vcf.gz, vcf.bz2, fastq.gz, fq.gz, bam VCF files that include both small variants and CNVs will be split to two separate files, one for small variants and one for CNVs. Lines with missing value on the alternate base(s) field (ALT), if any, will be filtered out from the variants are file(s).	3	_L01_9_2.fastq.gz	11 Sep 2023	•	Number of reads: 100000 Number of bases in reads: 7552007	•
select File(s)	4	_L01_9_1.fastq.gz	11 Sep 2023	•	Number of reads: 100000 Number of bases in reads: 7552007	۲
Go to Sample Definition	5	test_multi_cohort.vcf.gz	11 Sep 2023	•	Number of variants in vcf: 15	
Start upload	6	cohort-n-31.vcf.gz	11 Sep 2023	•	Number of variants in vcf: 15	۰.
		multsatha.vcf.gz	8 Sep 2023	•	Number of variants in vcf: 3	•
		default_germline_analysis.vcf.gz	8 Sep 2023	•	Number of variants in vcf: 9	•
	9	for_kon_update.vcf.gz	8 Sep 2023	•	Number of variants in vcf: 14	
						-



Click on "<u>Go to Sample Definition</u>" to define your samples from the uploaded files.

A Please note that files that have been uploaded and not used for more than 30 days will be automatically deleted from VarSome Clinical.

You can organize your analyses using Analysis tags. It is possible to create new tags or edit existing ones by clicking on the "Tags" option next to the User name.



To create a new tag, enter a name in the "Tag label", choose a color and click on save.

Tags Create and manage tags that you can use to label your analyses.					Dashboard / Tags
Available Tags				Create a tag	
2018	aurigen	cancer sample	control	Tag label	Color
demo tag	epilepsy	epilepsy	index		Save
mistake	multi		NO causative variants		
solved cases	unsolved cases	validated with SANGER			

To edit an existing tag, click on the edit icon displayed on each tag, and to delete a tag click on the trash icon:

Edit icon:	\mathbf{Z}
Trash icon:	

To add a tag to an analysis, click on the tag icon shown in the Analyses Table and select the tag you want to add to the sample:

	Analysis	State	Phenotypes	Туре	Genome	Date	Variants	Tags	ld	User	Assay
=	the set party party and the set of the set	(O New)		G	hg19	01 Aug 2023	14	00	41500000000		
≡>	the second	(O New)			hg19	01 Aug 2023	1	00	-4571000000		
≡>		(O New)	Breast Cancer (OMIM: 114480)	G	hg19	01 Aug 2023	26	10	41499000000		Twist Core Exome + RefSeq Spike in
=		(O New)	Breast Cancer (OMIM: 114480)	G	hg19	01 Aug 2023	0	10	41498000000		Twist Core Exome + RefSeq Spike in
= \	snort_genozip_mes_test	(O New)		6	ba19	01 400 2023	28	20	41497000000		Twist Core Exome + RefSeq Spike in +

Tag icon: 🚫

To filter samples using tags, click on the "Tags" box on the top of the page and select the tag that you want to use to filter your samples. Now, only analyses containing the selected tag are displayed:



Varsomeclinical Analyses Upload / view files Manage -	Launch analysi	s - Filter sets Gene lists	Help								Tags About +
Analyses											Dashboard / Analyses
Search subject id, phenotypes, diseases, variant Sort table by: v bate added Rows: 20 v	From		То		1	auton	nation			ć	1 2 3 4 5 30 >
Analysis	State	Phenotypes	·	Туре	Genome	Date	Variants	Tags	Id	User	Assay
E Description:	(O New)			G	hg19		1	10			
E> Description: test marins XXbh s44:	Ready			G	hg19		14	10			

1.1 Accepted input files

The accepted input files to run analyses on VarSome Clinical are either:

- FASTQ files only from Illumina or MGI sequencers
- VCF files which conform to the <u>VCF standard</u>, regardless of sequencing platform. Users may also optionally upload an alignment BAM file for the VCF sample which can be used to visualize the coverage of the variants provided in the VCF file.

1.1.1 Accepted file names for FASTQ

We expect files that conform to Illumina's or MGI naming convention.

When providing paired-end FASTQ files, we require that reads are properly coordinated between them. Paired-end reads provided in a single FASTQ file are not accepted.

For Illumina pair-end files, we will consider pairs to be files with the exact same name except for the number of the read, for example SampleName_S1_L001_R1_001.fastq.gz and SampleName_S1_L001_R2_001.fastq.gz. We accept files in which the read number is specified alone (for example SN1234_S1_L001_1.fastq.gz and SN1234_S1_L001_2.fastq.gz) or with an "R" before the number (for example SN5678_S1_L001_R2.fastq.gz) and SN5678_S1_L001_R1.fastq.gz). For further instructions in terms of naming conventions, please refer to <u>Illumina</u>.

For MGI pair-end files, we will parse the files as follows: [flow cell ID]_[lane ID]_[barcode ID]_(optional_id)_[read 1/2].fastq.gz and we accept the number of the read to be specified alone (for example, 12345_L02_48_1.fastq.gz and 12345_L02_48_2.fastq.gz) or with an "R" before the number (for example, 6789_L02_56_R1.fastq.gz and 6789_L02_56_R2.fastq.gz)

In cases where there are more than two paired-end files per sample, all the paired reads should be provided: R1 with R2, R3 with R4, R5 with R6 and so on.

1.1.2 Requirements for submitted VCF files



VarSome Clinical accepts VCF files for SNPs/INDEL and CNV annotation. You can upload VCFs containing only SNPs/INDELs or CNVs, but you can also upload VCFs containing both types of variants. If you upload a VCF containing both types of variants, it will be divided into two files: one file to annotate SNPs/small INDELs (**filtered.vcf.gz*) and one file to annotate CNVs (**cnv.vcf.gz*).

1.1.2.1 Required format for SNPs/INDELs annotation

VCFs containing SNPs and small INDELs can be used to launch a somatic or germline analysis: (Launch analysis > New analysis > <u>Germline/Somatic Illumina analysis from VCF</u>).

The VCFs uploaded to analyze SNPs/small INDELs variants must have the following requirements:

- 1. Are compliant with the VCF standard.
- 2. Include *specific* SNVs and INDELs. In order to annotate a variant, we need to know exactly what that variant is, so we cannot handle cases where the variant's sequence isn't specified. For example, we cannot annotate "NON_REF" variants:

#CHROM	POS	ID	REF	ALT
chr1	10052		С	<non_ref></non_ref>

Or variants with an "N" in the ALT field:

#CHROM	POS	ID	REF	ALT
chr22	30998425		С	CTTTTTNT

- 3. Include a valid genotype (GT) field for each variant entry.
- 4. The files should contain the variants found in a real human sample. We expect a maximum of around 4 or 5 million variants in a sample.

1.1.2.2 Required format for CNVs annotation

VCFs containing CNVs (deletions and duplications) can be used to launch a <u>CNV subanalysis from</u> <u>VCF</u>. The VCFs uploaded to annotate CNV variants must have the following requirements:

- 1. Are compliant with the VCF standard.
- 2. Include duplications and/or deletions where the type of copy number variant is shown in the ALT field:



##ALT=<ID=DEL,Description="Deletion">
##ALT=<ID=DUP,Description="Duplication">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT SAMPLE1
chr12 133040735 . C <DUP> . PASS SVTYPE=DUP;SVLEN=140;END=133040875 GT:CN 0/1:1.50 chr12
133049934 . G . PASS SVTYPE=DEL;SVLEN=78;END=133050012 GT:CN 0/1:0.50

3. According to the VCF Specification, the CNV category should not be used when a more specific category can be applied. Therefore, the following VCF format is not accepted:

```
##ALT=<ID=CNV,Description="Copy Number Variant">
chrX 133559227 . G <CNV> . . SVTYPE=CNV;SVLEN=140;END=133559366 GT:CN 0/1:1.50
```

- 4. Include a valid genotype (GT) field for each variant entry.
- 5. Do not include other types of SV variants such as large chromosomal rearrangements (e.g. inversions, translocations) or gene fusions. We currently do not support these types of SV variants.

There are various ways to ensure that the format of your VCF files is correct. For more details please see below:

Tip: checking the format of a VCF file

Ensuring that your VCF file is structured correctly and ready to be uploaded to <u>VarSome Clinical</u> is a recommended practice that could facilitate your analyses and save valuable time.

An easy way to check that your VCF file is valid is to try to run a <u>bcftools</u> command on it. Bcftools, a set of utilities that manipulate VCF files, is very sensitive to malformed VCFs, so it will fail if the file doesn't conform to the standard.

After installing Bcftools according to the instructions, the following command can be executed, where file.vcf represents your input VCF file:

```
bcftools norm -m -any -NO v file.vcf
```

This command will attempt to perform certain actions: check that REF alleles match the reference, split multiallelic sites into multiple rows, or recover multi allelics from multiple rows. If the fields in your file are complete, the command will be executed smoothly. However, if it comes across a non-compliant field like the following,

chrl 16366632	2	СС	GC,GT	193.02	PASS	AB=0.5;
VarSome Clinical User Man	ual Ve	rsion: 11.9.1 -	19th December	r 2023	Pa	ge 45 of 254



the command will fail. In the row above, the field allelic balance (AB) is incomplete, as this is a multiallelic site with two alleles in a single row and two numbers are expected. This information will be provided with an error message:

Error: wrong number of fields in INFO/AB at chr1:16366632, expected 2, found 1

Other alternatives to VCF validation:

- https://github.com/EBIvariation/vcf-validator
- http://vcftools.sourceforge.net/perl_module.html#vcf-validator

which can be used to locate other types of errors (e.g. a malformed or missing header).

Another quick test is to just see if a standard program like <u>bcftools</u> recognizes the file and doesn't complain.

VarSome Clinical API

The platform comes with full API, allowing you to automate each step of the data analysis process, including the data upload. <u>Documentation for VarSome Clinical API</u>

1.1.2.3 Required format for Repeat Expansion annotation

VCFs containing short tandem repeat (STRs) can be used to launch a <u>Repeat Expansion</u> <u>sub-analysis from VCF</u>. The VCFs uploaded to annotate STRs must meet the following requirements:

- 1. Are compliant with the VCF standard.
- 2. The number of repeats is shown in the ALT field as < STRn > where *n* is the number of repeats.
- 3. The INFO field contains the repeat unit in the following format: DisplayRU=CCG
- 4. The FORMAT field contains the number of repeats spanned by each allele.

##FORMAT=<ID=CN, Number=A, Type=Integer, Description="Number_of_repeat_units_spanned_by_the_allele">

1.2 Upload FASTQ files from Illumina BaseSpace.

This feature can be accessed by clicking on the "Illumina BaseSpace" option displayed in the drop down menu shown when you hover over your user name.

Varsomeclinical						
Tags About -						
💄 Profile						
··· Preferences						
ClinVar Submissions						
Illumina BaseSpace						
IIII Claim Assay Token						
Lock Session						
🔒 Logout						

There are two options to connect your Illumina BaseSpace account depending on which BaseSpace Sequence Hub your data is allocated in (EU or US). After that, you will be automatically redirected to Illumina's site in order to login to your account.

	Dashboard / BaseSpace Integration
You are currently viewing the US instance.	
If you have an Illumine Bandpace account and would like to trativity and files directly to ViceSome Clinical you need to connect your BaseSpace account. You will be redirected to Illumine Bandpace where your in the field is apin which your Bandspace account. If you are not already singled in to VarSome SSO you will need to privide your VarSome credentials as well first.	
Connect your BaseSpace account to the EU instance Connect your BaseSpace account to the US instance	

Varsomeclinical
illumına
Sign In Email address
Password
Remember email address
Sign In Don't have an account? Forgot password?
Private domain sign in >
Additional products
Verinata
Correlation Engine

Your projects will then synchronize with VarSome Clinical after you have granted the necessary permits.



	equesting permission to.
Access y	our basic information.
This in	ncludes your name and email address.
68 Global Br	rowse
This a	pp may browse the names of your Projects, Samples, and
App R	esults, but cannot view or download the associated files.

1 Please note: if you do not grant access to VarSome, then your projects cannot synchronize and you will not be unable to transfer files.

This functionality will allow you to download the FASTQ files directly to VarSome Clinical. To do this, please, click on the "Download" button to start downloading the reads into VarSome Clinical.

ame	Project	Updated
A12878-12p-5_L001	HiSeq 4000: TruSeq Exome (12plex replicates of NA12878)	15 June 2018
NA12878-12p-5_S5_L001 Not available 20 August 2017	_R1_001.fastq.gz	2
VA12878-12p-5_S5_L001	_R2_001.fastq.gz	Downic
20 August 2017		

Please, wait a few minutes while the files are being downloaded. The downloading time will depend on the size of your data and on your internet connection.

Data Sets



Data Sets

NA12878-12p-5_L001 HiSeq 4000: TruSeq Exome (12plex replicates of NA12878) The file will be available soon NA12878-12p-5 S5 L001 R1 001.fastq.gz	15 June 2018	HiSeq 4000: TruSeq Exome (12plex replicates of NA12878)
The file will be available soon NA12878-12p-5 S5 L001 R1 001.fastq.qz		
The file will be available soon NA12878-12p-5 S5 L001 R1 001.fastq.gz		
NA12878-12p-5 S5 L001 R1 001.fastg.gz		ole soon
NA12878-12p-5 S5 L001 R1 001.fastq.gz		
		D1_R1_001.fastq.gz
20 August 2017		

Once the download is complete, the status of the file will change from "Not available" to "Available", meaning that the FASTQ files are accessible from the "Upload/View files" page and can be used to launch analyses (germline or somatic) from FASTQ.

98	Project				Updated	
2878-12p-5_L001	HiSeq 4000: Tru	Seq Exome (12plex replic	ates of NA128	78)	15 June 2018	1
12878-12p-5_55_L001_R1_00	1. fastq. gz					Down
organit 2017						
2878-12p-5_55_L001_R2_00	1. fastq gz					Down
the second se						
ugust 2017						
kapuni 2017						
kupun 2017						
Rappor 2017						
loaded files		Uploaded	Status	Notes		Actions
loaded files		Uploaded on	Status	Notes		Actions
loaded files File NA12878-Rep2_S8,	L002_R2_001.fastiq.gc	Uploaded on 4 Jun 2021	Status •	Notes Number of reads: 3250533 Number of bases in reads: 463991083		Actions

Please note that any file that has been previously imported from BaseSpace to VarSome Clinical will have a status of "Available". If the file has already been used in an analysis, in order to analyze it again you will need to find the sample that contains it and select the option "Reuse sample files".



Sample phonotype/s)
Sample phenotype(s)
Sample/Analysis informat
Reuse sample files
View QC report
Downloads

Click on to disconnect the Illumina BaseSpace account from VarSome Clinical.

2. Manage

- <u>Samples</u>
- Workflows

Click on "Manage" to create, view, or edit your samples or workflows.

2.1 Samples

2.1.1 Create and define your samples

Go to Manage > Samples > Create new samples. This step is required before using the samples for analysis.

Once your files are uploaded, you need to create and define your samples. The sample creation involves the association of files to sample names and addition of optional sample metadata related to the patient (e.g. phenotypes).

<u>Uploaded FASTQ files</u> are recognized automatically in pairs (or groups) under the same filename prefix and according to the Illumina or MGI naming conventions. The table displays suggested sample names for each file or pair(s) of files. For VCF files, each file is associated with a suggested sample name.

The files shown in first place with a yellow background are the ones uploaded by the current user. Then, we display the files uploaded by other users (gray background). All files are sorted by their upload date.

Select the rows with suggested sample names in order to create your samples. You can select all by clicking on the "Select" option at the top of the table.



Once the files and suggested names are selected click on "Next".

Select s	samples		2 Next
SELECT	SAMPLE NAME	FILE NAMES	
1	S 200012346_12	S200012346_L01_12_1.fastq.gz (13/09/2023)	

In the following menu you can <u>optionally</u>:

- Modify the **sample type** (germline or somatic). Click on the "germline" or "somatic" boxes if you wish to change the sample type.
- Files: remove or add files.
- **Sample name**: you can edit the suggested name of your sample here. Please remember that the sample ID should not contain information about patient identity.
- **Description**: add a description of your sample. Please remember that, like the sample ID, this field should not contain any information that can be used to identify a patient.
- CNV VCF file for CNV sub-analysis: you can choose along with the VCF/FASTQ files of your sample a VCF file that contains only CNVs for the same sample to be annotated. The CNV results will be displayed as a sub-analysis accessible from the main analysis of the sample.
- Repeat Expansion VCF file for Repeat Expansion sub-analysis: you can optionally add a VCF file containing repeat expansion (RE) variants for annotation. The RE results will be displayed as a sub-analysis accessible from the main analysis.
- BAM file for alignment visualization (VCF samples only): you can optionally attach a BAM file that will allow you to visualize the alignment of the reads and access to the IGV or JBrowse from the variant table. <u>Please note that this file won't be used for variant</u> <u>calling, only for visualization purposes</u>.

You can define your sample as germline or somatic:

Depending on the sample type, there will be the following optional fields:

Germline samples:

Phenotypes: to select phenotypes of interest, start typing a term and the associated phenotypes will appear as a list that can be selected. Phenotypes shown in the dropdown list can be limited based on their source. When selecting "All" you will get terms from HPO, MONDO and OMIM® databases, while selecting "Only OMIM®" they will be retrieved solely from the OMIM® database. These terms can be used to create a gene list and filter the variant analysis for the genes matching the selected phenotypes. For more details see: Phenotype Matching.



Somatic samples:

- **Tissue type:** specify the tissue type of your sample. This field is optional, however we encourage you to fill it in, as this information is taken into account when annotating the variant with information from cancer databases.
- **Cancer type:** specify the sample's type of cancer. This field is optional, however we encourage you to fill it in, as this information is taken into account when annotating the variant with information from cancer databases.
- **Age (years):** specify the sample's individual age (in years). This field is optional, however we encourage you to fill it in, as this information is taken into account when annotating the variant with information from cancer databases.
- **Sex:** specify the individual's sex. This field is optional, however we encourage you to fill it in, as this information gets considered for the annotation with the cancer databases.

Please note that while all of the fields above are optional, we encourage you to fill them all in because this information is taken into account when annotating the variant with information from cancer databases.

Click on "Create" to create the samples. Your samples are now ready to launch a new analysis with them. Go to <u>"Launch new analysis"</u> to find more details.

FASTQ germline example



short2

Germline Somatic		
Files		
short2_L001_R1_001.fastq.gz - 26/11/2023 ×		
short2_L001_R2_001.fastq.gz - 26/11/2023 ×	~	Ť
Sample Name		
short2		
Description		
Phenotype names from All Only OMIM ®		
Search for phenotypes		\sim
CNV VCF file for CNV sub-analysis		
Select files		\sim
Repeat Expansion VCF file for Repeat Expansion sub-analysis		
Select files		\sim

VCF somatic example



variants.hg38

	Germline	Somatic		
		Comato		
Files				
variants.hg38.vcf.gz - 22/11	1/2023 ×		×	\sim
Sample Name				
variants.hg38				
Description				
Tissue Type				
Tissue type(s)				\sim
Cancer Type				
Search for cancer type(s)				\sim
Age (years)		Sex		
		Select		\sim
Soloct files	aiysis			
Select lifes				
Repeat Expansion VCF file for	Repeat Expansion	sub-analysis		
Select files				\sim
BAM file for alignment visualiz	ation (optional)			
Select files				\sim

2.1.2 View/Edit your samples

Go to Manage > Samples > View/Edit new samples

This table shows all created samples. If the current user has already created some samples, they will be shown first (yellow background) while the samples created by other users of the same group are displayed last (gray background). All samples are sorted by the upload date.



Select s	samples			Next
SELECT	SAMPLE NAME	FILE NAMES	CNV FILES	ACTIONS
	sample_ABC	sample_ABC.vcf.gz		 Ø ☑
	C my_sample	S200012346_L01_12_1.fastq.gz S200012346_L01_12_2.fastq.gz		∅ 𝔅 𝔅

In the actions column you can view/edit the sample information and remove the sample. When a user removes a sample, the associated files become available again to create a new sample (Create new sample menu).

Please note that you can only edit and remove your own samples. These actions are not granted for samples created by other users of your group.

2.2 Workflows

Go to Manage > Workflows

2.2.1 Workflow table

The table shows the workflows created within your group, the user who has created them, the date of their last update and the available actions to perform:

- Use workflow (>>): this is to launch an analysis using this workflow. When selected, you will be directed to the launch new analysis page where this workflow is already pre-loaded. You can optionally modify any workflow parameter (if permitted) and then click on "Next" to select the samples to analyze using this workflow.
- Copy a workflow: create a new workflow by copying the values of an existing workflow. Please note that the workflow names must be unique.
- Remove a workflow (): remove an existing workflow.
- Edit a workflow: edit the values of an existing workflow.

Only the group supervisor can add new workflows, edit, duplicate or remove the existing ones. Any user can view the workflow details and use the workflows for launching a new analysis.

2.2.2 Create a new workflow

As the group supervisor, you can create a new workflow.

Click on "Add new" to create a new workflow, then select the input file type (FASTQ or VCF).



Workflow Name	Workflow Type	File Type			
Q Search by name	All	All	~		🕂 Add New
NAME	USER		ТҮРЕ	UPDATED ON	ACTIONS
test default any gene list analysis			Germline - Single Sample	Sep 14, 2023	▷ © @ ₡
Workflow test		S	Germline - Single Sample	Sep 13, 2023	▷₲₫₡
Workflow		is	Germline - Single Sample	Sep 13, 2023	
Multi Sample Workflow (FASTQ)			Multiple samples	Sep 13, 2023	
Tumor Normal Workflow AT_Testing (Temporary)			Tumor Normal	Sep 13, 2023	
Germline Single Sample VCF Workflow AT_Testing			Germline - Single Sample	Sep 12, 2023	
Tumor Normal Workflow (FASTQ)			Tumor Normal	Sep 12, 2023	▷□₫
Couple Workflow (FASTQ)			Couple	Sep 12, 2023	
Family Trio Workflow (FASTQ)			Family Trio	Sep 12, 2023	

You can start filling the workflow fields. The workflow form is divided into two main parts.

Main analysis parameters

The first part correspond to the main analysis parameters:

- Name: workflow name. You can not use a workflow name if it already exists.
- **Analysis type:** select the type of the analysis from the drop-down list. This field can not be changed once the workflow has been created.
- Assay: select the capture, amplicon, whole genome library preparation method or kit corresponding to your analysis. This field is required if you are starting your analyses from FASTQ files. The assay's details will be used to calculate the coverage of the coding regions included in the kit. This information will be shown later in the Quality Control (QC) Report. If you are creating a VCF workflow this field is optional and has no effect other than having the assay mentioned in the QC report.
- Keep variants (only in FASTQ workflows): Do you really want to see all the variants? If you also need the variants that did not pass the quality filters, then you should choose "Variant list will contain all variants". By default, only the high quality variants will be reported. Selecting "Variant list will contain all variants" will increase the amount of annotated variants and it could slow down the analysis.
- **Sequencer** (only in FASTQ workflows): Select the sequencing technology used to obtain the FASTQ files. Two available options: Illumina or MGI.
- Targeted mode (only in FASTQ workflows):
 - **ON:** this mode is used to limit the results to the assay's target regions. In order to achieve this, the variant caller will discard any reads in the alignment BAM file that do not overlap with the assay's regions. If at least one base of a read overlaps with one of the targeted regions, then the read is kept and included in the analysis. If there is no overlap with any of the targeted regions, the read is discarded and not taken into account when calling the variants. The end result of this approach is a faster and more specific analysis. Any variant not falling in the target regions defined by the selected assay will not be called when using the targeted mode.
 - **OFF (untargeted mode):** all variants will be reported regardless of whether they fall within the target regions of the sequencing assay used. You may want to consider using the untargeted mode, especially for whole-exome data, since that



will ensure you see all variants that may be of interest, including those that may be just at the very edge of a target region (e.g. intronic variants).

Genome: choose the reference genome the reads will be aligned to. For VCF-based analyses run against hg19, if the VCF file contains variants reported on "chrM" (the name of the mitochondrial sequence in the hg19 genome), then those will be annotated with respect to the NC 001807.4 sequence, the original mitochondrial sequence of hg19. If the variant is instead reported on "MT" (the name of the mitochondrial sequence in the hg38 genome) then they will be annotated with respect to the rCRS. If the analysis is launched from FASTQ sample(s), using either hg19 or hg38, any mitochondrial sequences will be aligned to the standard mitochondrial genome (GenBank number: NC_120920.1), which is included in the hg38 human genome.

hg19: "chrM" = NC_001807.4, "MT" = NC_012920.1 (rCRS)

hg38: "chrM" = "MT" = NC_012920.1 (rCRS)

Inheritance from: which sources you want to use for the assessment of the mode of inheritance of a gene, which consequently affect the Germline Variant Classification. "All" includes OMIM[®], CGD, GenCC, Gene2Phenotype, PanelApp and Clingen disease validity and "only OMIM[®]" includes only data from OMIM[®]. In case no information about the mode of inheritance is available in the aforementioned databases, the mode of inheritance is selected from <u>DOMINO</u>.

Please note that PM2, BS2 and BP1 rules may change between both modes of annotation because the ACMG implementation uses different thresholds to evaluate these rules depending on the mode of inheritance.

Ethnicity: you can optionally specify one of the ethnicities proposed by Gnomad. When selected, the values shown in the "Frequency" column of the variant table will correspond to the provided ethnicity. When the ethnicity is set to "Not specified" we will display the overall frequency.

Filters and sub-analysis options

Once filled with the above information, you can optionally fill the bottom section. It contains fields related to filters and sub-analyses.

- **Filters**: which (dynamic) filter(s) you wish to apply to the results. When provided, the dynamic filter will be automatically added to the variant table. Please remember that you will need to click on "Apply" to apply the filter in the variant table and get the filtered variants, find more details in section <u>Create a Filter Set</u>.
- Run a gene list sub-analysis: you can provide one or multiple gene lists to filter the results. When provided, a new sub-analysis will be run on top of the main analysis where only variants falling in genes belonging to the list will be shown in the table. If multiple lists are provided, we will create a combined gene list containing all genes included in each of the lists. This option is not currently available for tumor-normal workflows.
- Run a gene list sub-analysis based on sample phenotypes: when selected, a gene list will be built using the genes associated with the sample phenotypes. You must choose to filter by genes associated with <u>all</u> the sample phenotypes or with <u>any</u> of them. This



option is not available for somatic analysis. Please note that if you provide samples without phenotypes, this sub-analysis won't run as there won't be any phenotypes to create the gene list.

- **Show only gene list sub-analysis results**: when selected, the main analysis results will be hidden by default and we will only display the gene list sub-analysis results.
- **Run algorithmic filter sub-analysis**: select one or multiple algorithm filters from the list to run a sub-analysis. Each algorithm filter will be run as a single sub-analysis. Click on "Options" to modify the filter parameters.

Lock workflow values

As the group supervisor, you will be able to create workflows that can be used by you or any other member of your group. While creating the workflow, you can decide if any of the fields has a fixed or an open value that can be changed by the person launching the analysis later. This is done using the lock icon ($\frac{1}{1}$).

When a field is set as open, it will have the default value set by the supervisor when creating the workflow, however this field can later be modified by the user using the workflow for the analysis.

If the lock icon is closed, then this field will have a fixed value that can not be changed when someone uses this workflow. This option is available in all workflow fields except the workflow name and analysis type which can not be changed when the user is using the workflow.

Name	Analysis Type (j)	
my_single_sample_workflow	Germline - Single Sample 🛛 🗙 🗸	
This workflow name can be safely used.		
Assay 👔 🔒	Keep variants (i)	
Twist Core Exome + RefSeq Spike-In	Variants that pass the quality filters $ imes$ $ imes$ $ imes$	
Sequencer 🛆	Targeted Mode 🚯 🛔	
Genome (i) A Inheritance from (i) A	Ethnicity 访 🔒	
⊖hg19 ●hg38 ●All ⊖OMIM	Not specified V	

For example, in the picture above, we are creating a workflow where the values assay, sequencer and targeted mode are locked. The other values, such as genome, will have a default option (e.g. hg38) but can be changed when this workflow is selected for an analysis.



Name	Analysis Type (i)
my_single_sample_workflow	Germline - Single Sample
Assay 🛈	Keep variants (j)
Twist Core Exome + RefSeq Spike-In	Variants that pass quality filters $\qquad \times \lor$
Sequencer	
Illumina	Targeted Mode 🛈 🚫
Genome (i) Inheritance from (i)	Ethnicity (j
	Not specified V

3. Launch analysis

- New analysis
- New cohort analysis

3.1 New analysis

You need to create your samples before launching a new analysis. If you have not defined your samples yet, please go to the <u>Sample Definition</u> section to find more information.

3.1.1 Use an existing workflow

You can start a new analysis from an existing workflow:

- <u>Workflows table</u>: click on the "Use" option (**>**) from the "Actions" column of the workflow of interest to launch an analysis using the selected workflow.



This option will take you to the "Launch analysis" menu where all parameters of the selected workflow are already loaded.



1	2	3	4

Workflow Details

Analysis from: OVCF OFASTQ

Name	Analysis Type 🚯
single_sample_workflow	Germline - Single Sample V
Assay 🛈	Keep variants (i)
Twist Core Exome + RefSeq Spike-In 🗸	Variants that pass quality filters $\qquad \times \qquad \lor$
Sequencer Illumina	Targeted Mode 🛈
Genome (i) Inheritance from (i)	Ethnicity (j)
hg19 hg38 OAll OMIM	Not specified V
Apply filters (i)	
basic-filter-1 X	× ~
Run a gene list sub-analysis ()	
ACMG SF v3.2 ×	× ~

- Extendention by so Free working as Extendention by so Free working as 1 2 3 4 Select Workflow Eg: Workflow - Germline Single Sample OR Use blank
- Launch analysis > New analysis



Here you can select a workflow from the available list of workflows shown in the drop-down list. When selected, all workflow parameters will be loaded.

Please note that when launching an analysis using an existing workflow, if you modify any of the default workflow parameters defined by your group supervisor (unlocked fields) this won't result in the modification of the stored workflow. This change only applies to the current analysis.

3.1.2 Launch an analysis without a workflow

This option can be used to define all analysis parameters without using an existing workflow. Please note that the **analysis parameters defined when selecting this option option won't be saved as a workflow for future analyses**. It will be used only for the analysis launched at that moment. If you wish to create a workflow and use it for future analyses please check the <u>Create a new Workflow</u> section.

To use this option, go to "Launch analysis" > "New analysis" and click on "Launch an analysis without a workflow". You will need to fill in all the details before selecting the samples. All fields (except the workflow name which is not applicable here) are similar to those presented when a workflow is created. Please follow the instructions given in the <u>Create a new Workflow</u> section to define the analysis parameters.

Analysis from: OVCF OFASTO

Analysis type 🛈	
Eg: Germline - Single Sample	
Assay 🛈	Keep variants 🛈
Select or search for an assay 🗸	All variants X V
Sequencer (;)	
Illumina × ~	Targeted Mode
Genome (i) Inheritance from (i)	Ethnicity 🛈
●hg19 ○hg38 ●All ○OMIM	Not specified ~
Filters (1)	
Search filters	~
Run a gene list sub-analysis 🕡	
Search gene list	~
Run a gene list sub-analysis based on sample phenotypes ③ Show only gene	list sub-analysis results 🕢 📃
Run algorithmic filter sub-analysis 🥡	
Search algorithmic filters	~



How to limit the "Launch an analysis without a workflow" option:

By default all users of your group are able to launch analysis and define the analyses parameters themselves. As a group supervisor, if you would like that your colleagues only use the workflows you have created, you can do it by going to "Preferences" and unchecking the box "Permit users to run analyses without using workflows".

lp	Tags Abo 1
	Profile
2	••• Preferences
Enable emails on Analysis finished	ClinVar Submissions
	> Illumina BaseSpace
Comma-separated email addresses Update	Claim Assay Token
	Lock Session
	🔓 Logout
Enable sensitive mode for germline	e CNV analyses
Enable sensitive mode for somatic	CNV analyses
3 Permit users to run analyses without	ut using a workflow

3.1.3 Select samples for analysis

Once selected the analysis parameters, click on Next to select the samples. **Please remember that the samples need to be created before this step**. Go to the <u>Create samples</u> section for more information.



Once specified the analysis settings, you can select which samples you wish to analyze

• Single sample (germline or somatic): analysis of several independent samples. Select the samples from the dropdown. All selected samples will be analyzed independently. Click on "Select all" to select all samples in one click.

Search samples	~
de_novo-1.2_SP_4454	Â
de_novo-1.3_SP_4454	
de_novo-1.1_SP_4454	
cnv_valid_test_10.filtered	
plain_variants_mix.filtered	
for_kon_update_brB7KwkLRL	
for_kon_update.C1367.0976787465	
test_sd	
IC initial 100/567000 1	

• **Couple (for carrier risk analysis)**: analysis of two unaffected samples (male and female). Select one sample per field. Click on "Add another set of samples" to add another couple.

Select Samples

Female CExisting Analysis New Sam	ple			
test_sd		×	\sim	Unaffected
Male CExisting Analysis New Sample				
mother_trio ab		×	\sim	Unaffected
	Add another set of samples			

• **Family trio**: analysis of one affected child (proband) and two unaffected parents. Click on "Add another set of samples" to add another trio.



Launch Analysis

12	3		4
Se	lect Samples		
roband O Existing Analysis 💿 New Sample			
HG005_son	×	\sim	Affected
Nother C Existing Analysis New Sample			
HG006_mother	×	\sim	Unaffected
ather O Existing Analysis O New Sample			
HG007 father	x	\sim	Unaffected

• **Multiple samples**: this is for multi-sample germline analysis that are not couple nor family trio (e.g. extended family like two parents and two siblings). You need to select the affected status for each member of the multi-sample analysis. Click on "add another sample" to add more samples to the analysis. Click on "remove" to remove any of the samples.

Note: If you wish to perform any type of multi-sample analysis with one or more samples already analyzed in VarSome Clinical, you don't need to upload them again. For multi-sample analysis you can click on the "Existing analysis" radio button to browse and select samples already analyzed in VarSome Clinical.



Launch Analysis

12		3	4
Select S	Samples		
Sample name O Existing Analysis O New Sample		Status	
HG005_son	x v	Affected	× ~
Sample name CExisting Analysis ONew Sample		Status	
HG008_sister	× ×	Unaffected	× ~
Sample name O Existing Analysis O New Sample		Status	
HG006_mother	× ~	Unaffected	× ~
Ren	nove		
Sample name CExisting Analysis New Sample		Status	
HG007_father	x ~	Unaffected	× v
Ren	nove		

• **Tumor-normal** (only for analysis starting from FASTQ): one tumor sample and its matched normal. Please note that you have defined your samples previously (the tumor one as somatic and the normal one as germline). Click on "Add another set of samples" to add another tumor-normal pair of samples for the analysis.

Select Samples

Tumour		Normal	
12345_tumor	× ~	12345_normal	× ~

Add another set of samples



3.1.4 Preview and launch analysis

In this step you will see a summary of the analysis settings (workflow details) and selected samples. You can also add tags to the analysis by clicking on the Tags drop-down and select one of the available tags.

Analysis from FASTQ
Workflow Details
Analysis type: Family Trio
Assay: Twist Core Exome + RefSeq Spike-In
Sequencer: Illumina
Targeted Mode: Enabled
Genome: hg19
Inheritance from: All
Ethnicity: Not specified
Run a gene list sub-analysis: ACMG SF v2.0 genes and associated phenotypes recommended for return of secondary findings in clinical sequencing (2016)
Run algorithmic filter sub-analysis: Compound Heterozygous for Trios
Keep variants: Variants that pass the quality filters
Show only gene list sub-analysis results: Disabled
Run a gene list sub-analysis based on sample phenotypes: Disabled
Samples
Sample #1 mother : HG006_mother proband : HG005_son father : HG007_father

You can click on "Previous" to go back to any of the previous steps and modify any of the analysis parameters.

3.2 New cohort analysis (multi-sample VCF)

If you have a multi-sample VCF file with more than 30 samples inside, you should go to this page in order to launch the analysis.

Sample information (left side of the menu):

- Select the multi-sample VCF file to use
- Description: optional description
- Sample Identifier: user sample name
- Phenotypes: optionally provide phenotypes to the cohort analysis

Analysis information (right side of the menu):

- Assay (optional)
- Ethnicity: ethnicity of the samples of the cohort
- Sources to be used for mode of inheritance: all or only OMIM
- Reference genome: reference genome version used to obtain the VCF (hg19 or hg38).



- Analyze the provided sample based on the gene lists: provide a gene list to run a gene list sub-analysis once the main analysis is done.
- Tags: provide tags to label your samples.

Click on "Start" to launch the analysis.

New cohort analysis			Dashboard
Cohort s	ample analysis with referenc	e genome hg38	
Sample Information			
Sample	Assay (optional)	Select	~
Select the multi-sample VCF file to use		Manage your assay preferences. Your assay is not in the list?	
multi_cohort.vcf.gz X V	Ethnicity	Salart	
Description		Jelett	
Cohor analysis	Sources to be used for mode of		
Sample Identifier	intertoinee (or eace adostrication)	Only OMIM ®	
cohort_35_samples	Reference Genome	ha29	
Phenotypes All O Only OMIM ®		1950	
Search for phenotypes V	Analyze the provided sample based on the gene lists	ACMG SF v2.0 genes and associated phenotypes recommended for return of secondary fi X	× ~
	Tags	Select	~
	Start		

4. Launch sub-analysis

- Gene list analysis
- Algorithmic filters
- CNV/SV analysis
- Repeat expansion annotation from VCF file

4.1 Gene list analysis

<u>Note:</u> The methodology to generate a gene list from phenotype(s) has been changed. Previously, when making a gene list from phenotypes, we would include only those genes that are directly annotated with that phenotype. We have now extended this and instead first collect all diseases linked to the phenotype and then all genes linked to those diseases, as well as any genes directly linked to the phenotype. We already worked this way when adding phenotypes to analyses, so this change ensures we are consistent and also makes sure we don't miss any genes when creating gene lists.

Filter the results of a finished analysis to show only those variants falling within (including 500 base pairs up and downstream) the genes given.



Launch a new gene list analysis	Dashboard / New gene list analysis
Launch a sub-analysis filtered by either an existing gene list OR by an ad-hoc gene list generated from phenotypes using y Analysis Cerne list Please select	ur selected data sources Phenotype names from: All Orbito OMMA Start filing in a phenotype (ype 3 characters minimum). Search multiple by separating with semicolon [∨] . Selected phenotypes (changing this list will not affect the phenotypes linked to this sample). Gene list analysis from phenotypes Gene list analysis based on all selected phenotypes Gene list analysis based on all selected phenotypes.
	Start analysis

- Analysis: choose the analysis to be filtered.
- Gene List: choose a gene list to filter by.
- **Phenotypes:** choose a list of phenotypes. A gene list will be built using the genes associated with the chosen phenotypes. When filtering by phenotype, you can also choose to filter by genes associated with *all* the phenotypes or with *any* of them.

4.2 Algorithmic filters

You can apply sophisticated filters to perform more complex variant filtering to finished analyses. For several of the existing filters we now provide the option to change specific parameters that used to be fixed. That way you can customize each of those filters according to your needs.

You can access Algorithmic filters under Launch analysis > Algorithmic filter analysis as shown in the picture below.



To start an algorithmic filter, choose the target analysis, select an algorithmic filter, and then click on **Start** analysis.



alysis	Available filters Display non applicable filter
Please select	Chromitish for an all of the second sec
	Camerinsk for couples (n= 2) Camerina (n= 2)
	Compound Heterozygous Candidates (n= 1) 0 12 10000
	De novo (strict) (n= 3) O
	De novo candidates (naive) (n= 3) 0
	Exonic and splicing (na 1) 0 (# Options)
	Family trio recessive (coding and rare) (n= 3) 0
	Fisher exact test (na 4) O Options
	Genes in common (nit 2) 0 27 Options
	GWAS Catalog (n≥ 1)
	Max other samples (nit 1) 0 [2] Options
	PharmGKB (nž 1) O CZ Options
	□ Segregating Variants (n≈ 2) 0 2 Cotions
	Trio Recessive (n= 3) 0 22 Options
	□ Variants in Common (ni≥ 2) 0 2 Cotors
	VarSome Picks (n= 1)

For further details on the algorithmic filter and the provided parameters please click on the info icon next to the "Options" button.

As you can see for several of the existing filters, we provide the option to change specific parameters. That way you can customize each of those filters according to your needs.

If you want to change a parameter click on the light blue "Options" button:



For example, if you want to change the default parameters for the filter "Max other samples $(n\geq 1)$ ", after you choose from the drop-down the analysis to which you wish to apply one or more filters, you then need to select the filter "Max other samples $(n\geq 1)$ " by clicking the relevant check-box, which will in turn activate the Options box.



Launch a new algorithmic filter analysis



Maxim number of	0	
samples	0	
oumproo		

Now you can change the default value for this filter, by changing the number in the *Maxim number of samples* field and then by clicking on "Save".

Once the analysis has run, you can view the filter options selected for the algorithmic filter analysis both in the sub-analysis name

		Analysis	State	Phenotypes
=	•	exonic and splicing	Set State	
	=	Exonic and splicing(Flanking Region=10)		

and in the "Sample analysis information" option of the "Analysis actions" menu:





Analysis 4			
Attribute	Value	Updated on	
Capture Method	Generic capture kit	1 Jun 2022	
Description		1 Jun 2022	
Reference Genome	hg19	1 Jun 2022	
Sample Type	germline	1 Jun 2022	
Sequencer	Unknown	1 Jun 2022	
Sample File	exonicandsplicing_filters_test.vcf.gz	1 Jun 2022	
Other			
Attribute	Value	Updated on	
Algorithmic filter	Exonic and splicing version 11.2 released on 16-Feb-2022	1 Jun 2022	
Hard Filter Parameters	{"flanking_size": 10}	1 Jun 2022	
Sample Input File Type	vcf	-	

Available Algorithmic Filters

4.2.1 ACMG Actionable Genes

Variants meeting the criteria described in <u>ACMG SF v2.0: Kalia et al., Genetics in Medicine,</u> (2017) 19, 249–255 (doi:10.1038/gim.2016.190).

4.2.2 Carrier risk for couple

Pathogenic variants for which both members of a healthy couple are heterozygous and so can be passed on to any offspring. Included are genomic (not mitochondrial) variants that are classed as pathogenic, likely pathogenic or of uncertain significance, are not homozygous in either prospective parent and either:

- Both individuals are heterozygous for the variant
- The variant is on the X chromosome of the mother
- The variant falls in a gene that has at least two variants where one is present and heterozygous only in one individual and the other is present and heterozygous only in the other individual. In such cases, all pathogenic, heterozygous variants for that gene are shown as they are candidates for compound heterozygosity.

Options:

• Strong VUS: If selected, keep variants of Unknown Significance (but only if one of the strong pathogenic Germline rules has fired for this variant). VUS variants qualify only if they trigger one of the Germline rules: PVS*, PS* or PP5. If not selected, keep all VUS variants (irrespective of Germline rules).

 \bigotimes


Strong rules are PVS1, PS1, PS2, PS3, PS4 and PP5. In addition, any rule whose strength has been raised to "strong", "very strong" or "stand alone" will be considered a "strong" rule, even if its original strength as per the Germline guidelines was lower. For example, PM1 can be raised from "supporting" to "strong" if the variant is located in a dense hot-spot.

A Please note that we recommend that you further filter by genes with recessive mode of inheritance or loss-of-function pathogenicity.

4.2.3 Compound Heterozygous Candidates

Variants classified as pathogenic, likely pathogenic or of unknown significance for which all of the following apply:

- they are heterozygous variants in genes that carry at least one other heterozygous variant and no homozygous pathogenic variants.
- are not in mitochondria

Options:

• **Strong VUS**: If selected, keep variants of Unknown Significance (but only if one of the strong pathogenic Germline rules has fired for this variant). VUS variants qualify only if they trigger one of the Germline rules: PVS*, PS* or PP5. If not selected, keep all VUS variants (irrespective of Germline rules).

Strong rules are PVS1, PS1, PS2, PS3, PS4 and PP5. In addition, any rule whose strength has been raised to "strong", "very strong" or "stand alone" will be considered a "strong" rule, even if its original strength as per the Germline guidelines was lower. For example, PM1 can be raised from "supporting" to "strong" if the variant is located in a dense hot-spot.

- Homozygous also: If selected it will also filter for variants which are homozygous.
- **Phased Mode**: If selected, it will only filter those variants with phasing information to identify compound heterozygous variant pairs in the same gene and phasing group and in different zygosities (1|0 vs 0|1).

4.2.4 Compound Heterozygous for Trios (n=3)

This filter is aimed to support the identification of compound heterozygous variants in an affected child when the genome of the two unaffected parents is also provided (Family Trio analysis (n=3)).

The filter will keep variants in the child classified as pathogenic, likely pathogenic or of unknown significance for which all of the following apply:

- they are heterozygous variants in genes that carry at least one other heterozygous variant and no homozygous pathogenic variants.
- are not in mitochondria
- If such variant pairs are detected, it will look for compound heterozygous pairs in the same gene in each of the parents.
- If either parent does have a compound heterozygous pair in the same gene (not necessarily the same pair as the child), then we discard the pair identified in step a, and



move to the next candidate. The assumption is that parents are unaffected, so the child would also be unaffected if one of the parents have a compound heterozygous pair in the same gene.

Options:

- Use phasing information (**phased mode**). When this option is enabled the filter will identify compound heterozygous variant pairs in the child—that is heterozygous variants in the same gene and phasing group and in different zygosities (1|0 vs 0|1).
- **Strong VUS**: If selected, keep variants of Unknown Significance (but only if one of the strong pathogenic ACMG rules has fired for this variant). VUS variants qualify only if they trigger one of the ACMG rules: PVS*, PS* or PP5. If not selected, keep all VUS variants (irrespective of ACMG rules). Strong rules are PVS1, PS1, PS2, PS3, PS4 and PP5. In addition, any rule whose strength has been raised to "strong", "very strong" or "stand alone" will be considered a "strong" rule, even if its original strength as per the ACMG guidelines was lower. For example, PM1 can be raised from "supporting" to "strong" if the variant is located in a dense hot-spot.
- Homozygous also: If selected it will also filter for variants which are homozygous.

4.2.5 De novo (strict)

Variants present in the proband and absent in both parents, where neither parent has any reads supporting the variant but only counting positions where the parents have a minimum coverage of 8.

4.2.6 De novo candidates (naive)

Variants likely to have arisen in the child from unaffected parents. Variants must meet either of the following conditions:

- the child is homozygous for the variant, but the variant is only present in one parent OR
- the variant is present in the child, but not present in either parent.

1 Please note that we recommend that you further filter for pathogenicity, coverage, frequency, and mode of inheritance.

4.2.7 Exonic and splicing

A special case of algorithmic filter is "Exonic and splicing" which only keeps exonic (including UTR and other non-coding exons) and splicing (no more than 10 nucleotides from a known splice site) variants. This filter comes with the option to change the distance from a known splice site according to your needs.

The filter may be launched manually, just like any other, but it will also be run automatically on any analyses with more than 500 000 variants.

This algorithmic filter provides the same results as would occur if you made a dynamic filter with the following:

Coding



Splicing Non-coding exon +3' utr Non-coding exon +5' utr

Function	
□ 3'flank	
☑ 3'utr	
□ 5'flank	
✓ 5'utr	
✓ coding	
intronic	
splicing (show all splicing variants; if this is the only function selected, any coding impact selections will be ignored)	•
✓ non-coding exon	
□ intronic but not splicing	

The aim of this filter, and the reason it will run automatically for large analyses, is to provide a smaller subset of results to the user which will be far quicker and easier to sort through. Since, even with WGS analyses, the variants of interest tend to be those that can affect the protein sequence, we feel that this filter will help our users quickly identify and focus on the variants of interest even on larger samples such as WGS.

Please note that while the filter will be run automatically for such large analyses, the full result set will still be available as usual. The filter will run as a sub-analysis and will not affect the results of the main, parent analysis in any way.

3' and 5' flank

The 3' and 5' flank refer to the transcript(s). Using Dynamic filter for Function, VarSome Clinical will mark a variant as flanking if it is within 500nt of a transcript's start or end position. UTRs are part of the transcript, and such shouldn't be an issue.

If you have a particular set of regions, we could filter the result for those regions. Alternatively, you could run a gene list analysis, limiting the results to only the list of genes.

Another possibility, if this is something you will often be doing, is setting up an Algorithmic filter. Algorithmic filters are more sophisticated filters that run as separate analyses and there we can write one catering to whatever specific filtering you require. In that case, we could search for coding variants in all genes and also include any found in the upstream regions of the target genes.

splicing

Splicing information is used by our Germine Variant Classification implementation to decide whether some rules should be applied or to boost their strength.



We use the scSNV as well as MaxEntScan databases for splice-site prediction. This is only available for single-nucleotide variants. Variants above the 'ADA Boost Splicing' threshold (0.708) and 'Random Forest Splicing' threshold (0.515) are considered candidate splicing variants. The scSNV splice-site prediction is used in the following rules:

BP7: A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice site consensus sequence nor the creation of a new splice site AND the nucleotide is not highly conserved. (Benign, Supporting). The splicing is checked as follows:

The variant is found more than 2 bases away from the next splice site.

It is not predicted splicing according to the scSNV database.

PP3: Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.) (Pathogenic, Supporting). We exceptionally boost the strength to 'Strong' if the variant is predicted splicing and rule PVS1 was not triggered.

Splicing is also considered in rule:

PVS1: null variant (nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single or multi exon deletion) in a gene where LOF is a known mechanism of disease. (Pathogenic, Very Strong).

In PVS1, intronic variants within 1 or 2 nucleotides from the exon of the transcript splice site are considered null variants as defined in this paper.

4.2.8 Family trio recessive (coding and rare)

The following criteria are applied in this order.

- 1. Keep only, either:
 - a. Coding variants that are frameshift or missense or nonsense or stop-loss or exon deletion or in frame or start loss or splice junction loss
 - b. Splicing variants (+/-10 bp from exon ends)
- 2. Remove variants where allelic balance < 0.2 or coverage <= 5
- 3. Remove variants present in both parents except if child is homozygous and both parents are heterozygous.
- 4. Remove variants with gnomAD population frequency over 1% based on the ethnicity of the proband. If ethnicity is not provided, the general population frequency is used.
- 5. Remove variants that are homozygous for the alternative allele in either parent.
- 6. Remove variants in mitochondria and chromosome Y.
- 7. Keep only variants where EITHER of these criteria apply:
 - a. Variant is homozygous for the alternative allele in the child.
 - b. Child is heterozygous and the following criteria BOTH apply:
 - i. There are two or more variants in the same gene. To qualify, a variant must be in a coding transcript of a gene with a Transcript Support Level consistent with the sample's settings. The variant must be Pathogenic, Likely Pathogenic or of Uncertain Significance (VUS).
 - ii. The variants did not all come from the same parent; some variants on the gene may have come from the mother and some from the father or are de



novo.

Options: All of the following fixed values originally set for the filter can now be changed:

- Maximum distance from splice site
- Minimum distance from splice site
- Minimum coverage
- Minimum allelic balance
- Maximum frequency

This filter specification was kindly contributed by Dr Erica Davis of Lurie Children's Hospital in 2021.

4.2.9 Fisher exact test

This filter will select any variants that are found more often (fisher exact test p-value <= 0.05) in the affected samples than in the controls.

4.2.10 Genes in common

All variants classified as pathogenic, likely pathogenic or of unknown significance (but only if one of the strong pathogenic Germline rules has fired for this variant), which are found in genes with at least one such variant in all samples of a merged analysis.

The filter works in the following way:

- It will identify all genes with at least one pathogenic, likely pathogenic or VUS variant in all samples.
- It will then return ALL pathogenic, likely pathogenic or VUS variants falling in that set of genes.

Options:

• Pathogenic only: If this option is selected, variants classified as Benign(B) or Likely Benign(LB) or VUS/LB-B will be excluded from the results.

4.2.11 GWAS Catalog

All variants with GWAS Catalog data.

Options:

• Pathogenic only: If this option is selected, variants classified as Benign(B) or Likely Benign(LB) or VUS/LB-B will be excluded from the results.

4.2.12 Max other samples

Remove variants that exist in more than the selected number of other samples from your group, and those with more than 5% population frequency.



4.2.13 PharmGKB

Variants associated with genes that have PharmaGKB information. For more information see in the section Pharmacogenomics Knowledge Base (PharmGKB).

Options:

• Pathogenic only: If this option is selected, variants classified as Benign(B) or Likely Benign(LB) or VUS/LB-B will be excluded from the results.

4.2.14 Segregating variants

For some of these filters, in order to add the parameterization functionality we have integrated similarly functioning filters into one. For example, you can now find under the filter "Segregating Variants", the filters previously known as:

- Segregating Variants (dominant, all VUS),
- Segregating Variants (dominant, strong VUS)
- Segregating Variants (recessive, all VUS)
- Segregating Variants (recessive, strong VUS)
- Compound Heterozygous Segregating Variants (all VUS)
- Compound Heterozygous Segregating Variants (strong VUS).

The segregating variants analysis falls in the category of algorithmic filters that can be performed on Family trio analyses as shown below:

amily trio analysis	New Analysis	
and the within the dealysis	New cohort analysis (multi-sample VCF)	
	Merge existing analyses 2 ~	- please refer to our agreement for the use of th
Family Trio Identifier	Family trio analysis 3	Proband
Family Trio Identifier	Carrier analysis for a couple	Please select
We do not accept names or other personally identifying information		Mother
Description	Merge analyses	Please select
Description	Gene list analysis	
Sources to be used for mode of inheritance (affects classification)	Algorithmic filter analysis	Father
All Only OMIM ®		Please select
In addition to the full variant list, apply the following filter (optional)	CNV analysis from FASTQ	Tumor samples cannot be selec

It is a parameterizable algorithmic filter with the following options:



Segregating V	ariants			×
Strong VUS				
Segregating recessive				
Compound heterozygous segregating				
				Save

Segregating dominant variants

Variants annotated as pathogenic, likely pathogenic, or of unknown function (but only if one of the strong pathogenic Germline Variant Classification rules has fired for this variant) that are present (in any zygosity) in all affected samples and absent from all healthy samples. Select the "Strong VUS" option.

When selected, "Strong VUS" will keep variants of unknown significance (VUS) only if one of the strong pathogenic ACMG rules has fired for this variant. If not selected, any VUS that meets the filter's other criteria will be kept (irrespective of ACMG rules).

Strong rules are PVS1, PS1, PS2, PS3, PS4 and PP5. In addition, any rule whose strength has been raised to "strong", "very strong" or "stand alone" will be considered a "strong" rule, even if its original strength as per the ACMG guidelines was lower. For example, PM1 can be raised from "supporting" to "strong" if the variant is located in a dense hot-spot.

Segregating recessive variants

Variants annotated as pathogenic, likely pathogenic, or of unknown function (but only if one of the strong pathogenic Germline Variant Classification rules has fired for this variant) that are homozygous in all affected samples but heterozygous or absent from all healthy samples. Select the "Segregating recessive" option.

Compound heterozygous segregating

Variants present (in any zygosity) in all affected samples and absent from all healthy samples and it will also find possible segregating compound heterozygous variants: those variants annotated as pathogenic, likely pathogenic, or of unknown function that are found in genes where all affected samples have at least two so annotated heterozygous variants. Select the "Compound heterozygous segregating" option.

This option cannot be used in combination with the Segregating recessive option.

Running the analysis with multiple affected samples

Image: Please note that this filter is not specifically designed for family trios, but for generalmulti-sample (e.g. cohort) analyses. When running it with multiple affected samples, it will lookVarSome Clinical User ManualVersion: 11.9.1 - 19th December 2023Page 79 of 254



for variants that are homozygous in all affected samples and absent or heterozygous in all healthy samples. The segregating variants (dominant) filter will look for variants that are present (het or hom) in all affected samples and absent in all healthy samples.

4.2.15 Trio Recessive

Pathogenic variants that may be causative of recessive disorders in the child of unaffected parents.

We follow this selection process:

- 1. We exclude variants that are homozygous for the alternative allele in either parent.
- 2. We exclude variants in mitochondria and chromosome Y.
- 3. We keep variants where EITHER of these criteria apply:
 - a. Variant is homozygous for the alternative allele in the child.
 - b. Child is heterozygous and the following two criteria BOTH apply:
 - i. The variants did not all come from the same parent; some variants on the gene may have come from the mother and some from the father or are de novo.
 - ii. There are two or more variants in the same gene. To qualify, a variant must be in a coding transcript of a gene with a Transcript Support Level consistent with the sample's settings. The variant must be Pathogenic, Likely Pathogenic or of Uncertain Significance (VUS).

Options:

- Strong VUS: If selected, keep variants of Unknown Significance (but only if one of the strong pathogenic Germline rules has fired for this variant). VUS variants qualify only if they trigger one of the Germline rules: PVS*, PS* or PP5. If not selected, keep all VUS variants (irrespective of Germline rules).
- **Missing from one parent**: Only keep variants that are missing from at least one parent (strict); in other words, neither parent has all variants that the child has in the same gene.
- **Maximum frequency for recessive**: We exclude any variants that are found with a frequency greater than the selected threshold. The default value is set to 1, so that no variants will be excluded based on the frequency.

1 Please note that we recommend that you further filter by genes with recessive mode of inheritance.

4.2.16 Variants in Common

Variants annotated as pathogenic, likely pathogenic or of unknown function (but only if one of the strong pathogenic ACMG rules has fired for this variant) that are present (in any zygosity) in all samples.

Options:

• Common only to all affected: If selected, keep all variants that are found (in any zygosity) in all affected samples; their status in non-affected samples is ignored.



• Pathogenic only: If this option is selected, variants classified as Benign or Likely Benign will be excluded from the results.

4.2.17 VarSome Picks

VarSome Picks is an advanced algorithmic filter, designed to empower bioinformatic analysis of genomic variants using AI. This tool goes beyond conventional variant prioritization methods by considering essential parameters such as phenotype, gene, and variant data, all within a disease-specific context. Its primary objective is to rank potentially causative variants, aiding researchers and clinicians in identifying significant genetic associations related to specific diseases.

Main Features

Comprehensive Data Sources

VarSome Picks leverages a diverse set of authoritative data sources to enhance the accuracy of its variant ranking. These sources include:

- HPO (Human Phenotype Ontology) is used to get the phenotype (HPO term) gene association
- GenCC (Genetic Clinical Characterization)
- CGD (Clinical Genomic Database)
- ClinGen Disease Validity
- MONDO (Monarch Disease Ontology) is used here to get the disease (MONDO term) gene associations.
- <u>OMIM</u>[®] (Online Mendelian Inheritance in Man)
- Gene2phenotype
- PanelApp

By integrating information from multiple databases, VarSome Picks ensures a comprehensive and well-informed list of variants that have been classified as potentially causative for the disease, even those previously labeled as Variants of Uncertain Significance (VUS).

Genetic and Phenotypic Parameters

VarSome Picks incorporates crucial genetic and phenotypic parameters to prioritize variants accurately. These include:

- <u>Phenotype(s) Selected by the User:</u> VarSome Picks ranks variants found in the Top 10 genes associated with the selected phenotype(s) by the user. This feature facilitates tailored analysis, focusing on variants with higher disease relevance.
- <u>Zygosity</u>: Variant-related information on zygosity is considered to assess the impact of heterozygous or homozygous variants on the disease phenotype.
- <u>VarSome Germline Classification</u>: The algorithm applies the ACMG (American College of Medical Genetics and Genomics) criteria for germline variant classification, ensuring robust variant categorization.



• <u>Mode of Inheritance:</u> Gene-related information on the mode of inheritance is taken into account to identify variants associated with specific inheritance patterns.

Quality Parameters

VarSome Picks considers essential quality parameters for variant evaluation:

- <u>Allelic Balance</u>: For germline variants, the algorithm applies a threshold for allelic balance to ensure reliable variant calling.
- <u>Coverage in the Sample:</u> Coverage information is evaluated to gauge the depth of sequencing and assess data quality.

How to use VarSome Picks

VarSome Picks supports germline analyses from FASTQ files and VCF files against hg38 and hg19 reference genomes. It is also available for gene list analyses.

The filter can be run on-demand (Analysis actions > New algorithm filter analysis) once the main analysis has finished. It can also run automatically for all analyses if the group supervisor enables this option in the analysis preferences.

	Permit users to run analyses without using a workflow
Run VarSome Picks automatically	~
	Update

In this case, the algorithmic filter will automatically run only when phenotypes are provided by the user. The results will be shown as the result of any other algorithmic filter, and in this case, with the variants being ordered by the VarSome Pick's assigned priority.

VarSome Picks runs only for:

- Single sample germline analysis
- Gene lists analysis



I		Description:			
	≡~	Description:		(Ready	1 (I
	=	VarSome Picks Based on:			
	≡	Description:	Ľ	Resolved	E

Whenever phenotypes are added or modified by the user in an existing analysis, if the supervisor has enabled the automatic analyses option, VarSome Picks will re-run automatically. The previous output is erased, ensuring up-to-date and accurate results for the latest phenotype selections.

It is also possible to run VarSome Picks even if phenotypes are not initially provided. In this scenario, the algorithmic filter will not be triggered automatically. However, a warning will be displayed to make sure our users are aware of the potential limitations of the analysis due to the absence of phenotype information:



Note to the User/Disclaimer

Experts and users should exercise discretion in evaluating the appropriateness of VarSome Picks for their specific use case, considering factors like non-specific or genetically heterogeneous phenotypes and idiosyncrasies in search functionality (e.g. "hearing loss" instead of a query using the term "deafness" may not return the relevant gene). Variable expressivity and age-related penetrance, while important considerations, are not currently accounted for by the algorithm. Therefore, the decision regarding the appropriateness of this approach has to be at the discretion of the experts/the users.

4.3 CNV/SV analysis

VarSome Clinical currently offers two CNV calling solutions for your samples that are suitable for all types of NGS samples: WGS, WES and panels.



4.3.1 CNV/SV calling (from FASTQ)

VarSome Clinical currently offers two types of CNV calling solutions:

- **Delly** suitable for single WGS samples and WGS tumor-normal CNV analysis.
- **ExomeDepth** suitable for cohorts of WES/panels and also for WGS samples.

To start a CNV analysis from FASTQ, please go to "Launch analysis" > "CNV analysis from FASTQ".

Upload / view files Manage 🗸	Launch analysis - Filter sets Gene lists
1	New Analysis
	New cohort analysis (multi-sample VCF)
	Merge existing analyses >
2	Gene list analysis
	Algorithmic filter analysis
	CNV analysis from FASTQ

Whole exome sequencing (WES) or targeted panel data

For such samples, we use the ExomeDepth CNV caller. The read depth based tool requires five or more (ideally between five and ten) germline or somatic samples that have already been analyzed on VarSome Clinical. These will be run as a cohort with each sample analyzed using the rest as a control. The samples should all have been sequenced using the same assay since CNV calls will only be made in the assay's target regions. For optimal results, the selected samples should:

- be from the same sequencing run
- come from individuals unrelated to each other and
- be of the same sex (either all male or all female). If the samples of the cohort are of not of the same sex, the CNV calls obtained for chromosomes X and Y will not be reliable

All samples will be analyzed together and the results (along with a <u>visual display</u>) of each sample will be shown as a sub-analysis of that sample.

▲ Please note that an inherent limitation of WES is that it produces reads only covering the ~2% of the human genome that falls in exons. Therefore, the full spectrum of CNVs and breakpoints may not be completely characterized. In addition, many large CNVs and VarSome Clinical User Manual Version: 11.9.1 - 19th December 2023 Page 84 of 254



cross-chromosome events may not be detected. For optimal results, we suggest either sequencing the entire genome (WGS), or a different experimental approach such as array CGH. Nevertheless, CNV detection based on WES data may give a quick insight into CNV patterns for a specific disease or phenotype. For more details on the limitations of calling CNVs in such data, please see R. Tanner *et al.*, 2014.

Sensitive mode

CNV calling for **non-WGS** CNV analyses is also available in "Sensitive mode". Compared to standard mode, a lower CNV detection threshold is applied, resulting in more sensitive calling and typically in a higher number of calls. CNV detection can be particularly challenging; for instance single exon CNVs can be hard to call. Still, in a clinical setting, the ability to detect such CNVs is of paramount importance. Sensitive mode is optimized for the needs of clinical laboratories. It allows a shift to the trade-off between recall and false discovery to benefit sensitivity, compared to the standard mode.

You can enable this feature for either somatic or germline samples (or both) in Preferences. Please, note that these settings are only available to the group administrator and any changes will be applied to all users of the group.



Whole genome sequencing (WGS) data

For WGS samples, VarSome Clinical offers two solutions:

- CNV calling for a single WGS sample. We use delly, an integrated structural variant (SV) caller tool that can detect both CNVs and other forms of Structural Variants (SVs) at single-nucleotide resolution in short-read genomic sequencing data. It combines 3 different approaches (paired-ends, split-reads and read-depth) to discover extensive genomic rearrangements. Quality passed CNV calls (deletions and duplications) are retained, while other types of SVs are currently not reported.
- CNV calling for a cohort of (2-5) WGS samples. The ExomeDepth caller has been adapted to also process WGS. The solution is suitable for samples with long CNVs (>50kb) that may not be reliably called by delly. For WGS, the assay target regions comprise the complete genome, split into 50Kb bins. As a result, this imposes a hard minimum size limit: no CNVs smaller than 50Kb can be detected using this approach. The requirements for non-relatedness between the samples and their processing by the same laboratory,



sequencer and ideally in the same batch, apply to WGS samples too. All samples are analyzed in a single CNV analysis and the results (along with a <u>visual display</u>) of each sample are shown as a sub-analysis.

A step-by-step example on how to run a CNV/SV analysis

Select "CNV analysis from FASTQ" from the "Launch analysis" drop-down menu on VarSome Clinical:



VarSome Clinical interface allows you to select a minimum of five (5) and a maximum of twenty five (25) already analyzed samples to be used as a cohort for CNV calling. For best results, we recommend you select **5-10 samples** from unrelated individuals of the same sex that were sequenced on the same sequencing run.

CNV analysis Start a new CNV analysis	DashboardCNV analysis
Existing tag Existing tag • • Salet at tag to retrieve avalges Analysis #1 Pless select • •	 You may start a CNV analysis from: 5 or more (deally, around 10 net-WGS germline or tumor samples, from the same sequencing out, as a cohot makyis. CNV calls on the X chromosome should only be trusted if the input samples are all the same set. If a mix of male and female samples are out of the same sequencing run, se a cohot mahysis. As above, CNV calls on the X chromosome value out oblication of the input samples are all the same set. If a mix of male and female samples are used, CNV calls on the X chromosome value out oblication of the input samples are all the same set, if a mix of male and female samples are used, CNV calls on the X chromosome will be unreliable. A faingful whole some samples are used, CNV calls on the X aromosome will be unreliable. A faingful whole some samples, from the same individual and using the same stage, as a Tumor/Normal analysis.



Each sample's results will appear as a sub-analysis of the main analysis.



4.3.2 CNV annotation (from VCF)

VarSome Clinical provides a pipeline to annotate CNVs from VCF files. There are two ways to annotate CNVs from VCFs.

- Provide a valid VCF file that contains **both** copy-number-variants (CNVs) and SNPs / small INDELs when launching a new analysis either from FASTQ or VCF. Files that include both small variants and CNVs will be split to two separate files, one for small variants and one for CNVs (i.e. **.filtered** and **.cnv**).
- Provide a valid VCF file containing only CNV variants.

The CNV analysis from VCF is launched as a sub-analysis of the main analysis. You can launch a CNV annotation by:

- Adding a CNV VCF file when defining your sample.



Create Samples

	sho	rt2	
	Germline	Somatic	
Files			
multi_vcf_test.vcf.gz -	09/11/2023 ×		× ~
Sample Name			
short2			
Description			
Phenotype names from All O Only OMIN	1 ®		
Search for phenotypes			~
CNV VCF file for CNV su	b-analysis		
Select files			~

- Launching the analysis once the main analysis has finished as a "New CNV sub-analysis" either from single or multi sample analyses.

	=		`
Based on: no variants by added on ld:	-	TruSight One May2014 5126 variants Description: by added on id:	SNVs & Indels Sample/Analysis information Reuse sample files View QC report Downloads
Based on: 94599 variants by added on id:	=	TruSight One May2014 5126 variants Description: Dy added on id:	2 Re-Annotate Analysis New Gene-List Analysis Add BAM file(s) New Algorithmic Filter Analysis New CNV Subanalysis Archive sample data
	≡.		
rt a new CNV subanalysis		3	Dasht
: CNV file:		Pue a	The second s

From the Dashboard menu:

From the Analyses menu:

SNVs & Indels	-
Sample phenotype(s)	
Sample analysis information	
Reuse sample files	
View QC report	
Downloads	
Archive sample data	
Re-annotate analysis	
New Gene-List analysis	
New algorithmic filter analysis	
New CNV Sub-Analysis	
New Repeat Expansion Sub-Analysis	
Add BAM file(s)	Multi_san

Start a new CNV subanalysis	3	Dashboard
Select CNV file:		Pun a CNV subanalysis
input.cnv.vcf.gz	× ~	 "A VCF file containing CNV calls (and only CNV calls) must be provided, originating from the initially analyzed sample."
	4 ►Start analysis	

4.4 Repeat expansion annotation from VCF file

VarSome Clinical can annotate short tandem repeats (STR) VCF files. Please go to the "<u>Requirements for Repeat Expansion VCF files</u>" to find out more information.

The STR analysis from VCF is launched as a sub-analysis of the main analysis. You can launch a STR annotation by:

- Adding an STR VCF file when defining your sample.



Create Samples

sample_ONT_chr15	
Germline Somatic	
Files	
sample_ONT_chr15.vcf.gz - 13/11/2023 x	× ~
Sample Name	
sample_ONT_chr15	
Description	
Phenotype names from All Only OMIM ®	
Search for phenotypes	~
CNV VCF file for CNV sub-analysis	
Select files	~
Repeat Expansion VCF file for Repeat Expansion sub-analysis	
Select files	~
RAM file for alignment visualization (ontional)	
Select files	~
Back	

- Launching the analysis once the main analysis has finished as a "New Repeat Expansion sub-analysis" either from single or multi sample analyses.





5. Merging analyses

- Family trio analysis
- Carrier analysis for a couple
- Merge analyses

To merge already finished analyses, click on "Launch analysis" and choose one of the options under "Merge existing analyses". When the analyses to be merged were run from FASTQ input files, VarSome Clinical will perform <u>joint calling</u> on all samples to be merged. The allelic balance and the coverage for each sample will be displayed in the resulting variant table. Clicking on the coverage numbers opens a window showing the read coverage (JBrowse) for the selected variant in that sample.

If any of the parent analyses were instead run from VCF input data, then joint calling cannot be performed and so VarSome Clinical will merge the results of the original analyses directly.

The finished analysis will be displayed as shown in the image below: For analysis containing more than one sample:



	Analysis	State	Phenotypes	type	Genome	Date	variants	Tags	Id	User	Assay
≡~	Automated_merge_family_trio(n=3) Description: Automated test for merge family trio	(O New)			hg19			00			
=	Gene List (n=3)				hg19			00			
=	De novo candidates (naive) (n=3)				hg19			00			
=	De novo (strict) (n=3)				hg19			00			
=	De novo (NP+Fisher+MinCov) (n=3)				hg19			00			
=	De novo (noSupportingReads) (n=3)				hg19			00			

The first entry corresponds to the main sample (Automated_merge_family_trio (n=3)), the merged analysis, and below it are shown any sub-analyses applied to the parent merged analysis (for example *De novo candidates (naive), De novo (strict)*, etc).

5.1 Family trio analysis

For a trio analysis, the mandatory fields are the analysis identifier, and the identifiers of the Proband, Mother and Father analyses. You can also select any of the listed algorithmic filters which will be applied to the analysis' results. Please click on the *i* icon to get the description of each filter. The algorithmic filters (eg. De novo candidates) are optional, and can be started at any moment. Click on "Start analysis" to launch the analysis.

Family trio analysis Start a new family trio analysis		Dashboard / New family trio analysis
This function may incur a charge - please I	efer to our agreement for the use of this platform.	
Family Trio Identifier	Proband	
Family Trio Identifier	Please select	 Affected
We do not accept names or other personally identifying information		
Description	Mother	
Description	Please select	Unaffected
Sources to be used for mode of inheritance (affects classification)	Father	
All Only OMIM	Please select	 Unaffected
In addition to the full variant list, apply the following filter (optional) De novo (strict)	Tumor samples cannot be selected for family trio analysis	
De novo candidates (naive)		
Exonic and splicing C (2' Options		
Family trio recessive (coding and rare) Izl Options		
Genes in common C C/Options		
GWAS Catalog 0 27 Options		
Max other samples 0 C Options		
PharmGKBO (C) Cpscors		
Segregating Variants (27 Options		
Trio Recessive 0 22 Crators		
□ Variants in Common0 😰 Crivers		

5.2 Carrier analysis for a couple

Choose an identifier for the analysis, and an optional description. Then, on the right hand side, select the female and male samples that will be analyzed. You can also select any of the listed algorithmic filters which will be applied to the analysis' results. Please, click on the **i** icon to get the description of each filter. The algorithmic filters (eg. De novo candidates) are optional, and can be started at any moment. Click on "Start analysis" to launch the analysis.



	This function may incur a charge - please refer to our agreement for the use of this platform.	
Couple Identifier	Female	
Couple Identifier	Please select	 Unaffected
le do not accept names or other personally identifying information		
Description	Male	
Description	Please select	▼ Unaffected
courses to be used for mode of inheritance (affects classification)	Tumor samples cannot be selected for carrier analysis	
All		
○ Only OMIM ®		
In addition to the full variant list, apply the following filter (optional)		
Carrier risk for couples		
Exonic and splicing 2 27 Options		
Genes in common 3 C Options		
GWAS Catalog		
Max other samples 2 (2) Options		
PharmGKB0 (2) Options		
Segregating Variants 0 2' Options		
Variants in Common Continue		
		Stat analysi
		Start analysis

5.3 Merge analyses

Select this option "Merge analyses" to merge analyses with no special settings. Simply set an identifier and select the analyses to be merged on the left hand side. Click on "Start analysis" to launch the analysis.

Multi sample analysis Start a new multi sample analysis		Dashboard / New multi sample analysis
This function may incur a charge - please re	fer to our agreement for the use of this platform.	
Multi-Sample Analysis Identifier	Analysis #1	
	Please select	N/A 👻
We do not accept names or other personally identifying information		
Description	Tumor samples cannot yet be selected for multi-sample analyses	
Sources to be used for mode of inheritance (affects classification)		
All		
O only OMIM ®		
 Martin 		
Existing tag		
· · · · · · · · · · · · · · · · · · ·		
Select a tag to retrieve analyses		
		Start analysis

6. VarSome Clinical Dashboard and Analyses pages

The **Dashboard** page displays your *latest* analyzed samples, while the **Analyses** window displays all the samples that you or your group have uploaded to VarSome Clinical.

In both pages you can see the analysis status (e.g. running). An email will be sent to you once the analysis finishes.



Varsomeclinical Analyses Upload/view files Mana	age 🗸 Launch analysis 🖌 Filter sets Gene lists Help			Tags About
Dashboard				Dashboard
Tools	Latest analyses - view all analyses			
Texel: 4959 / 9989 Meanhay: 21 / 991 Samples / Analyses	O sample Lest 162020 varints Description: by added on 13 Nev 2023 dr. 4555000000	=	CIVIX Results of sample Ltest Based or: 14727 variants by added on 13 Nev 2023 id: 43559000001	=
0/0 Filter sets / Filters	© 2061242722-CHA	Ξ	VarSome Picks Based on: faste hg38 no variants	=
310 Gene lists	Description: by added on 13 Nov 2023 id: 43558000000		added on 13 Nov 2023 id: 43554000001	=
Upload/view files	Fisher exact test Based on: fisher kp (w-300)	=	172 variants	
New Analysis	by added on 13 Nov 2023 id: -4719000001		by added on 13 Nov 2023 id: 43557000000	
New colhort analysis (multi-sample VCF) Family tito analysis Carrier analysis for a acougle Carrier analysis for a acougle Carrier site analysis Algorithmic filter analysis CNV analysis from FASTQ	Repeat expansion results of main_analysis_for_re Based or chain_analysis_for_re 2 variants by added on 13 Nev 2023 ist 43093000011	=	Fisher exact test Based on: multi-Fisher exact test (n=11) by added on 13 Nev 2023 id: -464000003	-

A Please note that when an analysis has no results to report, **no variants** will be displayed in the sample's grey box and no active link will be visible to access the results.

If however an analysis doesn't report any information on the variants called (or not), it doesn't necessarily mean that it has no results to show.

Accel-Amplicon EGFR Pathway Panel	Ξ
Description: by added on 7 Jan 2020 id: 10706000000	

When clicking on "Analyses", you will be redirected to the Analysis Table View, a table that contains all the analyses that you or members of your group have uploaded to VarSome Clinical. From the Analyses Table, you can see the status of each analysis you are running or have run, as well as access the main analysis information and select any actions on an analysis.



	h subject id, phenotypes, diseases, variant	From	то			Select ta	ıg(s)				۹ 🗠
rt tab	le by: Date added Rows:										2 3 4 5 221
											2 J T J 221 J
	Analysis	State	Phenotypes	Туре	Genome	Date	Variants	Tags	Id	User	Assay
≡>	sampletest Description:	(O New)		G	hg38	13 Nov 2023	162 020	00	43559000000	-	
=	2061242722-CHA Description:	(O New)		s	hg19	13 Nov 2023	173	00	43558000000	100	SOPHIA DDM Myeloid Solution
=	Description:	(O New)		s	hg19	13 Nov 2023	172	00	43557000000		SOPHIA DDM Myeloid Solution
=	P121498_1 Description:	(O New)		G	hg19	13 Nov 2023	0	00	43556000000	-	Twist Core Exome + RefSeq Spike-In
≡>	fastq hg38 Description:	() New	Acne, Adult (OMIM: 604324)	G	hg38	13 Nov 2023	12	00	43554000000	-	Twist Core Exome + RefSeq Spike-In
≡>	For_kon_update.cnv.C1391.5287413698 Description: VCF Analysis with CNV Sub-analysis	(O New)		G	hg19	13 Nov 2023	14	00	43553000000	-	Twist Core Exome + RefSeq Spike in
=	VP test-Sample defined with Phenotypes Description:	(O New)	Hypospadias HP:0000047	G	hg19	13 Nov 2023	29	00	43550000000		
≡>	VP test-Add phenotypes in completed analysis Description:	(O New)	Hypoglossia-Hypodactylia (OMIM: 103300)	G	hg19	13 Nov 2023	29	00	43552000000	-	
=	Description:	() New		G	hg38	13 Nov 2023	0	10	43248000000		
≡>	fisher kp (n=300) Description:	() New			hg38	13 Nov 2023	422 447	00	-4718000000	1	
=	short.C1425.7410915916 Description:	(O New)		G	hg19	13 Nov 2023	25	00	4318000000	-	Twist Core Exome + RefSeq Spike-In
≡>	hard_filters_mt Description:	(O New)	Hypogeusia HP:0000224	G	hg19	13 Nov 2023	132	00	43247000000		
_		(O New)		G	hg19	13 Nov 2023	5 126	00	43207000000		
	Description:										

For each analysis, you can view some basic information like the name and the description of the analysis, the state, the phenotypes linked to this analysis (if any), the type of the sample (germline [G] or somatic [S]), the reference genome used for this analysis, the date it ran, the number of variants that were called, the number of <u>tags</u> added to this analysis, the analysis ID, the name of the user that launched the analysis as well as the Assay used for this analysis.

VarSome Clinical allows the user to search for any of the analyses in which a variant has been found. This can be done from the "Analyses" page, if you type the variant you are looking for and press enter/click on search (\bigcirc):

Analyses										Dashboard / Ana	alyses
rs121913254 To 🖨 Select tagis)										a a	
Sort table by: • Date added Rows: 20 •									¢	1 2 3 4 5 17	>
Analysis	State	Phenotypes	Туре	Genome	Date	Variants	Tags	Id	User	Assay	
⇒ Description:	(O New)		G	hg19	07 Nov 2023	5.126	00				
E Description:	(O New)		s	hg19	07 Nov 2023	5.126	1 🛇				

Only those samples containing the variant will be shown in the "Analyses" menu.

In addition, it is also possible to identify samples containing the given variant within a subset of analyzed samples. To do this, you can combine the previous feature with the Sample Tags feature.

First, you will need to label the samples of interest using the Sample Tags. Then, you can go to the "Analyses" menu and search for a specific variant in a subset of samples labeled with a particular Tag.



In the example shown in the picture below, only those samples labeled with an "az" tag and containing the specific variant will be shown.

Analyses													Dashboard / Analyses
BRCA1(NM_007294.4):c.671-49del Sort table by: • Disc added Rows: 20 •	From		i	To 2			az				3	•	۹. ۵
Analysis	State	Phenotypes			Туре	Genome	Date	Variants	Tags	Id	User	Assay	
E Description:	B Set State				G	hg19	01 Mar 2023	50	26 🟷				
-													

The accepted formats for variant search queries are the following:

- Chromosome position ref seq variant seq (e.g. chr7:140453136:A:T)
- HGVS DNA-level variants (e.g. BRAF:c.1799T>A, BRAF(NM_004333.6):c.1799T>A)
- HGVS single amino acid substitutions (e.g. BRAF Val600Glu, BRAF p.Val600Glu, BRAF V600E)
- dbSNP rs id (e.g. rs113488022)

You can view the current status of each of your analyses from the colored bar on the left and by hovering the mouse over the bar:

alys	es										Dashboard / Ana
	Analysis	State	Phenotypes	Туре	Genome	Date	Variants	Tags	Id	User	Assay
≡>	samplecnv_test Description:	(O New)		G	hg38	13 Nov 2023	162 020	00	43559000000		
=	2061242722-CHA Description:	(O New)		s	hg19	13 Nov 2023	173	00	43558000000	Sec.	SOPHIA DDM Myeloid Solution
=	Description:	() New		s	hg19	13 Nov 2023	172	00	43557000000	The second	SOPHIA DDM Myeloid Solution
=	P121498_1 Description:	(O New)		G	hg19	13 Nov 2023	0	00	43556000000	The second	Twist Core Exome + RefSeq Spike
≡>	Fastq hg38 Description:	(O New)	Acne. Adult (OMIM: 604324)	G	hg38	13 Nov 2023	12	00	43554000000	-	Twist Core Exome + RefSeq Spike
≡>	For_kon_update.cnv.C1391.5287413698 Description: VCF Analysis with CNV Sub-analysis	() New		G	hg19	13 Nov 2023	14	00	43553000000		Twist Core Exome + RefSeq Spike
=	VP test-Sample defined with Phenotypes Description:	() New	Hypospadias HP:0000047	G	hg19	13 Nov 2023	29	00	43550000000		
≡>	VP test-Add phenotypes in completed analysis Description:	(O New)	Hypoglossia-Hypodactylia (OMIM: 103300)	G	hg19	13 Nov 2023	29	00	43552000000		
=	_phased_analysis Description:	(O New)		G	hg38	13 Nov 2023	0	10	43248000000	Concerning of	
≡>	Fisher kp (n=300) Description:	() New			hg38	13 Nov 2023	422 447	00	-4718000000		
=	shori Description:	(O New)		G	hg19	13 Nov 2023	25	00	43180000000		Twist Core Exome + RefSeq Spike
≡>	hard_filters_mt Description:	(O New)	Hypogeusia HP:0000224	G	hg19	13 Nov 2023	132	00	43247000000	-	
=	Description:	(O New)		G	hg19	13 Nov 2023	5 126	00	43207000000		
=	short.C1425.0106717201 Description:	(O New)		G	hg19	13 Nov 2023	0	00	43181000000		Twist Core Exome + RefSeq Spike
=	custom Description: custom.C5474.1254332084	(O New)		s	hg19	13 Nov 2023	0	00	43240000000	-	Twist Human Core Exome
=	custom	(O New)		s	hg19	13 Nov 2023	0	00	43239000000		Swift Biosciences Accel-Amplicon

The different colors indicate different sample/analysis status that follow this pattern:

Red: the analysis has failed

Solo.Ashk.son.L008.fastq.genozip test2	(O New)	G	G	hg19	28 Jul 2023	0	00	41445000000	-	Twist Human Core Exome
Analysis failed son L008 faste repozin									Marilena	

Green: the analysis has finished successfully



=	germline single vcf bam Description:	(O New)	Acne, Adult (OMIM: 604324)	G	hg19	28 Jul 2023	25	15	41446000000	
A	nalysis finished successfully genozip test2	New		6	ha19	28 Jul 2023	0	0FN	41445000000	Twiet Human Core Evome

Blue: The analysis is currently running

Analysis	State	Phenotypes	Туре	Genome	Date	Variants	Tags	ld	User	Assay
germline single vcf bam	(O New)	Acne, Adult (OMIM: 604324)	G	hg19	28 Jul 2023	25	1 🔿	41446000000		

Gray: The analysis is archived

Analysis	State	Phenotypes	Туре	Genome	Date	Variants	Tags	Id	User	Assay
Analysis archived	Bet State		G	hg19	28 Feb 2017	307 442	00	32000000		TruSight One May2014

You can open the analysis actions menu by clicking on the three horizontal lines next to the name of the analysis:

	lyses Upload / view files Manage 👻 I	aunch analysi	is 👻 Filter sets 🛛 Gene lis	ts Help								Tags About
nalyses												Dashboard / Analyse
Search subject id, phenotypes, disea	ses, variant	From	=	То		i	Select ta	g(s)				۹ ۵
ort table by: • Date added Rows:	20 🗸										< 1	2 3 4 5 217 >
Analysis		State	Phenotypes		Туре	Genome	Date	Variants	Tags	Id	User	Assay
main_analysis_for_re Description		O New			G	hg38	07 Nov 2023	20	00	43093000000	-	
Sample phenotype(s)		O New			G	hg19	07 Nov 2023	32	00	43092000000		Twist Core Exome + RefSeq Spike-In
Sample analysis information Reuse sample files		O New			G	hg19	07 Nov 2023	32	00	43091000000		Twist Core Exome + RefSeq Spike-In
View QC report Downloads		O New			G	hg19	07 Nov 2023	32	00	43090000000	-	Twist Core Exome + RefSeq Spike-In
Archive sample data Re-annotate analysis	Multi_sample_analysis_not_genotyped.42	O New				hg19	07 Nov 2023	0	00	-4707000000	-	
New Gene-List analysis New algorithmic filter analysis	sie	() New			G	hg38	07 Nov 2023	2 033	00	43080000000		
New CNV Sub-Analysis		(O New)			G	hg19	07 Nov 2023	5 126	00	43055000000	100	
Add BAM file(s) Description:	a <mark>2 no bam, s1 no bam</mark> (n=2)	() New				hg19	06 Nov 2023	54	10	-4706000000	-	
Description:		O New			G	hg19	06 Nov 2023	134	00	43067000000		Twist Human Core Exome

From the Analyses Table, you can also change the name or the description of your analyzed samples:



var	someclinical	Analyses	Upload / vie	ew files Manag	ge≁ I	Launch an	alysis -	Filter sets	Gene lis	ts Hel	p		
alys	es		/										
Search	h subject id, phenotypes	, diseases, vari	ant			Fro	m			То			
Sort tab	le by: <mark>→ Date added</mark> Ro Analysis	ows: 20	~			State	F	Phenotypes				Туре	Genome
≡>	Description:	g				() New						G	hg19
≡>	Multi sample analysis Description:	fron				O New	(<u>Syncope</u> HP:000 OMIM: 609289)	1279 <u>Syncop</u> 	<u>e, Familial</u>	<u>Vasovagal</u>		hg19
≡>	Multi sample analysis Description:	from				O New	(<u>Syncope</u> HP:000 OMIM: 609289)	1279 <u>Syncop</u> 	<u>e, Familial</u>	<u>Vasovagal</u>		hg19
Analysi	is		State	Phenotypes		Туре	Genome	Date Vari	ants Tags	ld	User	Assay	

Additionally, when adding phenotypes to the analysis these will be displayed on the Phenotypes column.

	Analysis	State	Phenotypes
≡>	Description:	O New	
≡>	Description:	(O New)	<u>Syncope</u> HP:0001279 <u>Syncope, Familial Vasovagal</u> (OMIM: 609289)
≡>	Description:	(O New)	<u>Syncope</u> HP:0001279 <u>Syncope, Familial Vasovagal</u> (OMIM: 609289)
≡>	oresemption.	O New	

When you launch any sub-analysis on your main analysis, you can click on the arrow on the left of the analysis to view the sub-analysis:

varsomeclinical	Analyses Upload / view file	es Manage - Launch ana	lysis 🗸 Filter sets 🛛 Gene lisi	s Help		
nalyses						
Search subject id phenotype	s diseases variant	From	#	То		
Sort table by: • Date added R	iows: 20 🗸					
		State	Phenotypes		Type	Genome
Description:		State	Phenotypes		Туре G	Genome
	s from	State New New	Phenotypes Syncope HP:0001279 Syncope (OMIM: 609289)	<u>, Familial Vasovagal</u>	G	Genome hg19 hg19



•	var	Somecinical Analyses Upload / view files Manage -	Launch analysis	- Filtersets Genelists I	Help			
Ar	nalys	es						
	Search Sort tabl	n subject id, phenotypes, diseases, variant le by: - Date added Rows: 20	From		То			iii
		Analysis	State	Phenotypes		Туре	Genome	
	≡≻	Description:	(O New)			G	hg19	
	≡>	Multi sample analysis from . Description:	() New	Syncope HP:0001279 Syncope, Fam (OMIM: 609289)	nilial Vasovagal		hg19	
	≡∽	Multi sample analysis from Description:	() New	Syncope HP:0001279 Syncope, Fam (OMIM: 609289)	nilial Vasovagal		hg19	
	=	Analysis Filtered by Gene list from		<u>Syncope</u> HP:0001279 <u>Syncope, Fan</u> (OMIM: 609289)	milial Vasovagal		hg19	
	=	CNV Results of Based on:		<u>Syncope</u> HP:0001279 <u>Syncope, Far</u> (OMIM: 609289)	milial Vasovagal	G	hg19	
	≡≻	Description:	(O New)			G	hg38	

The Analyses Table view also displays possible issues detected in the sample by <u>FastQC or other</u> <u>quality control</u> measures applied:

	es								
= >	Description:	Unew		G	ng 19	27 NOV 2023	y	00	
≡ >	Multi sample analysis from Description:	O New	Syncope HP:0001279 Syncope, Familial Vasovagal (OMIM: 609289)		hg19	27 Nov 2023	0	1 🛇	
≡ >	Multi sample analysis from Description:	O New	Syncope HP:0001279 Syncope, Familial Vasovagal (OMIM: 609289)		hg19	27 Nov 2023	5.138	1 🕟	
≡ >	Description:	(O New)		G	hg38	27 Nov 2023	4	1 🕟	
-	Multi sample analysis from Description:	O New			hg19	27 Nov 2023	0	0 🔿	
≡ >	Multi sample analysis from Description:	- O New			hg19	27 Nov 2023	23	00	
≡ >	Multi sample analysis from Description:	O New	<u>Ovarian Melanoma</u> (MONDO: 543) <u>Esophageal</u> <u>Melanoma</u> (MONDO: 1192)		hg19	27 Nov 2023	5.138	10	
≡ >	Description:	() New		s	hg38	26 Nov 2023	97	10	
-	Description:	(O New)		G	hg19	26 Nov 2023	32	00	*
_	Description	() New			hg38	26 Nov 2023	5	00	

Clicking on the error or warning icon will let you investigate the issue.



When an analysis is finished, it will appear on your **Dashboard** and, depending on the state of your clinical sample, you can select a predetermined status for it. The **Clinical Sample Status** is a way to organize analyzed samples within a team or between different teams that collaborate to yield a clinical result. In this way, when the analysis is finished and is ready for a curator team to examine the results and yield a diagnosis, the **Clinical Sample Status** can give information on the status of the analysis at a given moment by the previous team that worked on it. If a clinical diagnosis was completed for an analysis, the **Clinical Sample Status** can be changed to indicate

whether a diagnosis was made (Resolved) or if no diagnosis could be reached (Ourresolved).

For the analyses that were performed before 11.7.5 release a black status icon will appear next to the analysis name and by clicking on it, you can select a status for your older analyses:

WB wgs_hg38 WGS pcr-free WGS -PCR 5207330 variants	=
Description: by added on 19 May 2023	ľ
id:	•

After an analysis is finished, you can choose between four different statuses:

@ 23-05 -tes	01 ≡
ONEW	ľ
◎ READY	18°
	_
OUNRESOLVED	•

: The analysis has finished running and its results can be viewed.

• READY : The analysis is ready for review. This option is useful when multiple teams share the work of analyzing a sample, and the first team sets up basic filtering and then passes it on to



the next team. Setting the state to "Ready" indicates that the first pass has been done and the sample is now ready to be viewed by the next team.

RESOLVED

: A diagnosis has been made.

OUNRESOLVED: The analysis was inconclusive: no diagnosis can be made at this time.

The Analyses tab displays all the analyzed samples. You can use the "Search" field to look for sample names, users, phenotypes or diseases. Gene list analyses or algorithmic filters are displayed as sub-analyses of the main sample.





6.1 "Analysis actions" options

You can choose among a series of actions that allow you to retrieve information or generate different reports, providing an overview of your results, or further explore the analysis. You can also access this menu from the "Analysis Actions" button of the upper right corner at the variant table page.

SNVs & Indels

The results of the analysis. When clicked at the Dashboard leads to the Variant table.

Sample phenotype(s)

Using these fields it is possible to add and modify the clinical description of the sample as described below. You can choose this option if you want to provide phenotypes associated with your case from a drop-down menu or view the already selected ones. The phenotypes will be matched to the resulting variants.

Add / Modify phenotypes

You can also add phenotypes in an already analyzed sample by clicking on the button of the Analysis actions menu either when viewing your sample in the Analyses page:





Or you can edit the Phenotype information from the results page of an analysis by clicking on the "Analysis actions" button:

	1 Tags A	bout -
	🕇 Filters 🛛 🛓 Report ge	eneration Analysis actions 👻
٥ 2	Q 📋 🞜 🗧	SNVs & Indels Sample phenotype(s) Update samples affected status
	exon 10 of 21 positi	Sample/Analysis information Reuse sample files
	exon 21 of 21 positi	View QC report
	exon 2 of 5 position	Downloads
	exon 15 of 18 positi	Re-Annotate Analysis
	exon 2 of 11 positio	New Gene-List Analysis
	exon 19 of 21 positi	Add BAM file(s) New Algorithmic Filter Analysis
	exon 15 of 18 positi	New CNV Subanalysis
	exon 11 of 27 positi	New Repeat Expansion Sub-Analysis

You will then be directed to a new screen where you can add the phenotype information and then click on "Save" to apply the changes. After a few minutes, the "Phenotypes" column will be updated

Phenotype(s) information Provide phenotype(s) information for 101212809-Madre_74.119TESTING	Dashboard / Phenotype(s) information
Warning: Adding or removing phenotypes will not modify the ACMG classification, it will only affect the "Phenotypes" column with the number of phenotype associations.	
Start by filling in a phenotype. Click on phenotype from the drop down menu to add it to your phenotypes selection. All ● Only OMM ● Start Filling in a phenotype (type 3 characters minimum). Search multiple by separating with semicolon [∨] . Selected phenotypes	
	Save

There will be a column named "Phenotypes" in the variant table with a value per each variant. This column will contain the number of user input phenotype(s) associated with the variant



gene.

Germline Class 🛛 🍦	Germline Rules	O Phenotypes
Likely pathogenic	PP3 Strong PM2 Supporting PP5	2
Likely benign	894	0
Likely benign	BP4 Moderate PM2 Supporting	2
Likely benign	894	2
Likely benign	BP4 Moderate	0
Likely benign	BP4 Moderate	0
Likely benign	BS2 BP6 Moderate PP2 PP3 PP5	1
Likely benign	BP4 Strong	0
Likely benign	BP4 Strong	0
Likely benign	BS1 BP4	0

Hover over the number with the mouse to see the name of the matched phenotypes.

This column can be used to sort the table in descending/ascending order. You can also create a <u>dynamic filter</u> to filter variants based on these values following the steps below:



Tags About -						
	Filters 🛃 Rej	port generation	Analysis acti	ons 🕶		
for gene, chr (e.g. hr1 c	۹ 🖬 🕯	; Filters	Total Var	iants: 104		
Over 1 Genes	O Phenotypes	Select a filt	er set	~		
CASQ2	0		Add	Create		
TNNT2,ENSG00000286	0	CIRes	et Filters	Q Apply Filters		
TGFB2	0	C Mus				
RYR2	0					
PRDM16	0			2		
PCSK9	0					
ACTN2	0					
Number of Phenotypes Exclude variants that match Number of Phenotypes From	3 To	Add				
✓ OMIM Variants Filter	I	Add				
✓ Custom Variant Classifications		Add				
✓ ACMG Points Score		Add				

Update sample affected samples

It is possible to change the affected status of the merged analysis from the three horizontal lines next to the analysis name on the Analyses page.

This option is useful in order to run algorithmic filters for segregating variants, where you need to set which sample is the affected and unaffected respectively.

4 Save filter set ×



Multi sample analysis	Coloredge and Advantage of State
SNVs & Indels	
Sample phenotype(s)	
Update samples affected status	
Sample analysis information	e Fr
Reuse sample files	u
View QC report	
Downloads	
Re-annotate analysis	
New Gene-List analysis	
New algorithmic filter analysis	
New CNV Sub-Analysis	
New Repeat Expansion Sub-Analysis	
Add BAM file(s)	
Delete BAM file(s)	

You will be directed to a screen, as the one shown below:

Multi sample analysis from the same and th	Dashboard / Update sample status
Affected status N/A	v
N/A	~
N/A	~
	Update

You can delete the filters displayed, then update the affected/unaffected status of your samples and run the algorithmic filters you would like for your merged analysis.

Sample/Analysis information

It opens a pop-up table with an overview of information about the analysis and the sample, the sequencing (e.g. capture method, number of reads and bases, sequencer), the analysis (e.g. the versions of the databases and the implemented pipelines and software), databases, classification (e.g. ACMG annotator version) and other useful metrics such as those mentioned below.

Useful metrics:

<u>Predicted Sex:</u> As part of the main analysis, the sample sex is predicted based on zygosity in selected chromosome X SNVs. The algorithm assumes that females are XX and males are XY. Sex is predicted with a p-value (binomial test) <0.01, otherwise sex is undetermined. INDELs are not considered in the calculation due to their higher



false positive rate.

- <u>Rare Homozygous Count</u>: This metric provides the fraction of rare variants in homozygosity. A minimum of 50 genotypes are required to report results.
 - An SNV is considered rare when its frequency is below 1% or if it is unknown.
 - InDels are not considered in the calculation due to their higher false positive rate. The number of rare SNVs and their fraction in homozygosity is reported.
- <u>Runs of Homozygosity</u>: Samples are scanned for the presence of extended segments of homozygosity. Runs of homozygosity (**ROH**) are computed directly from the VCF file. These calculations are based on the information contained in the VCF file (either the user's or the one generated by VarSome Clinical) and specifically on the either genotype likelihoods (PL) or genotypes (GT). By default, genotype likelihoods are expected. ROH is calculated for different minimum region sizes (100Kb, 500Kb, 1.5Mb and 5Mb). For each size, we calculate the number of regions of that size or greater which show evidence of ROH.

Please note:

- ROH are computed for germline samples, but not for somatic
- Only bi-allelic sites are considered in ROH calculations.
- ROH are only reported for autosomal chromosomes.
- All ROH have a minimum of 50 variants
- INDELs are not considered in the calculation due to their higher false positive rate.
- Minimum ROH quality (Phred score) is 20. ROH quality represents the probability of the state assignment being incorrect. Larger values indicate a more confident call.
- ROH is calculated for all analyses. Please note, however, that ROH results for small, targeted panels are less reliable because of the sparseness of the target regions.

The following ROH related metrics are reported for each length threshold in Analysis Actions > Sample/Analysis Information:

NROH: The number of ROH sections detected in the sample

SROH: The total length of ROH in bp

maxROH: The length of longest ROH segment in bp

FROH: The fraction of autosomal genome in autozygosity

ROH are visible also in the <u>Sample View</u>.



Reuse sample files

This option is available if you wish to reuse an existing sample for a new analysis. A blue box message in the upper right corner of the page will let you know that you will receive an email when the input file(s) of the sample is/are again available for new analysis and which you can view in the "Upload / view files" page.

Gene Coverage

When clicking on Gene Coverage, a new window will open showing a tree with the coverage per gene. Please note that only information for 100 or fewer genes is shown since the depiction of more genes would be impractical. If you wish to see coverage information for analyses with more than 100 genes, please use a gene list to limit the displayed results to 100 genes or fewer.

The gene list can be selected when the new window opens by clicking on **T**Filters . If this option has been selected when analyzing a gene list, the tree will contain only the genes from that gene list.


Clicking on any gene will result in a tree of its transcripts and each transcript is also expanded to a tree showing the coverage of individual exons. Clicking on an exon will open a new tab, showing the read coverage from the corresponding bam file.

If you click on the filter icon on the top left corner this will take you to a new page, showing a long list of genes. You can filter for your gene of interest:





The following options concern the generation of different reports and can be useful when the user needs to check the quality of the sequencing in terms of coverage or retrieve information about the alignment.

View Quality Control report

You can view a quality control report about your analysis and download it as a PDF file or export it as a docx file. The report includes a list of information such as sequence technology, read alignment results, regions reported, coverage, number of identified variants by class, summary for ACMG rules and number of SNV found in coding regions.

In order to find the Quality Control (QC) Report of an analysis in VarSome Clinical you can go

Version: 11.9.1 - 19th December 2023



- to the Analyses page and select from the three horizontal lines next to the analysis name you wish the "View QC Report" or



- in the Variant Table page you can click on the "Analysis actions" button and select the same option

either:



About	•
peneration	Analysis actions +
÷ ₹-	SNVs & Indels Sample phenotype(s)
Functic	Sample/Analysis information
	Reuse sample files
	Gene coverage
_	View QC report
6	View FastQC report
0 💿 😳 (Coding coverage report
0	Region list coverage report
•	Downloads
•	Coverage report for targeted regions
6	Re-Annotate Analysis
6	New Gene-List Analysis
6	New Algorithmic Filter Analysis
•	New CNV Subanalysis
5f in	New Repeat Expansion Sub-Analysis
	Archive sample data
< 1	Delete FASTQ sample data
	Delete BAM sample data

A QC report includes all of the categories presented below:

General information

Simple information about the user, the reference genome, the sequencing technology, the file name and the type of pipeline used.

Software and Database versions

Information about Saphetor's and Sentieon's software versions used for alignment and variant calling and all available sources' versions that were used in the pipeline for annotation.

Read alignment results

This matrix reports information about the alignment of all reads and on the targeted regions of the selected assay.



Read alignment results for short.

	All reads	On target
Total number of reads:	18 284	13 105 (71.67%)
Number of duplicates:	3 645 (19.94%)	2 789 (15.25%)
Number of reads singly mapped:	13 807 (75.51%)	10 103 (55.26%)
Number of reads unmapped:	18 (0.10%)	0 (0.00%)
Number of reads multiply mapped:	25 (0.14%)	2 (0.01%)
Number of reads with gapped alignments:	431 (2.36%)	212 (1.16%)
Number of reads below minimum quality:	383 (2.09%)	1 (0.01%)
Number of reads with insert size:	14 166 (77.48%)	10 295 (56.31%)
Average insert size:	138 bp	139 bp
Number with supplementary mapping:	0 (0.00%)	0 (0.00%)
Number out of mapping range:	0 (0.00%)	0 (0.00%)
Number of bases in reads:	2 211 519	1 611 185

Regions reported and Average region coverage

Display of information regarding the assay used for the analysis and the average coverage of the targeted regions.

Regions reported:	TruSight Cancer (1736 regions, 320536 bp)
-------------------	---

Average region coverage: 76

<u>Coverage</u>

This matrix shows the coverage depths and percentages of all positions in total and of the targeted regions



Coverage	bp (All positions)	% (All positions)	regions (Target regions)	% (Target regions)
5x	309 140	96.4	1 483	85.4
8x	307 135	95.8	1 469	84.6
10x	306 320	95.6	1 467	84.5
20x	296 875	92.6	1 435	82.7
30x	275 785	86.0	1 346	77.5
50x	211 401	66.0	1 010	58.2
100x	82 567	25.8	397	22.9
300x	1 682	0.5	9	0.5

Variant type summary

Variant type summary

Туре	Variants observed	Clinical	ly reported	In di	SNP	In Gn	omAD	Novel*		
Total	20 953	915	4.37%	15 767	75.25%	9 655	46.08%	5 051	24.11%	
SNV	12 894	200	1.55%	9 354	72.55%	3 809	29.54%	3 507	27.20%	
Deletion	4 172	369	8.84%	3 565	85.45%	3 199	76.68%	530	12.70%	
Insertion	3 135	345	11.00%	2 824	90.08%	2 647	84.43%	286	9.12%	
Substitution	752	1	0.13%	24	3.19%	0	0%	728	96.81%	

"Novel" refers to variants not found in dbSNP, gnomAD, or clinical databases.

Number of identified variants by ACMG class*:

The matrix presents the number of automatically classified variants in each of the 5 standard pathogenicity classes.

Number	of iden	tified	variants	by	ACMG	class*:
--------	---------	--------	----------	----	------	---------

ACMG class		
Benign	23	
Likely Benign	2	
Uncertain Significance	0	
Likely Pathogenic	0	
Pathogenic	0	
Total	25	

* (NOTE this classification is of the variant itself, and doesn't say how it affects this patient. For example the variant can be Pathogenic but it may not be causative if it's recessive and the patient is heterozygous)



Summary for ACMG Rules

Display of the number of times (and its percentage) each Germline Variant Classification rule was triggered.

Total	65	
BA1	22	33.85%
BP4_Strong	21	32.31%
BP6_Moderate	6	9.23%
BP6_Very Strong	4	6.15%
BP4	3	4.62%
BP7	2	3.08%
PM2_Supporting	2	3.08%
BP3	1	1.54%
BP6	1	1.54%
BP6_Strong	1	1.54%
BS1	1	1.54%
BS2	1	1.54%

Summary for ACMG Rules

Variants found in Coding Regions

Variants found in Coding Regions :

Туре	Total	In	ClinVar	Ir	dbSNP	In GnomAD		
SNV	4	3	75.00%	4	100.00%	4	100.00%	
Insertion	1	0	0.00%	1	100.00%	1	100.00%	

-							-	
Туре	Total	In	ClinVar	Ir	dbSNP	In GnomAD		
SNV	7	6	85.71%	7	100.00%	4	57.14%	
Deletion	1	0	0.00%	1	100.00%	1	100.00%	
Insertion	1	1	100.00%	1	100.00%	1	100.00%	



Quality Control report for CNV analyses

The CNV quality control report is a useful tool to evaluate the performance of a CNV analysis and to check the correlation between the control samples chosen for each test sample in a given cohort.

It can be accessed from the CNV main analysis as shown above, and provides the following information:

- <u>SampleId:</u> VarSome Clinical unique ID assigned to the sample
- <u>User sample name</u>: the name given by the user for each test sample.
- <u>Median fragment count</u>: It is the median fragment count in each genomic interval of the assay .bed file. These are data generated by exomedepth and used to ensure that a sufficient number of reads is present in each sample. Samples should have a similar range of values (same order of magnitude).
- <u>Number of reference samples:</u> number of reference samples used as controls for this test sample.
- <u>Reference sample names</u>: the names of the samples used as a reference (control) set.
- <u>Correlation</u>: correlation coefficient between the test sample and its reference samples.
- <u>Sex:</u> sex of the samples in the cohort

SampleId	User sample name	Median fragment count	Number of reference samples	Reference sample names	Correlation	Sex
37127	NA12878_WGS_PIPE_262	1225	1	ERR174310_WGS_PIPE_262	0.9847	F
37141	ERR174310_WGS_PIPE_262	3438	2	NA24385_WGS_PIPE_262 NA12878_WGS_PIPE_262	0.9781	м
37142	NA24385_WGS_PIPE_262	8722	1	ERR174310_WGS_PIPE_262	0.9738	м

We highly recommend users check the CNV QC report first. This report provides useful information on how the reference set of samples was generated for each of the samples of the cohort. It also checks that the median coverage of each sample was sufficient.

Each input sample is compared against an optimized set of reference samples chosen from among the other samples of the cohort. The set of reference samples is automatically generated by the algorithm for each sample of the cohort (test sample). Please bear in mind that the reference set of samples might not include all the other samples of the cohort. The reference set is chosen based on how well the coverage correlates with that of the test sample. The algorithm will choose the largest set of samples from the reference with the greatest correlation coefficient.

A sample will not be included in the reference set if it results in a decrease of the correlation coefficient.

In the CNV QC report, the user can check how many samples were used to construct the reference set for each test sample and their correlation coefficient. A high number of samples (at least 2) used for the reference set and a high correlation coefficient (> 0.97 for gene panels and > 0.98 for exomes) would be indicative of a good performance and a higher reliability of the CNVs found in that test sample. If the coefficient of correlation is too low for a test sample, the algorithm won't be able to generate meaningful results, and no CNVs will be called in that test



sample.

Calling CNVs on the sex chromosomes can create issues if the test sample and the reference samples it is being compared to are not gender matched. To make reliable calls on the X and Y chromosomes, you need to make sure that all samples are of the same sex.

The information from the three different metrics that can be found on the column <u>call quality</u> <u>control</u> can be used in combination with the CNV QC report.

For example, when the second call quality metric (number of reference samples) fails for all variants of the variant table:

γPι	iblic CNV sample	6 - CNV	Resu	Its of	Public CNV	sample 6 (hg19)	Public CNV sar	nple 6	CNV Results of Public CN	V samp CNV Result	s of Public CNV sample 6	+2 more D
0	0- 🖸 A 🐴 K	->									Sear	ch for gene, chr (e.g. chr1
0	Variant 0	Length ()	Call Q	uality	Copy Number	Туре	Ge ()	0	Quality Score	ACMG Class	CNV Rules	Number of genes
	chrX:1486166951210078	36 348 409	× 3	< ×	1.38957	deletion	ACAA2		51.1	(attegers)	Core Linizare Overlap	605
	chrX:670215301331194	66 097 967	1	c ×	1.51197	deletion	AARSD		29.1	Pathogenic	Gene Literature Overlap	956
	chrY:67379336737973	40	1	• •	3.02857	duplication	AMELY		15.8	Benign	Literature	1
	chr1:1734908317371404	22 321	1	e 🗸 .	1.53804	deletion	SDHB		15.7	Likely Pathogenic	Overlap	1
0	chr8:145742414145743	775	1	• •	2.28688	duplication	RECQL4		7.36	Pathogenic	Literature	1
	chr6:31366575.31366616	41		• •	1.39835	deletion	MICA-AS1		4.47	Berign	Gere	1
κ.				-				-				
Warnis	Raming: Number of reference samples low.											

we should go to the QC report and check the size of the reference set. In this example, the reference set used to call CNVs on "Public CNV sample 6" has only one sample and it is therefore below the threshold (2 samples), and this is why the second call quality metric fails for all variants. In this case, we suggest repeating the CNV analyzes with other samples that could be better correlated with the test sample.

Sample name	Coverage (Median)	Number of reference samples	Reference sample names	Correlation	Gender (Predicted)	Sex (User defined)	wsex (Consensus)
Public CNV sample 1	604	2	Public CNV sample 3 Public CNV sample 2	0.9975	_	_	_
Public CNV sample 2	673	3	Public CNV sample 4 Public CNV sample 1 Public CNV sample 7	0.9988	_	_	_
Public CNV sample 3	739	2	Public CNV sample 1 Public CNV sample 4	0.9976	-	_	_
Public CNV sample 4	784	3	Public CNV sample 3 Public CNV sample 2 Public CNV sample 9	0.9987	-	-	-
Public CNV sample 5	1256	1	Public CNV sample 8	0.9954	-	-	-
Public CNV sample 6	1099	1	Public CNV sample 8	0.9882	-	_	_

View FASTQC report

A quality control report for high throughput sequence data. For more information please see the documentation of the FASTQC tool.



In order to find the FastQC report of an analysis in VarSome Clinical you can go either:

- to the Analyses page and select from the burger menu of the analysis you wish the "View FastQC Report" as shown below, or



- in the Variant Table page you can click on the "Analysis actions" button and select the same option.

		Tags About	-
	🕇 Filters 🛛 🛓	Report generation	Analysis actions 👻
ır1 o	۹ 📋	ଟ + ₹-	SNVs & Indels Sample phenotype(s)
	Inheritance	Function Zy	A Sample/Analysis information
	AR	8 🙃 👦	View QC report
			View FastQC report
	AR	C 😳 😳 👽 🔻	Region list coverage report
		33 10	Downloads
	. 2	30	Re-Annotate Analysis
		35 30 in	New Gene-List Analysis
		•	New Algorithmic Filter Analysis
		•	New CNV Subanalysis
		30 in nc	New Repeat Expansion Sub-Analysis
	AR	50 in	Archive sample data
		af au (in	Delete FASTQ sample data
	-	0.00	Delete BAM sample data



The different analyses (modules) performed for the FASTQC report are summarized on the left of the screen, for each FASTQ file used for the analysis.

Summary	
Basic Statistics	
Per base sequence quality	
Per tile sequence quality	
Per sequence quality scores	
Or Per base sequence content	
Ser sequence GC content	
Per base N content	
Sequence Length Distribution	
Sequence Duplication Levels	
Overrepresented sequences	
Adapter Content	

The symbol on the left of each module corresponds to a flag of "Passed", "Warn" or "Fail".

Here, we include examples for each module with what should be expected as a result:

Basic Statistics

Simple information about input FASTQ file: its name, type of quality score encoding, total number of reads, read length and GC content.

Basic Statistics	
Measure	Value
Filename	short_S1_L001_R1_001.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	9142
Sequences flagged as poor quality	0
Sequence length	35-151
%GC	56



Per base sequence quality

A box plot showing aggregated quality score (Phred score) statistics at each position along all reads in the file.



Per tile sequence quality

The graph allows you to look at the average quality scores from each tile across all of your bases to see if there was a loss in quality associated with only one part of the flow cell.





The picture above shows an ideal case, where no quality loss is reported in any tile of the flow cell. On the contrary, in the picture below you can see that certain tiles show consistently poor quality. The colors are on a cold to hot scale, with cold colors being positions where the quality was at or above the average for that base in the run, and hotter colors indicate that a tile had worse qualities than other tiles for that base.





Per sequence quality scores

A plot of the total number of reads vs the average quality score (Phred score) over the full length of that read.



Per base sequence content

This plot reports the percent of bases called for each of the four nucleotides at each position across all reads in the file.





Per sequence GC content

Plot of the number of reads vs. GC% per read. The displayed Theoretical Distribution assumes a uniform GC content for all reads.



Per base N content Percent of bases at each position or bin with no base call, i.e. 'N'





Sequence Length Distribution

Shows the distribution of reads lengths over all sequences.



Sequence Duplication Levels

Percentage of reads of a given sequence in the file which are present a given number of times in the file.



Percent of seqs remaining if deduplicated 91.62%



Overrepresented sequences

List of sequences which appear more than expected in the file.



Adapter Content

Cumulative plot of the fraction of reads where the sequence library adapter sequence is identified at the indicated base position.



Please note that if an error or warning is reported by FASTQC, VarSome Clinical will also provide a warning to the users. For more information see <u>User messages</u>.

Coding coverage report

You can use a previously created <u>gene lists</u> to produce an Excel document reporting coverage information of the coding regions included in the analyzed gene list. This report can be exported for all component samples of a multi-sample analysis.



1	А	В	С	D	E	F	G	н	1	J	К	L	м	N	0	Р	
1	Gene	Region	Coding Length	Mean Cover	#bp<5x	#bp≥5x	%Cover≥5x	#bp<10x	#bp≥10x	%Cover≥10x	#bp<20x	#bp≥20x	%Cover≥20x	#bp<30x	#bp≥30x	%Cover≥30x	
2	ACTC1	chr15:3508688135087009	129	100	0	129	100	0	129	100	0	129	100	0	129	100	
3	ACTC1	chr15:3508544635085770	325	321	0	325	100	0	325	100	0	325	100	0	325	100	
4	ACTC1	chr15:3508460935084770	162	269	0	162	100	0	162	100	0	162	100	0	162	100	
5	ACTC1	chr15:3508429135084482	192	294	0	192	100	0	192	100	0	192	100	0	192	100	
6	ACTC1	chr15:3508331535083496	182	181	0	182	100	0	182	100	0	182	100	0	182	100	
7	ACTC1	chr15:3508261335082756	144	223	0	144	100	0	144	100	0	144	100	0	144	100	
8	ACTC1	total	1134	249	0	1134	100	0	1134	100	0	1134	100	0	1134	100	
9	ACTN2	chr1:236849974236850099	126	207	0	126	100	0	126	100	0	126	100	0	126	100	
10	ACTN2	chr1:236881158236881272	115	148	0	115	100	0	115	100	0	115	100	0	115	100	
11	ACTN2	chr1:236882194236882313	120	344	0	120	100	0	120	100	0	120	100	0	120	100	
12	ACTN2	chr1:236883405236883491	87	248	0	87	100	0	87	100	0	87	100	0	87	100	

Coverage report for targeted regions

Generates an Excel document for the coverage of the regions captured by the specific assay that you selected when launching the analysis.

	A	в											M	N			Q	R			
1	Coverage re	port on targe	ted regions																		
2	Region	Coding Lef	Mean Cove	Minimum	#bp<5x	#bp≥5x	%Cover≥5	#bp<8x	#bp≥8x	%Cover≥8	#bp<10x	#bp≥10x	%Cover≥1	#bp<20x	#bp≥20x	%Cover≥2	#bp<30x	#bp≥30x	%Cover≥3#	#bp<50x	#bp≥50x
3	chr1:69091	918	0	(918	0	0	918	0	0	918	() (918	0	0	918	0	0	918	0
- 4	chr1:62109	939	0	(939	0	(939	0	0	939	() (939	0	0	939	0	0	939	0
5	chr1:86132	72	0	(72	0	(72	0	0	72	() (72	0	0	72	0	0	72	0
6	chr1:8655 3	182	0	(182	0	(182	0	0	182	() (182	0	0	182	0	0	182	0
7	chr1:86641	51	0	(51	0	(51	0	0	51) (51	0	0	51	0	0	51	. 0
8	chr1:87119	125	0	(125	0	(125	0	0	125	() (125	0	((125	0	0	125	0
9	chr1:87442	90	0	(90	0	(90	0	0	90	() (90	0	0	90	0	0	90	0
10	chr1:87469	186	0	(186	0	(186	0	0	186	() (186	0	0	186	0	0	186	i 0
11	chr1:87652	163	0	(163	0	(163	0	0	163	() (163	0	0	163	0	0	163	0
12	chr1:87751	116	0	(116	0	(116	0	0	116	() (116	0	0	116	0	0	116	i 0
13	chr1:87779	79	0	(79	0	(79	0	0	79	() (79	0	0	79	0	0	79	0
14	chr1:8779ૐ	500	0	(500	0	(500	0	0	500	() (500	0	0	500	0	0	500	0
15	chr1:87863	125	0	(125	0	0	125	C	0	125	(0 0	125	C	C	125	C	0	125	0

You can find it under the "Analysis actions" tab, as shown below,





or on the Analyses page if you follow the steps shown below:



Region list coverage report

You can use a list of public/custom regions of interest to produce an Excel document reporting coverage information about these regions. Contact us in case you want to add regions to your analysis. In order to find the Region list coverage report of an analysis in VarSome Clinical you can go either:

- to the Analyses page and select from the burger menu of the analysis you wish the *"Region list coverage report"* or

~	varsomeclinical	
i di la	Description.	
	SNVs & Indels	
	Sample phenotype(s)	nple
	Sample analysis information	
	Reuse sample files	
	Gene coverage	
	View QC report	PI_sa
	View FASTQC report	
	Coding coverage report	PI_sa
	Region list coverage report 2)
	Downloads	
	Coverage report for targeted regions	
	Archive sample data	ər_koi
	Delete FASTQ sample data	
	Delete BAM sample data	HT_fa
	Re-annotate analysis	
	New Gene-List analysis	PI_sa
	New algorithmic filter analysis	
$\overline{}$	New CNV Sub-Analysis	FAST
•	New Repeat Expansion Sub-Analysis	Somati
	= 5000000000000000000000000000000000000	
	Description:	
Region List Coverage Report		3
Please select a region list from the available options		
Select		
Download		

in the Variant Table page you can click on the "Analysis actions" button and select the same option

V	arso	meclinical
Tags	About	*
a 🛃 Report ge	eneration	Analysis actions -
1 8 C	• T •	SNVs & Indels Sample phenotype(s)
Inheritance	B Function	Sample/Analysis information
AR	G A	Reuse sample files
	••	Gene coverage
		View QC report
AR	6	View FastQC report
AR	0000	Coding coverage report
	-	Region list coverage report 2
AR	i	Downloads
-	6	Coverage report for targeted regions
-	6	Re-Annotate Analysis
	i 0	New Gene-List Analysis
AR		New Algorithmic Filter Analysis
	•	New CNV Subanalysis
AR	5 in	New Repeat Expansion Sub-Analysis
		Archive sample data
	< 1_	Delete FASTQ sample data
		Delete BAM sample data



Downloads

This option directs you to a new screen where you can see any files associated with this analysis that are ready to be downloaded.

Sample Data Files:

- VCF file: A compressed (*.vcf.gz) VCF file will be downloaded with the results of the variant calling.
 - For sub-analysis (gene list analysis and algorithmic filters) there is the new option to "Generate VCF" that contains **only** the filtered variants.
- BAM file: Download the bam file (the sample's reads aligned against the reference genome) used in the analysis. For multi sample analyzes you will find the BAM (and BAI) files for each component sample.
- **BAI file:** The BAI file format is the index file of a BAM file. This is a companion file for



your previous BAM file, which doesn't contain any sequence data but acts as an external table of contents allowing a computational tool to navigate in the BAM file and locate specific parts.

Quality control File:

- QC report (PDF)
- QC report (docx)

Download PGX report

<u>PharmCAT</u> is a pharmacogenomics clinical annotation tool that generates a report containing genotype-based prescribing recommendation. This option is available for WGS samples, run against hg38 and using one of the WGS capture methods: 'WGS+PCR' or 'WGS-PCR'.

- The report comprises 4 sections:
 - Genotype Summary: This section has a summarized table of the matched genotypes and the following associated clinical annotations: Drug, Gene, Allele Functionality and Phenotype
 - Prescribing Recommendations: This section includes a list of recommendations per associated drug, along with bibliographical reference material
 - Allele Matching Details: Detailed information about how data in the sample VCF matches up with haplotype definitions
 - Disclaimers: PharmCAT disclaimer (in blue) as stated in their <u>website</u> and the disclaimer about the default running parameters in Varsome Clinical (in yellow), as shown below.



Typically in WGS samples, a large fraction of the PGX relevant positions that are considered by pharmCAT are absent from the input vcf. If those missing positions are assumed to be "no-call", the resulting PGx report results in the output of multiple possible genotypes, and hence less specific treatment recommendations.

In VarSome Clinical, we assume that all missing PGx relevant positions are homozygous
VarSome Clinical User ManualVersion: 11.9.1 - 19th December 2023Page 130 of 254



reference. However, this may not reflect reality, because such positions may in fact be unreadable or uncallable. Running PharmCAT with positions as missing vs reference can lead to different results.

Re-annotate Analysis

Reannotating will not change the list of identified variants (we do not perform the calling again), it will simply update the annotations for the same list of variants using the data available on the day of re-annotation. This means that some Germline Variant Classification might change if they are affected by newer data. Re-annotating an analysis is charged at 50% of the price of the original analysis.

New Gene-List Analysis

This option will re-start the analysis with the same sample using a previously generated gene list of selected genes or a gene list created from selected phenotype(s) or disease(s). It will appear as a sub-analysis of the initial analysis.

Add BAM file(s)

Select a BAM file for alignment visualization. If your sample type is VCF, you can upload and add a BAM file to visualize the alignment.



Once the BAM file is added to a sample, IGV and JBrowse links will be displayed on the "Coverage" column of the Variant Table. These links will allow you to visualize the alignments in each variant position.

Please note that the BAM file will be taken into account when calculating storage fees. For more details please read the <u>Storage Management</u> section.

New Algorithmic Filter Analysis

It opens a new screen with the available algorithmic filters, as seen in <u>Algorithmic filters</u> section. Please click on the info icon, to get a description of each filter. Algorithmic filters create a snapshot of the parent analysis, with a filtered subset of variants, according to each Algorithmic filter. You can run a sub-analysis choosing a filter from the list or you can ask for a custom made filter.



New CNV Sub-analysis

Launching a CNV analysis from VCF as seen in <u>CNV annotation (from VCF)</u>.

New Repeat Expansions Sub-analysis

Launching a CNV analysis from VCF as seen in Repeat expansion annotation from VCF file.

Archive sample data

The analysis will be archived and a VCF file will be stored.

Delete FASTQ sample data

Allows to delete the fastq files used for the analysis. Once this option has been used, this will be disabled meaning that the files have been deleted from the server.

Delete BAM sample data

Allows to delete the bam file produced by the alignment during the analysis. Once this option has been used, this will be disabled meaning that the files have been deleted from the server. For more information on storage fees please refer to this article: <u>Data Deletion and Data Storage Fee explained</u>.

6.2 User messages

VarSome Clinical introduces a monitor to highlight quality issues related to user input data and ensure that they won't be missed. The easily accessible display consists of badges of different colors based on their severity (errors-red, warnings-amber and information-blue). These messages are visible both on "Dashboard"/"Analyses" and "Variant table" menus, per sample and provide useful information about the status of each sample during the quality control step.

Quality control is performed at the FASTQ (FASTQC tool) and the BAM level, to ensure that the raw data is good enough to reliably analyze the results. For each analysis, warning is or error icons may appear on the "Analyses" / "Dashboard" menu to inform the users that the quality control of the FASTQ file produced warnings and/or errors.





To reduce the number of error/warning messages displayed on the "Analyses"/ "Dashboard" menu and improve their practicality, FASTQC errors have been divided into three groups according to their frequency: Frequent, Standard and Rare. The Frequent messages include Per base sequence content as well as Per sequence GC content. Frequent errors will not display any warning or error message, therefore their messages will not appear as warnings or errors on the "Analyses" / "Dashboard" menu.

On the other hand, Standard errors include:

- Per tile sequence quality
- Overrepresented sequences
- Adapter Content
- Sequence Duplication Levels

and will display a single warning when there are more than two present in the report.

Note: Sequence Duplication Levels will be ignored when the analysis uses an amplicon kit.

		\otimes
	5 0 Please note that the information displayed below may be updated with new messages until a few minutes	
A products	arter the analysis has inlined.	
Read File	.rastq.gz: Per base sequence content	
\Lambda Read file	.fastq.gz: Per sequence GC content	
\Lambda Read file	.fastq.gz: Per sequence GC content	
? Read file	fastq.gz: Per tile sequence quality	
? Read file	.fastq.gz: Sequence Length Distribution	
? Read file	.fastq.gz: Sequence Duplication Levels	
? Read file	fastq.gz: Overrepresented sequences	
? Read file	fastq.gz: Per tile sequence quality	

Finally, Rare errors include:

- Per base sequence quality
- Per base N content
- Per sequence quality scores
- Sequence Length Distribution

and will display an error message when there is one or more of them present in the report.

To know more about what parameters of FASTQC have raised the flag, you can click on the



icons, and a pop-up window will appear showing you a detailed description.



The same information is also available in the variants table for an individual analysis:

2	± 0+ @ ≓ •+ d	(hg19) 01		_			
□	Variant 🔶	Variant Type	🕚 Gene Symbol	0	Germline Class 🝦	Germline Rules	HGVS
	chr21:44488824 delAA 🎭	Deletion (2)	CBS 🌏	••	Likely benign	894	ENST000003981
	chr21:44478031 C⇒T	SNV	CBS 🌯		Likely benign	BP4 Strong PM2 Supporting	ENST000003981
	chr21:44476003_4 insT	Insertion (1)	CBS 🌏		Benign	BA1 BP4	ENST000003981
	chr21:44477878 G⇒C	SNV	CBS 🌏		Benign	BA1 BP4 Strong PM2 Supporting	ENST000003981

In the same way, there are some quality checks run at the alignment level that will produce warnings and/or error messages in case there are issues with the coverage of the sample. Additionally, if one sample included in a CNV cohort analysis does not have enough coverage, the whole CNV analysis will fail and an error message will be displayed informing about the

problematic sample. VarSome Clinical User Manual



	Analysis	State	Phenotypes		Type Ger	nome Date	Variants	Tags	Id	User	Assay		
\equiv >	CNV analysis from n1 n2 n3 n4 n5 (n Description:	(O New)			h	ng19	0	00	0				
	·												
													(\mathbf{x})
	CNV analysis from		n1	1	ן 2ר		n3		Г	אר		n5	\cup
				1Δ									
	Please note tha	t the infor	mation displayed	d below may be u	update	d with new me	ssages un	til a few m	inutes				
			arter	r une anatysis na:	51111511	leu.							
2	1 The coverage of sample(s)	n1	n2 i	n3 n4		n5 is too l	ow to be	e used ir	n a CN	NV ana	lysis		

6.3 Tabs

If multiple analyses have been performed using the same sample, they will be organized as tabs adjacent to the current analysis's title, with the first tab being the main analysis.

When there are more than three analyses, a total of four tabs will be displayed. These include the main analysis, the two most recent sub analyses, and a tab indicating the remaining number of subanalyses.

•	varsomeclinical	Analyses U	pload / view files	s Manage + Launo	ch analysis 👻	Filter sets Gene lists	Help						Tag	is About	
~	_WGS_	- Cl	VV Results	of	_WGS_	(hg19) CN	IV analysi	s fromWGS	CNV Results of	_WGS_P 🕲	CNV Results ofW	GS +1 more D	🕇 Filters 🕹 Report	generation Analys	is actions 👻
	0- 🕐 🕐 🍕 🤇	>									Search	for gene, chr (e.g. chr1 c	৭ 📋 🖉	• • •	∎- <u>≵</u>
Ο.,	Variant ϕ	Length 0	Call Quality	Copy Number	Туре 🕴	Genes 🕴	0	Quality Score	Germline Class	CNV Rules	Number of genes 🕴	Number of exons	Reads expected 🕴	Reads observed	Read
	chrY:26000019900000	7.299.999	× x x	0	deletion	AGPATSP1,AMELY,A	• •	2090	Pathogenic	Gene Overlag	132	691	58.209	1.363	0.0234
	chrX:61750001154700000	92.949.999	× × ×	2.61353	duplication	AARSD1P1,ABCB7,		2390	Pathogenic	Gene Overlap	1.510	11.215	1.418.283	2.166.554	1.53
	chrX:270000152150000	49.449.999	× x x	2.61353	duplication	ACAA2P1,ACE2,AC		1280	Pathogenic	Gene Overlap	759	6.173	776.078	1.187.039	1.53
0	chrY:1390000128850000	14.949.999	× x x	0	deletion	ACTG1P11,ACTG1P2		3460	Pathogenic	Gene Literature	345	1.891	86.190	1.092	0.0127
	chrX:5250000158500000	5.999.999	× x x	2.61353	duplication	ACTG1P10,ALAS2,A		153	Pathogenic	Gene Overlap	140	1.150	89.760	137.181	1.53
0	chr14:106200001106750000	549.999	× × ×	2.22651	duplication	ADAM6,ATP6V1G1P			Pathogenic	Gene	96	171	5.022	5.878	1.17
	chr22:2240000123250000	849.999	× × ×	1.07587	deletion	ABHD17AP5,ASH2L		46.7	Uncertain Significance	Gene Literature	119	284	16.340	8.617	0.527
	chrY:1315000113750000	599.999	× × ×	1.42868	deletion	ACTR3BP1,DUX4L1		55.3	Uncertain Significance		13	21	21.429	14.432	0.673
	chr10:5120000151400000	199.999	× x x	3.09085	duplication	ENSG00000174194,		8.91	Uncertain Significance	-	13	71	207	440	2.13

To view and access all the sub analyses associated with the main analysis, users can click on the arrow located to the left of the analysis title in or on the last tab with the title "+ n more",

where n is equal to the remaining number of sub analyses (for example). This action will reveal the analysis tree, presenting all the analyses in chronological order.



7. Variant table

The results are displayed in the variant table. Rows contain the identified variants, and columns contain core annotations for each variant (Variant, Variant type, Class, Genes, Function, Zygosity, Allelic balance, Coverage). However, none of the columns is mandatory, you can choose the ones to be displayed using the "Columns" icon . The length of each column of the variant table can be adjusted by dragging the sides of the column headers.

Tip: The column order in the variant table is user-specific, meaning each user can set up a custom order and visibility of columns.

You can hover over the info icon i next to the column names to display the column information. The variant table is sorted by the Germline Variant Classification by default showing the most pathogenic variants first on the list. You can use the icon to sort the variant table by different columns (e.g. phenotypes, variant position, coverage) in ascending or descending order. Use the "Reset/refresh" icon to return the variant table to its original state.

The variant table can be accessed by the user who requested the analysis or by other people belonging to the same group.



V S (hg19) © s © Analysis filtered by Gene list from. O 1								🝸 Filters 🛛 🛓 Report g	generation 🧳	Analysis actions 👻				
	± 0• ⊵ ≓ •	• • • • ?;	100 C							Sear	ch for gene, chr (e.g. chr1	ୁ ବ 📋 😂	• T- +	- ≣ - <u>≵</u>
α.	Variant 0	Variant Type	Gene Symbol	0	Germline Class	Germline Rules	Phenotypes	HGVS	HGVS Protein	HGVS Coding	Transcript Position	Overlapping Genes	Inheritance	e 🕒 Function
	chr1:116247826 T⇒C 🎭	SNV	CASQ2	•••	Likely pathogenic	PP3 Strong PM2 Supporting PPS	2	NM_001232.4:c	D309G(p.(Asp30	c.926A>G	exon 9 of 11 positio	CASQ2	AD/AR	8 😳
	chr1:25880262 delAAAAA	Deletion (5)	LDLRAP1		Likely benign	574	0	NM_015627.3:c	p.?	c.89-137_89-133del	intron 1 of 8 positio	LDLRAP1	AR	(1) (1)
	chr1:237656125 delAAGG	Deletion (4)	RYR2		Likely benign	BP4 Moderate PM2 Supporting	2	NM_001035.3:c	p.?	c.1828-86_1828-83del	intron 18 of 104 pos	RYR2	AD	6
	chr1:237958751 delT	Deletion (1)	RYR2		Likely benign	8P4	2	NM_001035.3:c	p.?	c.13956+131del	intron 96 of 104 pos	RYR2	AD	6
	chr1:6819612 delAAAAA	Deletion (5)			Likely benign	BP4 Moderate	0	•	•	•	•	-	-	
	chr1:142803275_6 insAAC	Insertion (3)	·		Likely benign	BP4 Moderate	0	•			227 bp before trans	ANKRD20A14P,ENSG00		8 🙃
	chr1:201331068 A→G 🎭	SNV	TNNT2	••	Likely benign	BS2 BP6 Moderate PP2 PP3 PP3	1	NM_001276345	I231T(p.(Ile231T	c.692T>C	exon 14 of 17 positi	TNNT2	AD	8 8 8 8
0	chr1:31124691 T→C	SNV	•		Likely benign	8P4 Strong	0	•			•	-	-	
	chr1:31124690 A⇒C	SNV			Likely benign	8P4 Strong	0	-			-	-	-	
0	chr1:35057230 delA	Deletion (1)	•		Likely benign	851 894	0			-		-	-	

7.1 Description of results page functionalities

Columns for Germline/Somatic samples:

- Variant: The variant's sequence and genomic location.
- Variant type: SNV (single nucleotide variant); for INDELs and substitutions, the number of nucleotides affected are shown.
- **Gene Symbol:** Gene used for annotation and classification of the variant for ACMG (& AMP for somatic samples)
- User variant classification: custom classification for variants marked by the user. User classifications are also available for AMP and for ACMG rules when the user clicks on "Save as manual classification" below the ACMG/AMP verdict. The custom classifications are linked to the variant and will be displayed in other analyses of a user's group if the same variant is found.

More specifically, you can classify a variant using either any of the five standard classes of pathogenicity ("Benign", "Likely benign", "Uncertain significance", "Likely pathogenic" and "Pathogenic") or any of the pre-set classifications as shown below, by clicking on the

••• icon:



Class	sify variant					
	0 - ि ≓ ● • 6	•				
	None					
□.	Global classifications	Your classifications	.ss 🔶	ACMG Rules	HGVS	HGVS Prot
	😑 Benign	 Artifact 	nic	PP5 Very Strong PM1 Strong	NM_004333.6:c	L597V(p.Lei
	🔵 Likely Benign	bluered				
	🛑 Uncertain Significance	Pharmacogenomics	enic	PM1 Strong PM5 Strong PP5	ENST0000269	R175L(p.Arg
	Likely Pathogenic	QWEROIUQWA098E	78900WA 09	PVS1 PM2 PP5	NM_001164665	R18Afs*64(
	🛑 Pathogenic	SampleClassification	_			
	Common Artefact	🛑 test	enic	PM1 Strong PM2 Strong PM5	ENST0000269	V172D(p.Va
	Drug Response	 Test 		DM1 Strong DS2 DM2 Support	ENST0000269	0167K(n GI
	Disease Association	test_manual_classific	ation	PMI Strong PSS PM2 Support	EN310000203	Q10/1K(p.O
	😑 Risk Factor	to review	ogenic	PVS1 PM2	NM_001127208	P463Lfs*23
	Protective		_	-	511070000000	
	 Phenotype Association 		nificance B	BP4	ENS10000269	-
	🛑 Tier I		nificance B	BP4	NM_000256.3:c	
	🛑 Tier II					
	🛑 Tier III		D	BS1 BS2 PVS1 Strong	NM_022489.4:c	V1247Dfs*8
	Tier IV					

You can also create your own classification tags. To do this, click on the 🚺 icon located above the variant table:

~	of the Link of	endigen	(hg19) 🕝 s		🕜 Analysis fil	tered by Gene list fro 01	D					🕇 Filters 🕹 Report	generation An	alysis actions 👻
0	± 0+ @ ≓ •	- # 14 13								S	earch for gene, chr (e.g. chr1	ୁ ବ 🝵 😂	• T- +	≣• ≛
۵.	Variant 🕴	Variant Type	Gene Symbol	0	Germline Class	Germline Rules	Phenotypes	HGVS	HGVS Protein	HGVS Coding	Transcript Position	Overlapping Genes	Inheritance	Function
	chr1:116247826 T⇒C 🍋	SNV	CASQ2	•••	Likely pathogenic	PP3 Strong PM2 Supporting PPS	2	NM_001232.4:c	D309G(p.(Asp30	c.926A>G	exon 9 of 11 positio	CASQ2	AD/AR	G 😳
	chr1:25880262 delAAAAA	Deletion (5)	LDLRAP1		Likely benign	6 P4	0	NM_015627.3:c	p.?	c.89-137_89-133del	intron 1 of 8 positio	LDLRAP1	AR	00
	chr1:237656125 delAAGG	Deletion (4)	RYR2		Likely benign	BP4 Moderate PM2 Supporting	2	NM_001035.3:c	p.?	c.1828-86_1828-83d	el intron 18 of 104 pos	RYR2	AD	0
	chr1:237958751 delT	Deletion (1)	RYR2		Likely benign	6 P4	2	NM_001035.3:c	p.?	c.13956+131del	intron 96 of 104 pos	RYR2	AD	0
	chr1:6819612 delAAAAA	Deletion (5)			Likely benign	BP4 Moderate	0		•	÷				
	chr1:142803275_6 insAAC	Insertion (3)			Likely benign	BP4 Moderate	0		•	•	227 bp before trans	ANKRD20A14P,ENSG00		00
	chr1:201331068 A⇒G 🌯	SNV	TNNT2	••	Likely benign	BS2 BP6 Moderate PP2 PP3 PP3	1	NM_001276345	I231T(p.(Ile231T	c.692T>C	exon 14 of 17 positi	TNNT2	AD	0 0 0 0
	chr1:31124691 T⇒C	SNV			Likely benign	BP4 Strong	0		•		-	•		
	chr1:31124690 A→C	SNV			Likely benign	BP4 Strong	0	-	-	-	-	-	-	
	chr1:35057230 delA	Deletion (1)	+		Likely benign	853 894	0							

A pop-up window will be shown where you can add your custom tags. You need to give a name to the tag and a unique classification code. Finally, you can choose a color for your tag and then just click "Save".



lassifications				
Available Variant Classificati	ions			Create a classification
Artifact (12345678)	bluered (bluered)	Custom_Classification_1 (CC1)	Custom_Classification_2 (CC2)	Name Name
Pharmacogenomics (123)	QWEROIUQWA098E7890 OWA 09OWE (12123)	SampleClassification (123)	Test (T)	Code
				Fill in a unique classification code - 10
lest (1)	TeST (test)	(123)	(136)	Color
to review (R)				
				Description
				Description
				Save

The new custom classification will be available together with rest of the options:



Custom Germline & Somatic Variant Classifications

The Germline and Somatic Variant Classifications allow users to modify the set of triggered rules if they do not agree with the verdict. Once you have modified the set of rules, you click on the "Save as manual classification" icon.



This will be saved as a custom classification attached to the variant. If you hover over with the mouse you will be able to see the manual classification with the set of rules used and the user who made the classification (in brackets).

Germline Variant Clas	ssification Version: 11.8		Terms of use Documentation
		Pathogenic 17 points = 17 P - 0 B (0) Using transcript NM_007264.4, MANE select.	Submit to ClinVar Save as manual classification
Sample Information		Findings	
B Mode of Inheritar	ce	AD/AR, based on gene information from CGD, ClinGen Disease Validity, Gen/CC, Mondo, OMIM and 2 more.	
Automated criteria	Show summary view 🔵 Enable clinical evidence 0		
Rule	Explanation		Show failed criteria
PVS1 0 Very Strong V	Null variant (nonsense) in gene BRCA1, predicted to cause NMD. Loss-of-function is a kno pathogenic variants. The truncated region contains 1 018 pathogenic variants.	own mechanism of disease (gene has 3 404 reported pathogenic LOF variants). The exon affects 1 functional domain: UniProt protein BRCA1_HUMAN region of Int	erest 'Disordered'. The exon contains 87
PP5 0 Very Strong V	ClinVar classifies this variant as Pathogenic, 3 stars (expert panel, reviewed Jul '23, 44 sub LOVD classifies this variant as Pathogenic. VarSome users have classified this variant as Pathogenic, citing 31331294, and as Pathog	bmitsions), cling 11 and/cls (2044311, 3308/7929, 3285571, 32854451, 32467255 and 6 more). genic (automatically lifted over from chr17-43062434 G=A on hg/30, cling 31331254.	
PM2 0 Supporting V	GromAD genomes homozygous allele count = 0 is less than 2 for AD/AR gene BRCA1, go GnomAD exomes bemozygous allele count = 0 is less than 2 for AD/AR gene BRCA1, go	od gnomAD genomes coverage = 31.8. d gnomAD exomes coverage = 71.6.	
PS1 0 Strong v	Not a missense variant.		
PS2 0 Strong v	No phenotypes or diseases provided.		

Potential artifacts

This can help you to identify potential artifacts by comparing the number of times a variant is observed within the samples of the group. The system will label the variants as "Automatically Tagged Likely Artefact (VarSome)" if:

- The group has more than 20 samples analyzed
- The variant has a population frequency lower than 1% (according to gnomAD)
- The variant is found in more than 10% of the samples.

Chr2:215802332 G⇒A SNV		Automatically Tagged Likely Artefact (VarSome) - updated: 2023-12-08					
chr11:5248004 G⇒A	SNV	HBB	• • •	D	Pathogenic	e	

Comment on variants

You can attach comments to variants by clicking on the "Comments" icon:



9	0- ₪ ≓ ⊴-	et				
□	Variant	Variant Type	Gene Symbol	0	Somatic Tier 🔶	Somatic Rules
	chrX:76972563 delA	Deletion (1)	ATRX		Tier IV	Drug I) Germ IV Soma Path
	chrX:76972554 delAA	Deletion (2)	ATRX		Tier III	Drug II Path I Freq I
	chrX:76954630 delT	Deletion (1)	ATRX	•	Tier IV	Drug I) Germ IV Path I Freq IV
	chrX:76954591 delA	Deletion (1)	ATRX	••	Tier IV	Drug II) Germ IV Path I
	chrX:76953295 delAC	Deletion (2)	ATRX		Tier IV	Drug II Germ IV Path I Freq I
	chrX:76940534 A⇒G	SNV	ATRX		Tier IV	Drug II) Germ IV Path I Type IV Freq II Pred IV

A pop-up window will be shown where you can add your comment. The comment will be attached to the variant and it will be shown in other samples of your group where the same variant is found. The comments are also private for your group by default.

Make a comment on this variant
Write your comment on the variant. By default comments you make can only be viewed by you and people in your department unless you select to share the comment in which sample anyone will be able to view it
Share comment outside your group
By default only users in your group can see your comments
The comment is specific to this sample only
Specific to sample comments appear only in analyses related to the sample and not any other analyses even if the variant is the same
any other analyses even if the variant is the same

- Share comment outside your group: if you click this option the comment will be shared with the VarSome community of users.
- The comment is specific to this sample only: if you choose this comment, the comment will be attached only to that sample and it won't be shown in other samples even if the same variant is found.

All custom classifications, along with other users' actions can be monitored by the group's supervisor via the audit trail tab.

varsomeclinical										
Tags	Audit Trail	About	•							
			Dashboard							

If your group does not have a group supervisor or you would like to transfer this role to another user please contact our <u>support team</u>.

The Audit Trail shows the record of different actions that have been made from all users of a group on the samples analyzed. This log is only visible to the group supervisor.

The group supervisor can retrieve specific actions recorded in the Audit Trail by using the filters which allow to filter by user, action type and date.

Search for user:		
Filter by action:		
Select		· · ·
From Date:	To Date:	

The Audit Trail shows the following action types performed by users of the same group:

- Sign in / sign out: who has signed in and out and the IP address.
- View / access a URL: who has accessed which analysis.
- Accepted analysis terms: users who have accepted the analysis terms before accessing the results. Note that the users need to accept the analysis terms the first time they access the results. This is not required for analyses for which they have already accepted the terms.
- View analysis messages: sometimes there are messages associated with the analyses. These messages are related to the quality of the input samples and are aimed to warn users of potential quality issues that may require the user's attention. If a user checks these messages, this action is recorded in the Audit Trail.
- Analysis launch: users who have launched an analysis.
- Variant classification: when users add, modify or remove manual variant classifications, this is also stored in the Audit Trail.
- Analysis filtering: users who have applied filters in analyses.



• Variant export: variants selected for export and which user has selected them.

Audit Trail		Dashboard
search nor user:		
Filter by action:		
Expected variance) X Classification Addition X Launched an analysis X \times		
From Date: To Date:		
regor (mont)		
Description	Date added	Action
Mr Qa Automation selected for export variant chr7.140453138 AmC in analysis for_kon_update.C1367.7412587415	Nov 14, 2023	Exported variant(s)
Mr Qa Automation classified variant (http://140453136 A++G as Pathogenic in analysis for_kon_update.C1367.7412587415	Nov 14, 2023	Classification Addition
Mr Qa Automation selected for export variant chr7.140453136 A=C in analysis for_kon_update.vcf hg19.C25.2189559268	Nov 14, 2023	Exported variant(s)
Mr Qa Automation classified variant chr7:140453138 A+G as Pathogenic in analysis for_kon_update.vcf.hg19.C25.2189599288	Nov 34, 2023	Classification Addition
Mr Qa Automation classified variant dv?.140453136 A+vT as Tier IV in analysis HD827 TUMOUR AMP.C362 after_cron_job	Nev 13, 2023	Classification Addition
Mr Qa Automation selected for export variant chr7:140453136 A++C in analysis for_kon_update.C1367.7412587415	Nov 13, 2023	Exported variant(s)
Mr Qa Automation classified variant circl 234675807 A++G as Benign in analysis for_kon_update.C1367.7412587415	Nov 13, 2023	Classification Addition
Mr Qa Automation classified variant chr?:140453136 A+G as Pathogenic in analysis for_kon_update.C1367.7412587415	Nov 13, 2023	Classification Addition

Import list of custom classifications and variant comments

VarSome Clinical can also take your list of custom classifications (i.e. manually curated list of variants) and upload it privately to your account, along with variant comments.

If you have your own, custom pathogenicity classifications for certain variants or wish to add information in the form of a comment, you can send us a file with this information. We will then include this information and display it in any of your analyses where the relevant variants appear. However, the provided file needs to fulfill certain requirements in order to be processed correctly:

- The file provided by the user should be in CSV (comma-separated values) format with content in the following order: chr,position,ref,alt,classification,comment(s). The classification should be one of 'P' (pathogenic), 'LP' (likely pathogenic), 'VUS' (variant of unknown significance), 'LB' (likely benign) or 'B' (benign).
- Multiple comments should be separated by semicolon (;) and the size of each comment is limited to 2000 characters.
- One classification per variant is expected.
- Each CSV file corresponds to one specific user. Multiple files need to be provided for multiple users.
- It is essential that the corresponding reference genome (hg19 or hg38) is specified by the user along with the provided file(s).

This is an example of a valid CSV file:

```
chr5,125562,C,T,P,found in a young adult with melanoma;
chr6,34229972,T,G,B,Common artifact
```



Classification Nomenclature

Standard classifications must be declared with the following names to comply with the nomenclature used by VarSome Clinical. They can be referred to either by the full name or its code:

- B or Benign
- LB or Likely Benign
- VUS or Uncertain Significance
- LP or Likely Pathogenic
- P or Pathogenic
- CA or Common Artefact
- DR or Drug Response
- DA or Disease Association
- RF or Risk Factor
- PR or Protective
- **CNV:** this column is shown when the user runs a CNV analysis with the sample. It will have green ticks if any of the following conditions meet:
 - there is a CNV that overlaps with the position of the current variant
 - there is a CNV in the same gene where the current variant was found (not necessarily in the same position of the variant)
- **Class:** Variants are ordered by our pathogenicity classification:
 - 5 = Pathogenic,
 - 4 = Likely pathogenic,
 - 3 = Uncertain significance,
 - 2 = Likely benign,
 - 1 = Benign.
- **ACMG Rules:** The set of triggered ACMG rules are displayed in clickable bubble icons that include the rule's description and explanation for triggering.
- **HGVS:** HGVS nomenclature for the variant.
- **HGVS Protein:** HGVS nomenclature for the protein sequence change compared to the reference.
- **HGVS Coding:** HGVS nomenclature for the variant.
- **Transcript position:** Variant described on the DNA level in relation to a specific **gene** based on the coding DNA reference sequence.
- **Overlapping Genes:** The name of any gene(s) the variant falls within.
- Inheritance: Mode of inheritance of the gene from the CGD, OMIM, ClinGen Disease Validity, gene2Phenotype, GenCC and Domino databases:
 - AD: autosomal dominant
 - AR: autosomal recessive
 - XL: X-linked
 - BG: blood group
- **Function**: The position of the variant with respect to the gene it falls within, and its coding effect (if any).


5f	5' Flank	nc	non-coding exon	
5u	5' UTR	fs	Frameshift	
3f	3' Flank	ms	Missense	
3u	3' UTR	ns	Nonsense	
С	Coding	sl	Stop Loss	
In	Intronic	sy	Synonymous	
Sp	Splicing	ed	exon deletion	
lf	in frame	nmd nonsense-mediated d		

VarSome Clinical allows the use of custom transcripts for annotation.

In terms of variant function, VarSome Clinical annotates variants against all the transcripts available (Ensembl and Refseq), and shows the results in the Function column when browsing results:

ilysis filtered by Gene li	st fro 🔗 CNV Re	sults of	00	Ţ	Filters 🛃 Report	generatio	n A	nalysi
-		Search for	gene, chr (e.g. chr1	с	ବ 📋 ଟ	• T	- +	:
• O Inheritance	Function	Zygosity	Phasing Grou	p 🔶	Frequency	🖲 Alle	lic Balan	ce
AD	0 🚥	000			0.000004032	0	0.073	0
AD	0 🚥	000			0.000004028	0	0.13	0
AD	C 🚥	000		-	0.000004011	0	0.022	0
AD		000		-	0.0000398	0	0.1	0
AD/AR	3 💿 😳	000		-	0.00003982	0	0.094	0
AD	8 🚥	000		-		0	0.29	0
AD	C 👓 🚥	000		-		-		-
AD/AR	C 🔜 🕞	000		-		0	0.29	0
AD/AR	C and (5	000		-		0	0.43	0
AD/AR	c and fs	000		-	-	0	0.098	0

However, for the Germline Variant Classification, as a part of the annotation of variants, VarSome Clinical considers only one transcript - by default the transcript with:

- 1. the most severe coding impact,
- 2. otherwise the MANE Select transcript
- 3. if the above is not available, the longest canonical transcript
- 4. otherwise, the MANE Plus transcript



- 5. failing that, the longest transcript
- 6. and finally, the RefSeq transcript

For Ensembl transcripts, all must have TSL $[\underline{1}] = 1$ or null.

Germline Variant Classification Version: 11.8		Terms of use Documentation
Sumily Information	Pathogenic 17 points = 17 P · 9 B P Using transcript NM_007294.4, MANE select.	Submit to ClinVar Save as manual classification
Mode of Inheritance	ADI/AR, based on gene information from CGD, ClinGen Disease Validity, GenCC, Mondo, OMIM and 2 more.	
Automated reliaria - Shrw summary view 🔿 - Enable clinical evidence o		

- **Gene Symbol:** Gene used for annotation and classification of the variant for ACMG (& AMP for somatic samples).
- Zygosity:

Homozygous with the alternative allele

- O Homozygous with the reference allele
- Heterozygous (unphased)
- Heterozygous phased 1
- Heterozygous phased 2
- Failed Quality Filters / Non genotyped

If the variant did not pass the variant caller <u>quality filter</u>, , the zygosity is shown in the table as (failed quality/non-genotyped).

Representation of zygosity for large cohort analyzes

The zygosity representation shown below applies for multi-sample analyses larger than 10 samples and brings together information about both the zygosity and the status (affected - unaffected) of the samples at first glance.



In this layout, each type of zygosity found for a variant is shown only once with the respective box. Next to each box reside two numbers: the number in green font corresponds to the unaffected samples having this specific type of zygosity for this variant. The number in red font corresponds to the affected samples having this type of zygosity in the dataset.

Each box is associated with this pair of color-coded numbers which provides information about both the number of occurrences in the dataset and the status of these samples in a compact manner.

The "Nearby variants in this sample" tab is modified, as well to fit more efficiently large cohort analyses of more than 10 samples using the aforementioned layout described for the "Zygosity" column.

Nearby variants in this sample		
Variant	Zygosity	Gene(s)
chr7:117530974 C⇒T	©000©2630©000©100©000	CFTR
chr7:117518106 A⇒G	©000 2630 000 100 000	CFTR
chr7:117509093 G⇒A	@000@2630@000@100@000	CFTR
chr7:117504391 A⇒G	◎000 2630 000 100 000	CFTR
chr7:117504290 C⇒T	@0000\2630@000@100@000	CFTR
chr7:117530975 G⇒A	◎0 000 26300002100€00 0	CFTR
chr7:117535318 A⇒G	◎000 2630 000 200 100 300	CFTR
chr7:117535451 A⇒G	◎ 000 ○ 2230 ● 100 ♀ 400 ♀ 000	CFTR
chr7:117536514_5 insGATT	@0000\2630@000\$100\$000	CFTR
chr7:117536515 delGATT	00000 1520 400 810 000	CFTR
chr7:117536684 C⇒T	◎0 00 ○2630●000②10 0 ②0 00	CFTR

Finally, the tab "Cohort Zygosity", on the variant table, reports exclusively the names of



the samples carrying the variant, also color-coded in red or green font to inform about the sample status and their respective type of zygosity. Note that this tab appears in all multi-sample analyses, even those with fewer than 10 samples.

Zygosity	Sample	Allelic Balance	Coverage
Samples containing this variant			

• **Phasing group**: this column indicates the phasing group of the variant if it exists. Note that this column will be available only for those variants that are phased. If you would like to filter the variant table and visualize only the variants that share the shame

phasing group with your variant of interest, you can click on the following icon _____. If you would like to know more about phasing please go <u>here</u>.

- Allelic balance: Proportion of reads that support the variant. For example, if a variant's location is covered by 100 reads, of which 25 support the variant and 75 do not, then the variant would have an allelic balance of 25/100 = 0.25.
- **Frequency:** Frequency of the variant in the general population or (if applied) the specified ethnicity.
- **Coverage**: Number of reads that align to the variant's position. For analyses of FASTQ samples, the blue numbers are links to JBrowse, showing the read alignments at the variant's position.
- **Filters**: filters that have been applied to the data. Filters are associated with the variant calling quality filters that have been applied to the variant to decide whether it has a <u>call</u> <u>status</u> of PASS or FAIL.
- **rsID**: the reference SNP cluster ID of this variant

Extra columns for *somatic* samples:

- AMP Tier: Variants are ordered by an aggregate AMP score (see AMP Implementation documentation for more details), from most pathogenic to benign. Tier I = Cancer with approved drug therapies, Tier II = Cancer but no approved therapy, Tier III = Uncertain Significance, Tier IV = Benign or not related to cancer.
- **AMP Rules:** The set of triggered AMP rules are displayed in clickable bubble icons that include each rule's description and explanation for triggering
- **Sample Metrics:** Each icon represents the sample information introduced by the user. They light up when there is data in one of the cancer-related databases matching the variant to the relevant sample characteristic:
 - <u>Cancer type</u>: this highlights any variants for which evidence is found linking to the same cancer type as the sample.
 - <u>Tissue</u>: This will highlight any evidence associating the variant or gene to the sample tissue.
- <u>Age</u>: This will display the patient's age relative to an age histogram VarSome Clinical User Manual Version: 11.9.1 - 19th December 2023 P



in reported somatic samples.

- <u>Ethnicity</u>: It will report the variant's frequency in the relevant ethnic group.
- <u>Sex</u>: more than 50% of reported cases across somatic sample databases match the sample's sex.
- <u>Variant Allele Frequency</u>: variants with a low VAF are most likely tumor variants, whilst VAFs of 50% and 100% indicate germline variants.
- **Somatic Samples:** Sum of available affected samples from databases included in Cancer Sample Summary (ICGC Somatic, COSMIC, CBioPortal, Cancer HotSpots, GDC).

Clinical Cards

On the bottom of the variant table the clinical cards display complementary information of the *variant* or the *gene*.

Please note that, when selecting a variant through the check box on the Variant table in VarSome Clinical, the Clinical Cards do not change because the selection of the variant is different from clicking on it to view the variant information. Therefore, to view the clinical cards and the information on a variant, simply click anywhere on the row containing the variant and not the check box to select it for any action.

Variant clinical cards:

Variant Gene	🖲 Variant 🗌 Gene 🌣 🛛										
General Information SNV PIK3CA(NM_006218.4):c.1633G>A (p.Glu545Lys)	Somatic Classification	ICGC Somatic Samples: 354 Type(s): Breast Cancer, NOS, Es	PharmGKB No data available	Germline Classification Pathogenic 27 points = 27 P - 0 B	OMIM ® Related Phenotypes: 1	ClinGen New No data available	Clinical				
Community Contributions Classifications: 2 Comments: 0	JAX CKB AMP: Tier I Evidence: 266 Breast Cancer	Cancer Hotspots Samples: 575 Type: Breast Cancer, Colorectal	DoCM Pathogenic Diseases: 26 Drugs: 0	ClinVar Pathogenic/Likely Path 22 24	IARC TP53 Germline No data available	GWAS No data available	Population Frequencies				
Publications	CIVIC **** Her2-receptor Positive Breast C	GDC Samples: 262 Type: Infiltrating Ductal Carcino	Cosmic Samples: 3 Prediction: Pathog Top: Carcinoma	LOVD Pathogenic New Affects function	Conservation Scores phyloP100: 9.477	Structural Variants 🚝	#Samples Yours: 0 (hom) 7 (het) 197 (som) All: 0 (hom) 7 (het) 197 (som)				
Transcripts NM_006218.4 - missense MANE Select	PMKB Tier I Evidence: 11 Diffuse Large B Cell Lymphoma	cBioPortal Samples: 3330 Type: Breast Cancer, Colorectal	Sample View 🚖 🛛 🔍	Frequencies exomes: f = 0.00000403 (cov: 42.7) genomes: not found (cov: 31.4)	Uniprot Variants Pathogenic	Protein Viewer 🕏					
Region Browser 물	Cancer Samples Summary Samples: 7534 Sources: 5 Tier I	IARC TP53 Somatic No data available	Expression Data Top: cells_ebv_transformed_lym Tissues: 54	MitoMap No data available	Pathogenicity Scores 12 9 4	Nearby Variants (Sample)					

- **General Information:** Information about the genomic location of the variant, its type, cytoband, HGVS notation, sequence.
- **Community Contributions:** VarSome's community public contributions for this variant.
- **Publications:** publications from PubMed related to the selected variant or gene where genes, variants, diseases, phenotypes, chemical compounds, drugs (if exist) are tagged by our internal AI tool.
- **Transcripts:** Chromosomal location, link to UCSC genome browser, dbSNP (rs number), Refseq transcripts containing the variant, HGVS notation, etc. Available transcripts for the selected genes are displayed (information from RefSeq and Ensembl, along with the MANE transcript). The transcript used in classification is highlighted in the Transcripts summary card.



General Information SNV	AMP Classification Tier II	Cancer Samples Summary	GDC	DoCM Pathogenic	ACMG Classification Pat	thogenic IA		Structural Variant	Browser 🕮	Population Frequencies	
Community Contributions	JAX CKB	Cosmic	cBioPortal	Region Browser 🚖	ClinVar Pathogenic	Co	inservation Scores	Protein Viewer 😪		#Samples	
Publications	CIVIC	ICGC Somatic		Expression Data	Frequencies gnomAD Ex	xomes Ur	iprot Variants	Nearby Variants (S	Sample)	OMIM	
Transcripts	РМКВ	Cancer Hotspots	PharmGKB		MitoMap	Pa	thogenicity Scores	Clinical			
kanscripts											
RefSeq Transcripts Version: 210										^	
Transcript	c	Coding impact G	ene	HGVS coding	HGVS Protein	Locatio	n	Protein position	Spl	ice distance	
NM_004333.6 See on NCBI	n	vissense B	RAF	c.1799T>A	V600E (p.Val600Glu)	exon 15 (coding)	of 18 position 58 of 119	600 of 767	58		
NM_001354609.2 See on NCBI	п	alssense B	RAF	c.1799T>A	V600E (p.Val600Glu)	exon 15 (coding)	of 19 position 58 of 119	600 of 768	58		
NM_001374244.1 See on NCBI	п	nissense B	RAF	c.1919T>A	V640E (p.Val640Glu)	exon 16 (coding)	of 19 position 58 of 119	640 of 807	58		
NM_001374258.1 See on NCBI	п	lissense B	RAF	c.1919T>A	V640E (p.Val640Glu)	exon 16 (coding)	of 20 position 58 of 119	640 of 808	58		
✓ Show 9 more											
										- Canonical transcript	
			This functional	annotation is calculated on the fly for al	II variants, known or unknown						
Ensembl Transcripts Version: 107										^	
Transcript	c	Coding impact Gen	e HG ¹	/S coding HC	3VS Protein	Location	Protein p	osition	Splice distance	TSL	
ENST00000288602.6 See on Ensembl	n	vissense BRA	F c.17	97×A V6	sooe (p.Val600Glu)	exon 15 of 18 posit (coding)	ion 58 of 119 600 of 76	7	58	1	
ENST00000479537.1 See on Ensembl		BRA	F			exon 2 of 6 position coding exon)	58 of 119 (non-		58	5	
ENST00000496384.2		income Die		075-A 547	07E (n. 10/2070 n.)	exon 6 of 10 position	n 58 of 119 207 of 27		co		

- **Region Browser:** Genomic region browser, lollipop graph of the pathogenicity of each variant, frequencies from gnomAD and Bravo, variant visualization with filtering according to databases and coding impact.
- **Sample View**: Sample's region browser presents information about SNVs, transcripts, ROH, conservation scores/position (and CNVs) of the sample. For further information please refer to section <u>Sample View</u>.
- **PharmGKB:** Information on the impact of genetic variation on drug response from PharmGKB database.
- **Expression Data:** Tissue-specific variant expression data from the Genotype-Tissue Expression (GTEx) project.
- **GWAS:** Associations of specific genetic variations with particular diseases from the genome-wide association study (GWAS) database.
- **ACMG Classification:** The Germline Variant Classification and its triggered rules for the selected variant.
- ClinVar: Information about clinical associated phenotypes connected to the selected variant. Description/Submitter column displays ClinVar Comments (View ClinVar Comments).
- **Frequencies:** If known, Gnomad frequencies for the selected variant and for any other known variants that overlap with it.
- **MitoMap:** Information from the Human Mitochondrial Genome Database.
- **Conservation Scores:** Conservation scores from different resources.
- UniProt Variants: Variant information from UniProt.
- In-Silico Predictors: : Variant pathogenicity predictions produced by *in silico* predictors. In this clinical card, the pathogenicity is displayed with a different color and there are two bar charts, one for the meta-score (predictor which combines multiple predictors into a single score) and one for the individual predictors.
- **ClinGen:** Curated data from ClinGen variants database. Information will be available when the queried variant has already been curated by any of the ClinGen expert panels.
- **Structural Variants:** Structural variant browser for the detected variants.
- Protein Viewer: The 3D protein viewer tool is available to map variants onto the protein



structure. By clicking on the "Protein Viewer" card, a new window will open with the 3D Protein Viewer tool. For more information please refer to <u>3D Protein Viewer</u>.

- **Nearby Variants:** Variants in the genomic neighborhood of the selected variant. This variant list is not affected by the filters applied to the sample.
- **Clinical:** ClinVar and Cosmic annotations for the selected variant and for any other known variants that overlap with it.
- **Population Frequencies:** gnomAD frequencies and coverage, from Gnomad Exomes and Gnomad Genomes.
- **#Samples:** The number of samples in which a specific variant has been found. This column gets updated daily. The number of homozygotes and heterozygotes in Saphetor for the variants are shown, but only sample IDs of samples analyzed by you and your group are reported.

VarSome Clinical comes with powerful sample cross-referencing, which links all your

samples together on the variant level. It can be found in each tab and it reports two things:

- 1. the number of samples of all VarSome Clinical users in which a specific variant has been found and
- 2. in which other samples of you or your group this variant is present.

This information is updated daily. The number of homozygotes and heterozygotes in VarSome Clinical for the variants are shown, but only sample IDs of samples analyzed by you and your group are reported.

Variant Gene		/	Current Annotation 🔅 😭
General Information Deletion Region Browser Sample View Con	ommunity Contributions Publications Transcripts PharmGKB Germline Classification Pathogenic	Uniprotrariants ClinVar Pathogenic AAAA LOVD Pathogenic	Frequencies MitoMap Conservation Scores
Pathogenicity Scores OMIM ClinGen Expression Data GWAS	AS Structural Variants Protein Viewer Nearby Variants (Sample) Clinical Population Frequencies	s #Samples	



Samples with this variant as of 11 May 2022

In all samples		
Homozygous	Heterozygous	Tumo
0	65	0
In your samples		
Homozygous (0)		
Heterozygous (65)		
tr1 Allelic balance: 0.29		
tr1 Allelic balance: 0.29		
tr1 Allelic balance: 0.29		

- **OMIM:** Information about phenotypes related to the selected variant, as retrieved from Online Mendelian Inheritance in Man[®].
- LOVD: Presents the clinical evidence available from the <u>Leiden Open Variation Database</u> (LOVD) for a given gene and/or variant and it has also been incorporated into the Germline Classification as a new source of clinical evidence.
- **DVD:** Variant annotations from the <u>Deafness Variation Database</u>. DVD provides a comprehensive guide to genetic variation in genes known to be associated with deafness.
- **Multi:** Displays the components and their corresponding files used to run the multi-sample analysis, and if these have been selected as affected or not.
- Audit trail: Shows the record of the actions that have been made from all users of a group on the samples analyzed. Only the group supervisor has access to this information.

Additional cards for somatic samples:

- **AMP Classification**: The AMP tier and the set of triggered rules for the selected variant.
- JAX CKB: Somatic gene variant annotations and related content provided by The Jackson Laboratory Clinical Knowledgebase.
- **CiViC:** Somatic variant annotations retrieved from CiViC.
- **PMKB:** Clinical interpretations of somatic variants retrieved from PMKB.
- Cancer Samples Summary: Aggregated information across different data sources.
- **Cosmic:** Somatic variant annotations from COSMIC database.
- **ICGC Somatic:** Somatic variant annotations from ICGC database.
- **Cancer Hotspots:** Somatic annotations from Cancer Hotspots database.
- **GDC:** Somatic variant annotations from GDC database.



- **cBioPortal:** Summary of samples matched in cBioPortal with the selected somatic variant.
- IARC TP53 Somatic & Germline: Somatic annotations of TP53 gene mutations in human cancers.
- **DoCM:** Information retrieved from DoCM, about known, disease-causing mutations associated with the variant.

Gene clinical cards:

Variant OGene PIK3CA	Variant 🖲 Gene PIK3CA 🗸									
Gene basic Info	Gene function	GenCC Disease: Cloves Syndrome Inheritance: Somatic mosaicism	Human Phenotype Ontology Number of diseases: 21	JAX CKB Role: oncogene	Pharm GKB Drugs: 8	Community Contributions				
Region Browser 🖻	Known gene variants P+LP: 385, VUS: 754, B+LB: 393 Frequency cut-off for rule BS1 : 0.0001	ClinGen Disease Validity	Human Protein Atlas Top: Colorectal Cancer and 5 more (1 Tissues: 76	CIVIC ***** Breast Cancer	FDA Drugs: 1					
Structural Variants 🚝	dbNSFP	DOMINO Probability of AD: 1	Fusion GDB Fusion Genes: 6	PMKB Tier II Evidence: 3 Brain	DGI Drugs: 104					
Transcripts NM_006218.4 ENST00000263967.3	GnomAD Genes Loss of function: 7.209 Missense: 5.599	PanelApp gene lists	Gene Expression	Cancer Gene Census Tier I Somatic: True Germline: False	CPIC No data available					
Publications	EBI Gene2Phenotype Disease: Cloves: Congenital Lipomato Inheritance: monoallelic_autosomal	Clinical Genomic Database Conditions: Cowden syndrome 5 Inheritance: AD	Protein Viewer 🕏	OMIM ® Related Diseases: 14	AACT Clinical Trials Recruiting: 27 Active, not recruiting: 13 Not yet recruiting: 5					

- Gene basic info: Description, synonyms, cytoband, links to clinical or other resources.
- **Transcripts:** Strand, chromosomal location, length, mRNA length, UniProt accession number, etc.
- **Publications:** Publications from PubMed related to the selected gene where genes, variants, diseases, phenotypes, chemical compounds, drugs (if exist) are tagged by our internal AI tool.
- **Gene function:** Functions related to the selected gene, as provided by Genetics Home Reference (GHR).
- Known gene variants: Variants in the selected gene with known pathogenicity.
- **dbSNFP:** Functional prediction and annotation of all potential non-synonymous single-nucleotide variants.
- **GnomAD Genes:** Data summary from a wide variety of large-scale sequencing projects associated with the selected gene.
- **EBI Gene2Phenotype:** Gene association with a disease entity based on an allelic requirement and a mutational consequence.
- **GenCC:** Curated information about the gene-disease relationship.
- NHI ClinGen Disease Validity: Gene-disease association validity information.
- **DOMINO:** Probability of the selected gene to cause dominant changes.
- PanelApp gene lists: Catalog of available gene panels including the selected gene.
- **Clinical Genomic Database:** Age affected, condition, inheritance, indicated intervention categories, publications as retrieved from Clinical Genomic Database.
- **Human Phenotype Ontology:** Disease and their phenotypic abnormalities associated with the selected gene.
- Human Protein Atlas: Protein expression information by cell and tissue type
- **Fusion GDB**: Functional annotation of fusion genes in cancer and their related drugs
- **Gene Expression:** Tissue-specific gene expression data from the Genotype-Tissue Expression (GTEx) project.



- **Protein Viewer:** The 3D protein viewer tool is available to map variants onto the selected protein structure. By clicking on the "Protein Viewer" card, a new window will open with the 3D Protein Viewer tool. For more information please refer to the section <u>3D Protein Viewer</u>.
- JAX CKB: Evidence and clinical trials content related to the selected gene as provided by the section 3D Protein Viewer Jackson Laboratory Clinical Knowledgebase.
- **OMIM:** Information about phenotypes related to the selected gene, as retrieved from the Online Mendelian Inheritance in Man[®].
- **PharmKGB:** Information on the impact of genetic variation on drug response from PharmKGB database.
- **FDA:** Approved drugs associated with the selected gene, from FDA.
- **DGI:** Information about drug-gene interactions interpreted by the "Drug Gene Interaction" Database.
- **CPIC:** CPIC levels to genes/drugs retrieved from "Gene Drugs Interactions and Levels"
- **AACT Clinical Trials:** Information about every clinical study registered in ClinicalTrials.gov associated with the selected gene as provided by AACT.
- **Community Contributions:** VarSome's community public contributions for this variant.

Additional cards for somatic samples:

- **CiViC:** Cancer-related clinical evidence as derived from CiViC database.
- **PMKB:** Clinical interpretations of gene variants retrieved from PMKB.
- **Cancer Gene Census:** Information about gene's mutations that are causally implicated in cancer as retrieved by COSMIC database.

A Please note that grayed out tabs are disabled if no related information is available.

Current Annotation:

VarSome provides "Current Annotation" of **germline and somatic** variants using the latest data and the latest germline or somatic classifier. This feature allows you to see whether additional data are available for a variant, and whether the automated classification is altered by new evidence or refinements to the classifier. *This feature is provided for research purposes only in VarSome Clinical*.

Original view:

f	or germline vo	ariants:					
	Variant Gene						Current Annotation 💠 😢
	General Information SNV BRAF(NM_004333.6):c.1799T>A (p.Val600Glu)	PharmGKB Drugs: 2	Germline Classification Pathogenic 27 points = 27 P - 0 B	Frequencies exomes: f = 0.00000398 (cov: 82.5) genomes: not found (cov: 30.7)	OMIM ® Related Phenotypes: 1	Protein Viewer 오	#Samples Yours: 0 (hom) 9 (het) 204 (som) All: 0 (hom) 9 (het) 204 (som)
	Region Browser 코	Publications	Uniprot Variants Pathogenic	MitoMap No data available	ClinGen (New) No data available	Nearby Variants (Sample)	
	Sample View 🚖 🛛 👫	Transcripts NM_004333.6 - missense MANE Select	ClinVar Pathogenic ★★☆☆ 20 12 1 2	Conservation Scores phyloP100: 9.201	GWAS No data available	Clinical	
	Community Contributions Classifications: 7 Comments: 0	Expression Data Top: testis Tissues: 54	LOVD Pathogenic (New) Affects function	Pathogenicity Scores 10 8 4	Structural Variants 🔚	Population Frequencies	



for somatic variants:

Variant ○ Gene									
General Information SNV TMPRSS2(NM_005656.4):c.478G>A p.(Val160Met)	Somatic Classification	ICGC Somatic Samples: 5 Type(s): Bowel Cancer, NOS, Lun	PharmGKB	Germline Classification Benign -16 points = 0 P - 16 B	OMIM ® No data available	In-Silico Predictors BP4: Benign 1 9 11 22	Nearby Variants (Sample)		
Community Contributions	JAX CKB No data available	Cancer Hotspots No data available	DoCM No data available	ClinVar Benign ★☆☆☆ 1	Deafness Variation Database (New) No data available	ClinGen No data available	Clinical		
Publications	CIVIC No data available	GDC No data available	Cosmic Samples: Prediction: Pathoge Top: Rhabdomyosarcoma	LOVD No data available	IARC TP53 Germline No data available	GWAS No data available	Population Frequencies		
Transcripts NM_005656.4 - missense MANE Select	PMKB No data available	CBioPortal Samples: 8 Type: Peripheral Nervous Syste	Sample View 🚖	Frequencies exomes: f = 0.245 (cov: 70.6) genomes: f = 0.28 (cov: 30.9)	Conservation Scores phyloP100: 1.573	Structural Variants 🚝	#Samples Yours: 0 (hom) 0 (het) 0 (som) All: 0 (hom) 0 (het) 0 (som)		
Region Browser 🖻	Cancer Samples Summary Samples: 13 Sources: 2 Tier III	IARC TP53 Somatic No data available	Expression Data Top: stomach Tissues: 21	MitoMap No data available	Uniprot Variants Benign	Protein Viewer 🕏			

Current annotation view:

for germline variants:

Variant Gene		() Cur	rent annotation data. <u>Disclaimer</u>			Current Annotation 💠 😢
General Information SNV BRAF(NM_004333.6):c.1799T>A (p.Val600Glu)	PharmGKB Drugs: 2	Germline Classification Pathogenic 27 points = 27 P - 0 B	Frequencies exomes: f = 0.00000398 (cov: 82.5) genomes: not found (cov: 30.6)	OMIM ® Related Phenotypes: 1	Protein Viewer 😂	#Samples Yours: 0 (hom) 9 (het) 204 (som) All: 0 (hom) 9 (het) 204 (som)
Region Browser ≘	Publications	Uniprot Variants Pathogenic	MitoMap No data available	ClinGen (New) No data available	Nearby Variants (Sample)	
Sample View 🚖 🛛 🔍	Transcripts NM_004333.6 - <mark>missense</mark> MANE Select	ClinVar Pathogenic ★★☆☆ 20 12 1 2	Conservation Scores phyloP100: 9.201	GWAS No data available	Clinical No data available	
Community Contributions Classifications: 7 Comments: 0	Expression Data. Top: testis Tissues: 54	LOVD Pathogenic New Affects function	Pathogenicity Scores 9 8 4	Structural Variants \Xi	Population Frequencies No data available	

for somatic variants:

Variant Gene			Urrent annotation da	ata. <u>Disclaimer</u>			Current Annotation 🔅 🤨
General Information SNV TMPRSS2(NM_005656.4):c.478G p.(Val160Met)	Somatic Classification	ICGC Somatic Samples: 5 Type(s): Bowel Cancer, NOS, Lun	PharmGKB	Germline Classification Benign -16 points = 0 P - 16 B	OMIM ® No data available	In-Silico Predictors BP4: Benign 9 11 22	Nearby Variants (Sample)
Community Contributions	JAX CKB No data available	Cancer Hotspots No data available	DoCM No data available	ClinVar Benign ★☆☆☆ 1	Deafness Variation Database (New No data available	ClinGen No data available	Clinical No data available
Publications	CIVIC No data available	GDC No data available	Cosmic Samples: Prediction: Pathoge Top: Rhabdomyosarcoma	LOVD No data available	IARC TP53 Germline No data available	GWAS No data available	Population Frequencies No data available
Transcripts NM_005656.4 - missense MANE Select	PMKB No data available	cBioPortal Samples: 8 Type: Peripheral Nervous Syste	Sample View 🚖	Frequencies exomes: f = 0.245 (cov: 70.6) genomes: f = 0.28 (cov: 30.9)	Conservation Scores phyloP100: 1.573	Structural Variants 🚝	#Samples Yours: 0 (hom) 0 (het) 0 (som) All: 0 (hom) 0 (het) 0 (som)
Region Browser 🚖	Cancer Samples Summary Samples: 13 Sources: 2 Tier III	IARC TP53 Somatic No data available	Expression Data Top: stomach Tissues: 21	MitoMap No data available	Uniprot Variants Benign	Protein Viewer 😒	

A Please note that the **current annotation view** is now available for germline and somatic analyses. The variant link to VarSome has been removed from these types of analyses.

Compact View:

It is possible to make the clinical cards much smaller by clicking on the "Display Options" wheel on the top right hand side of the cards panel, and selecting "Compact View". This removes the summary data from the cards and reduces their size.

	Vars	someclin	ical		
)763	459			
)984	Compact V	√iew		
	7 8 9	Apply			
	#Samples	•	• •		
● Variant Gene General Information SNV Community Contributions Publications Transcripts Region Browser DxCM Germine Classification Pathogenic Clinival Pathogenic 4+4+1 LOVD Pathogenic Nearby Variants (Sample) Clinical Population Prequencies #Samples Cosmic	Sample View Somatic Classification 1 Frequencies MitoMap OMIM 8 IA	Ter II JAX CKB CIVIC PHIKB C RC TP53 Gemiline Conservation Scores	ancer Samples Summary ICGC Soma Uniprot Variants Pathogenicity Scc	Cancer Hotspots GDC cBioPortal ress ClinGen Expression Data CWAS	ARC TPS3 Somatic PharmGKB Structural Variants Protein Viewer
Some useful icons:					
• Click on this box	to get acce	ess to your sa	wed filter se	ts	
• Click on the Report ge	neration icor	🛓 Report o	jeneration	o see the list	of variants

You can download a Report of the selected variants in PDF format. To do so, click on the

above icon, then select	Generate Report	and you will be directed to the following screen.
fastq.germline.twistcore.trusight.hg19.5030706306 clinical report (hg19)		Dashboard Samples Analysis clinical report
Clinical report		Lowedond as PC: Reset

	Varsomed	nica
306	Predicted sex:	
	Diseases	
		•
		∎ ⊕ ∎ ⊕

By clicking on the 🚨 icon, a Report widgets menu will be shown in order to customize the report.

selected for export.



Accumulated variant information

All variants' information

Specific Variant Information

NM_004656.4:c.659+192_659+193delCT

Variant information

Variant references

BAP1 Gene information

BAP1 PMKB info

You can drag and drop the information you prefer to include in the report. At the last option you can find PMKB information for the gene that includes the variant. See more in section 15 <u>Final</u> <u>Report Generation</u>.

• The options provided here are the same as described in <u>"Analysis actions"</u> options.



Green menu

- Search: You can search through your results, by querying according to the VarSome search format. The query can include any of the following:
 - gene: e.g. PIK3CA,
 - chromosome: e.g. chr3 or 3
 - o chromosome position: e.g. chr3:178947865, chr3-178947865, chr3 178947865 or 3



178947865.

- genomic range: e.g. chr3:178936091:178942431, chr3:178936091..178942431, chr3-178936091-178942431, chr3 178936091 17894243 or 3 178936091178942431.
- variant (DNA): e.g. chr3:178936091 G⇒A, chr3:178936091-G-A, chr3-178936091-G-A, chr3 178936091 G A, 3:178936091 G⇒A, 3:178936091-G-A...
- variant (HGVS): e.g. NM_004448.4:c.1947-3C>A
- variant (protein): e.g. BRAF:V600E or BRAF V600E.
- rsIDs ("rs" followed by a number)
- COSMIC IDs

This will filter the table and show only the results for that query.

- **Clear search:** This will empty the search box and show all variants again.
- **Reset variant list to original order:** Clicking on this icon resets the sorting order of the columns to the default (the variants will be ordered by **Class**).
- **Multiple sorting:** The list of variants can be sorted by multiple columns. A pop-up window will appear and multiple columns of interest can be selected in order to sort the variants in ascending or descending order.

▲ Please note that multiple column sorting will return informative results as long as the **first column,** which is selected to sort the variants, has **numeric** values (Frequency, number of samples, Phenotypes etc). For example, the user should *not* sort first by "AMP Tier" or "ACMG Class" and then sort by other values like allelic balance, frequency, etc. However "ACMG Class" and "AMP Tier" can be used as second or later in order of columns to sort by.

- **Display variants matching classification:** Filters for custom variant classifications.
- Add or edit your variant classifications: Open the Custom Tag creation menu. Custom tags allow you to classify variants using user-defined tags.
- **Columns**: Remove or add columns to the table. This functionality can be used to remove columns that are not relevant for the analysis.
- Download all filtered variants from the table below (max 50000) in Excel format: Download the list of variants (max. 50000) that pass any currently applied filters in Excel format. The Excel file also contains information about the filters used to obtain the exported table.

Blue menu

•	var	som	neclini	ca
Open resu a linked wir	lts in ndow att	VCF tributes	Commer	Select for nts export
9	• •	V CP 	= 🔩	🧈 🛷 🛛
	Classify variants	Trans	scripts V	f /iew in arSome

Open results in a linked window: This functionality allows you to utilize multiple screens by generating "linked" sub windows that contain the results of an analysis.

Clicking on the **i**con will split out the variant detail data into a separate window which you can reposition on a second monitor.

		varsomeclinica	Analyses	Upload / view fi	les M	anage 🚽 Launch	analysis - Filter sets Gene lists Help		
,	~		(h	ig19) 🕲		CNV Results	of OD		
	9		- e						
	□	Variant 🔶	Variant Type	Gene Symbol	0	Somatic Tier 🛛 🝦	Somatic Rules	O Sample Metrics	
		chr3:178947865 G⇒A 🌏	SNV	PIK3CA 🌏	••	Tierl	Crtd Drug Germ Some Pubs Path Freq Pred	\$ • 0 ≗ এ • ∿	
		chr7:140453136 A⇒T 🌏	SNV	BRAF 🎭	••	Tierl	Crtd Drug Germ Some Pubs Path Freq Pred	\$ 1. 이 이 가	

A pop-up box will inform you of the functionality and by clicking on the "Open New" button, you will see a new window and the result data will be split between the two windows:



~		(Ի	ng19) 🕑		CNV Results o	o. <mark>0</mark> 0			
a =	0- ⊡ ≓ 9	- e							
□ v	′ariant 🔶	Variant Type	Gene Symbol	0	Somatic Tier 🛛 🝦	Somatic Rules			Sample Metr
Ch ch	ır3:178947865 G⇒A 🌏	SNV	PIK3CA 🌏	••	Tierl	Crtdl Drug Germ Soma F	Pubs Path Freq Pr	red II	\$ <u>5</u> ≜ ⊕ • •
C ch	ır7:140453136 A⇒T 🌏	SNV	BRAF 🌏	•••	Tierl	Crtd Drug Germ Soma F	ubs Path II Freq Pr	red II	💠 占 🚨 💿 o -
C ch	ır3:178936091 G⇒A 🌏	SNV	PIK3CA 🎭	•••	Tierl	Crtd1 Drug1 Germ1 Soma1 F	Pubs Path Freq Pr	red II	🕈 占 🚨 😑 o -
C ch	ır3:178952085 A⇒G 🌏	SNV	PIK3CA 🎭	••	Tierl	Crtdl Drug Germ Soma F	Pubs Path Freq		🕈 占 🚨 😑 o -
C ch	ır10:43615641 A⇒C	SNV	RET	•	Tier I	Drug] Germ II Soma II Path II	Freq Pred		¢ <u>b</u> ≗ ⊜ o -
			VarSome C	linical	- Google Chron	ne			o x°
ss133-clinical.v	varsome.com/analy	sis-data/analy	sis/43565/results	s/separa	ate-window/				≊
	Ana	alvsis ID:	chr	3:1789	947865 G⇒A				×
		.,							p ·
icer Samples Summ Nar Pathogenic * prot Variants In-S iical Population F riant	ary ICGC Somatic ★★☆ LOVD Likely I Silico Predictors PP3: Pc Frequencies #Sample:	Cancer Hotspots Pathogenic Fre thogenic Strong s Cosmic	GDC cBioPorta equencies MitoMap ClinGen Pathoge	al IARC OMIN enic E	CTP53 Somatic Ph.	ton Database Arror Variants As Structural Variants	nline Classification 3 Germline Protein Viewer	Pathogenic ervation Scores Nearby Variant	s (Sample)
romosome chr3 1; UCSC genome br Mastermind	Position REF S 78947865 owser	equence A G	ALT Sequence A	Variant t SNV	ype Cytoband 3q26.32	HGVS PIK3CA(NM_006218 p.(Gly914,	.4):c.2740G>A Arg)	RS II rs587776932) dbSNP

Note: Only one new window can be open simultaneously with the main analysis tab. If you are inspecting more than one analysis in different tabs of your browser, the "dual-monitor" window will sync with the tab of the analysis that is active each time.

You can close an existing pop-up window either by hitting the "Close Linked" button in the main screen, or by just hitting "X" on the window. Also, you can click on the "Create New" button to refresh the existing window with the synced data from the variant/ analysis you are inspecting.



- **Classify variants:** add your own classification to a variant.
- VCF attributes: pop-up window describing the quality details for each software tool used to identify the variants. For more information see the section Explanation of <u>VCF attributes</u>.
- Transcripts: pop-up window with all the RefSeq transcripts containing the variant.
 It also shows the location of the variant (intron/exon, amino acid position), its HGVS notation, and genomic function (intronic, exonic, splicing, UTR ...). Canonical transcripts are shown in red.
- **Comments**: It is possible to attach a **short** comment to a selected variant (long comments will not be added and will return an error message). These comments will be linked to the variant or the gene and will be displayed in other analyses if the same

variant is found. Variants with comments will have an sicon in the Variant or Gene column. Comments are shared only within your group, unless you decide to make your comments public by selecting the "Share comment outside your group" option. You can also select the "The comment is specific to this sample only" option and the comment will be available only to this specific sample analysis. If however the variant is present in other analyses, the sample specific comment will not be shown.





Write your comment on the variant. By default comments you make can only be viewed by you and people in your department unless you select to share the comment in which sample anyone will be able to view it
 ✓ □Share comment outside your group
 By default only users in your group can see your comments
The comment is specific to this sample only
Specific to sample comments appear only in analyses related to the sample and n

×

- View in VarSome: link to our free knowledge base and database aggregator, VarSome.
- Select for export: Clicking on this box selects the variant for export, and information about the variant can be exported in Word and Excel format ("Export variant list" box)



• Gene coverage: a pop-up window showing the average coverage for the selected gene and its different transcripts. Clicking on the nodes will expand or collapse the tree.



Gene coverage for Chromosome: chr1, Position: 218519929



Also, by clicking on one of the Exons, a new tab will open with a JBrowse (jbrowse.org) window showing the alignment details from the analysis' bam files. JBrowse is a software tool installed on our secured servers.

browser		10 000 000				100 000 000		100 000 000	101.010.000 014.0		
	20,000,000	40,000,000	60,000,000	000,000,08	100,000,000 Q Q Q Q	120,000,000	140,000,000 07324_218607556 (233.b)	180,000,000	190,000,000 200,0	10,000 220,0 0 ,000	240,000,000
arence sequent	218,607,350		218,607,375	218,007,400	218,607,4	25 2	18,007,450	218,607,475	218,607,500	218,007,525	218,607.5
ig Sequences	5										
877_TruSight	rCardio BAM Alignmen	115	_				NM_00113	35599 2			
						×					

- **Read Alignment on JBrowse**: Opens a new tab with a JBrowse representation of the BAM files
- **Read Alignment on IGV**: Opens a new tab with an IGV representation of the BAM files



7.2 Description of CNV analyses' functionalities

CNV variant table

CNV/SV analysis variant table contains the following information:

8 0 -	- • • • • •	>										Search for gene, chr (e	.g. chr1 Q 📋	ଟ ⊺ -	+ ≡- ±
🗆 Varia	ant 🔶	Length 🕴	Call Quality	Copy Number	Туре 🔅	Genes 🕴	0	Quality Score	ACMG Class	CNV Rules	Number of genes	Number of exons 🔶	Reads expected	Reads observed	Read ratio
ChrY:2	264000110110000	7 469 999	× × ×	0	deletion	AGPAT5P1,AM		9750	Pathogenic	Gene Overlap	143	704	403 042	44 682	0.111
ChrX:	61680001152260000	90 579 999	× × ×	2.47508	duplication	AARSD1P1,A		9060	Pathogenic	Gene Overlap	1 370	9 233	11 576 855	16 040 111	1.39
ChrX:	270000152110000	49 409 999	× × ×	2.46467	duplication	ACAA2P1,AC		4860	Pathogenic	Gene Overlap	757	6 164	6 334 218	8 756 060	1.38
ChrY:	1387000128820000	14 949 999	× × ×	0	deletion	ACTG1P11,AC		15700	Pathogenic	Gene Literature	345	1 891	496 503	4 721	0.00951
ChrX:	5250000158560000	6 059 999	× × ×	2.47508	duplication	ACTG1P10,AL		609	Pathogenic	Gene Overlap	140	1 150	741 158	1 028 886	1.39
ChrX:	152450001154700000	2 249 999	× × ×	2.44361	duplication	ABCD1,ARHG		217	Pathogenic	Gene Overlap	136	1 964	236 389	322 035	1.36
Chr2:8	8916000189440000	279 999	111	2.73985	duplication	ENSG000023		60.1	Likely Pathogenic	Gene	27	45	13 715	22 903	1.67
Chr22	2:2255000123240000	689 999	111	1.38957	deletion	ASH2LP1,ASH		86	Uncertain Significance	Gene Literature	97	253	88 618	58 011	0.655
chr9:3	3924000139910000	669 999	× × ×	2.42223	duplication	CNTNAP3,EN		56.8	Uncertain Significance	-	17	54	17 679	23 771	1.34
Chr9:4	4089000141450000	559 999	× × ×	2.64155	duplication	ATP5F1AP1,E		61.6	Uncertain Significance	-	10	33	10 292	16 015	1.56
•															Þ
Showing 1 to 1	10 of 125 rows 10 -	rows per pa	ige									< 1 :	2 3 4 5 6	7 8 9 10	13 >

- Length: the length in bp of the region considered as a structural variation.
- Call Quality: Three quality control metrics collected for CNV analyses starting from FASTQ data (e.g. WES or gene panel). Each CNV call is assigned ticks, green, and "X"s, red, so at a glance you can see which has passed and failed the quality checks. The first and second metrics will have a grey dash for CNV results of analyses starting either from VCF or WGS data. From left to right, these are:
 - Test sample coverage: this quality control metric ensures a minimum coverage of the test sample at the CNV call region when calling duplications. Green: duplications with a coverage at least equal or higher than the minimum coverage threshold¹ (¹/₁ Please note that all deletions pass this filter and therefore will always have a green color). Red: duplications with a coverage lower than the minimum coverage threshold.
 - Number of reference samples: this is to ensure that a minimum number of samples from the reference set have a minimum coverage¹ in the CNV call region. Green: the reference sample set has at least two samples with coverage higher than the minimum coverage threshold, in the CNV region. Red: fewer than two reference samples with sufficient coverage in the CNV call region.¹
 - CNV call overlapping camouflaged region: this is to check whether an overlap exists between the region of the CNV and the Camouflaged Regions. Camouflaged regions contain duplicated genomic sequences where confidently aligning short reads to a unique location is not possible. Green: no overlap. Red: overlap with a Camouflaged Region (Ebbert et al., 2019)



□	Variant 🔶	Length 🔶	Call Quality	Copy Number 🛛 🍦	Туре 🔶	Genes 🔶	0	Ouality Score
	chr16:5323421258311	725.969	x	3	duplication	ANTKMT,C1QTNF8,		15
	chr1:8178601366223	548.363	🗸	3	duplication	ACAP3,AGRN,B3GA		6
	chr19:9849161281711	296.795	x	3	duplication	ABCA7,ARHGAP45,		4
	chr20:6327391063573784	299.874	🗸	3	duplication	ARFGAP1,CHRNA4,		3
	chr1:24035622650428	246.866	🗸	3	duplication	ENSG00000224387,		7
	chr22:5015678750342733	185.946	🗸	3	duplication	DENND6B,ENSG00		3

¹ Minimum coverage threshold (number of reads): the lowest value between 10 or sample median coverage/10.

- User CNV classification: custom classification for CNV variants for ACMG and AMP rules. For user-submitted VCFs with CNVs, only variants with a copy number value can be manually classified.
- **Copy Number:** estimated copy number of the CNV call calculated from the reads expected vs reads observed ratio assuming a diploid state.
- **Type**: type of CNV, can be either deletion duplication
- **Genes:** genes overlapping the CNV call region.
- Number of genes: number of genes overlapping the CNV call region.
- Quality Score: A measure of statistical support for each CNV call. Specifically, it is the log10 of the likelihood ratio of data for the CNV call divided by the null (normal copy number). The higher the Quality Score the more confident one can be about the presence of a CNV. While it is difficult to give an ideal threshold, and for short calls the scores may be unconvincing, the most obvious large calls should be easily flagged by ranking them according to this score.
 - For WGS CNV analyses, the quality score is given by delly, if it is a single sample, or by ExomeDepth for multiple samples .
 - For CNVs from VCF, provided for annotation only, the Quality Score displays the QUAL value from the VCF (if included).

□	Variant 🔶	Length 🍦	Call Quality	Copy Number 🍦	Туре 🔶	Genes 🔶	0	Quality Score	Germline Class
	chr16:5323421258311	725.969	· · · · ×	3	duplication	ANTKMT,C1QTNF8,	•	15	Pathogenic
	chr1:8178601366223	548.363	🗸	3	duplication	ACAP3,AGRN,B3GA	•	б	Pathogenic
	chr19:9849161281711	296.795	x	3	duplication	ABCA7,ARHGAP45,		4	Pathogenic
	chr20:6327391063573784	299.874	🖌	3	duplication	ARFGAP1,CHRNA4,	•	3	Pathogenic
	chr1:24035622650428	246.866	🗸	3	duplication	ENSG0000224387,		7	Pathogenic

- ACMG CNV class and CNV rules: the ACMG CNV classification and the set of triggered ACMG rules. These rules are displayed in clickable bubble icons that include the rule's description and explanation for triggering.
- Number of exons: number of exons overlapping the CNV region.
- **Reads expected, reads observed and reads ratio:** these columns contain the values for VarSome Clinical User Manual Version: 11.9.1 19th December 2023 Page 165 of 254



each CNV call of the reads expected, the reads observed, and the read ratio The column read ratio is calculated by dividing the number of observed reads by the number of expected reads. Since the number of expected reads is calculated from the reference set of samples, we highlight the importance of having an appropriate reference set, as we mentioned in the QC report section. Given a good reference set of samples with a good correlation between them, and all three call quality metrics passing the filters, the read ratio value can be used to rank the variants according to the strength of the signal.

CNV Rules	Number of genes 🍦	Number of exons 🛛 🍦	Reads expected 🔶	Reads observed	Read ratio 🛛 🍦	🕒 Frequ
Gene Literature Overlap	66	576	0	0		-
Gene Literature Overlap	61	470	0	0		-
Gene Literature Overlap	45	561	0	0		-
Gene Literature Overlap	80	1.003	0	0	-	-
Gene Literature Overlap	61	884	0	0		-
Gene Literature Overlap	63	742	0	0		-
Gene Overlap Literature	2.533	24.516	0	0		-

- **Frequency:** frequency of overlapping CNVs in the same genomic region. The gnomAD database is used to get the general population frequencies for a given structural variant. Depending on the type of variant, the frequencies are calculated as follows:
 - Deletions: we use gnomAD variants if they fully overlap with the given variant.
 - Duplications in coding regions: we compare at the gene level and we use those gnomAD variants that encompass the same coding genes as the given variant.
 - Duplications in non-coding regions: we use gnomAD variants if they are at least covering 85% of the variant region.
- **Cytoband:** The cytoband of each CNV is displayed. In case of long CNVs spanning more than one cytoband, then they are displayed as a range.

CNV tabs

• **Genes:** the gene information for all the genes overlapping the CNV region is available at the right side of the window under the "Gene" option.

	Variant	Gene	
Variant 🜔 Gene			¢ 0
Gene basic Info	dbNSFP	Clinical Genomic Database	Known gene variants
			NIH ClinCon

- **CNV Details**: Summary information about the selected variant (position, type, overlapping genes etc)
- Sample View: Sample's region browser which presents information about the



overlapping transcripts in the CNV region, conservation scores per position and SNVs of the sample. For further information please refer to section Sample View.

• **Transcripts**: A list of all the affected transcripts that overlap with each CNV is displayed on the bottom of the Variant Table, under the "Transcripts" tab. Transcripts can be filtered based on coding status and/or gene name.

Transcripts			^
FILTER RESULTS All \vee Select a gene \vee			
Transcript	Gene	Overlap	Function
NM_001143.1	AMELY	whole transcript	Coding
ENST00000421178.1	FAM197Y1	whole transcript	Coding
ENST00000450145.1	FAM197Y5	whole transcript	Coding
NM_032973.2	PCDH11Y	whole transcript	Coding
NM_001008.4	RPS4Y1	whole transcript	Coding
NM_003140.3	SRY	whole transcript	Coding
NM_033284.2	TBL1Y	whole transcript	Coding
NM_139214.3	TGIF2LY	whole transcript	Coding
NM_003308.4	TSPY1	whole transcript	Coding
ENST00000320701.4	TSPY2	whole transcript	Coding
SHOWING 1 - 10 OF 17 PAGE SIZE:			

- **CNV Classification:** In this tab we show the ACMG classification for each CNV and the set of triggered ACMG rules. Click on "Show full detail" to find out the criteria not met.
- **Publications**: publications from PubMed related to the selected CNV or genes where variants, diseases, phenotypes, chemical compounds, drugs (if exist) are tagged by our internal AI tool as well as from the VarSome community users and our curation team verifies them .
- CNV Browser: An interactive browser showing a wider region around the position of the CNV call as well as its location on the chromosome level. The user can zoom in and out using the mouse scroll and select among different chromosomes, genomic positions, samples and CNV calls. Data points represent read ratios (observed/expected read counts). These are colored blue or red, depending if they fall within the gray shaded area 95% confidence interval or not, respectively. Call genomic location is indicated by coordinates and annotated for overlapping gene structures (exons/introns). The coverage track, at the bottom of the interactive plot, shows the trend of the coverage on a logarithmic or linear scale across all cohort samples. By hovering the browser, there is useful CNV call information including genomic location and span, as well as links to the same region in other analyses of the same cohort. You can find further information in the <u>CNV visualization</u> article.



	varsomeclinical	Analyses	Upload / view	files Manage	- Launch ana	alysis - Filte	r sets	Gene lists Help						Та	gs About	-
~	_WGS_	- Cl	VV Results	s of	WGS	(hş	(19) CN	VV analysis from I	CNV Results	s of	CNV Results of		- +1 more D	🍸 Filters 🕹 Repor	t generation Analy.	sis actions 👻
	9 0+ e e 🦘 <	>										Search fo	r gene, chr (e.g. chr1 c	৭ 📋 😂	• T- +	∎• <u>≵</u>
	Variant 🔶	Length 🝦	Call Quality	Copy Number	Туре 💠	Genes	θ	Quality Score	Germline Class	CNV Rules	Number of	genes 🖕	Number of exons 👋	Reads expected	Reads observed	Read
	chrY:26000019900000	7.299.999	× x x	0	deletion	AGPAT5P1,A	••	2090	Pathogenic	Gene Literature On		132	691	58.209	1.363	0.0234
0	chrX:61750001154700000	92.949.999	× x x	2.61353	duplication	AARSD1P1,		2390	Pathogenic	Gene Overlap		1.510	11.215	1.418.283	2.166.554	1.53
0	chrX:270000152150000	49.449.999	× x x	2.61353	duplication	ACAA2P1,A		1280	Pathogenic	Gene Overlap		759	6.173	776.078	1.187.039	1.53
	chrY:1390000128850000	14.949.999	× x x	0	deletion	ACTG1P11,A		3460	Pathogenic	Gene Literature		345	1.891	86.190	1.092	0.0127
	chr14:105200001.105750000	5.49 999	× × ×	2.01000	duplication	ADAM6 ATP		133	Pethogenic	Gene Overlap		96	1.150	5.022	5.878	1.35
0	chr22:2240000123250000	849.999		1.07587	deletion	ABHD17AP5		46.7	Uncertain Significance	Gene Literature		119	284	16.340	8.617	0.527
0	chrY:1315000113750000	599.999	× × ×	1.42868	deletion	ACTR3BP1,		55.3	Uncertain Significance)			13	21	21.429	14.432	0.673
	chr10:5120000151400000	199.999	~ × ×	3.09085	duplication	ENSG00000		8.91	Uncertain Significance			13	71	207	440	2.13
0	chr14:106200001106350000	149.999	× × ×	0	deletion	ATP6V1G1P	•	20.6	Uncertain Significance	Gene Literature		28	51	2.300	423	0.184 -
Show	ing 1 to 10 of 31 rows 10 • •	rows per page							t							¢ 0
Varian CNV Det CNV Brow	t Gene ails Transcripts CNV Classi	fication Patho	genic Publicat	tions Known CN	Vs Sample Vie	ew CNV Plot	CNV E	Browser	\$				(int) 600,000 to	000.000		•
Varian CNV Det CNV Brow hg19 strand: 1	t Gene ails Transcripts CNV Classi vser	fication Patho	genic Publicat	tions Known CN	IVs Sample Vie	ew CNV Plot	CNV E	Browser	\$				y:2,600,0019	,900,000) - +) (◆ ↓ ↓
Varian CNV Det CNV Brow hg19 strand: 1	t Gene ails Transcripts CNV Classi vser	fication Patho	genic Publicat	tions Known CN	Vs Sample Vie	w CNV Plot	CNV E	Browser	\$				y:2,600,0019 show all chrY: 10.	.900,000 Samples	-+(> E3 × Save
Varian CNV Det CNV Brow hg19 strand: 1	t Gene aila Transcripts CNV Classi vser	S,7 M	genic Publicat	tions Known CN	IVs Sample Vie	w CNV Plot	CNV E	Browser	Э 7,7 М				y:2,600,0019 show all chrY: 10. 9,7 M	900,000 Samples You can press on to previe	d MOLD the eye kon @ w other samples	> [] × Save
Varian CNV Det CNV Brow hg19 strand: 1 CNV ca	t Gene alis Transcripts CNV Classi vser t t t t t t t t t t t t t t t t t t t	S,7 M	genic Publicat	tions Known CN	Vs Sample Vie	7 M	CNV E	Browser	\$ 7,7 M				y:2,600,001.9 dow all chrY: 10. 0.7 M	900,000 Samples You can press an to previe	HOLD the eye icon Θ w other samples	
Varian CNV Det CNV Brov hg19 strand: 1 CNV ca Read R	t Gene alis Transcripts CNV Classi ssee tsee ts	S.7 M	genic Publicat	tions Known CN	Vs Sample Vie	zw CNV Plot	CNV E	Browser	3 7.7 M				y:2,600,0019 abow all chrY: 10. 9.7 M	900,000 Samples You can press an to previe wors	-+ (d HOLD the eye icon @ w other samples	
Varian CNV Det CNV Det CNV Brow hg19 trand: 1 CNV ca Read R 14 12	t Gene alis Transcripts CNV Classi rsee c c c c c c c c c c c c c c c c c c	ication Patho	genic Publicat	Bions Known CN	Vs Sample Vir	sw CNV Plot	CNV E	Browser	3 7,7 M				(y:2,600,0019 dow all chrY: 10. 0.7 M	900,000 Samples You can press on to previe wcs_ 	HOLD the eye con @ d HOLD the eye con @ v other samples [EEF]	
Varian CNV Det CNV Brov bg19 strand:1 + CNV ca + Read R L4 L2 1 .8	t Gene alis Transcripts CNV Classi rsee ts	3.7 M	genic Publicat	tions Known CN	Vs Sample Vi	7 M	CNV E 7,3 Mb	Browser	7.7 M				y:2,600,001.9 dow all chrY: 10. 9.7 M	900,000 Samples You can press on to press wors wors wors	(HOLD the ope icon () e other samples	
Varian CNV Det CNV Brow hg19 strand: 1	t Gene alis Transcripts CNV Classi vser t t t t t t t t t t t t t t t t t t t	a,7 M	genic Publicat	ions Known CN	V/s Sample Vir	7 M	CNV E	Browser	3 7.7 М				y2,600,001.9 dow all chrY:10.	900,000 Samples You can press on to preve wcs UKCS	(1000 the ope lease or other samples	
Varian CNV Det CNV Brow hg19 trand: 1 t CNV ca t Read R 14 1 1 .	t Gene alis Transcripts CNV Classi reer tree ts	3.7 M	genic Publicat	tions Known CN	Vs Sample Ve	rw CNV Piot	CNV E	Browser	3				yz,600,001.9	900,000 Samples You can prest on to prest wes UNCS UNCS	I HOLD Uhr ayr son 🛛 ir HOLD Uhr ayr son 🖓 ir other samples [RE7] (
Varian Varian CNV Det CNV Det CNV Brov Find Find Find Find CNV ca Find CNV ca Find CNV ca CNV	t Gene alis Transcripts CAV Classi recer tss control of the second secon	a.7 M	genic Publicat	tions Known CN	Vs Sample Vi	rw CNV Plot	CNV E	Browser	3 7.7 M				y2,660,001.5	900,000 Samples You can press on to previo wcs 	HOLD the ope con HOLD the ope con w other samples	
Varian Varian CNV Deto CNV Brow transf CNV ca transf CN	t Gene alis Transcripts CNV Classi vser tss ts	3.7 M	genic Publicat	tions Known CN	Vs Sample Vi	w CNV Plot	CNV B	Browser	3 7.7 M				y2,660,001.5	900,000 Samples You can press on to previo wcs wcs	I HOLD the eye con d HOLD the eye con w other samples	
Varian CNV Det CNV Det CNV Brow Forward: T = = = = = = = = = = = = = = = = =	t Gene also Transcripts CAV Classi vser tts endo endo endo endo endo endo endo endo	3.7 M	genic Publicat	ions Known CN	Vs Sample Va	W CNV Plot	CNV B	Browser	3				y2,600,001.9 dowal dowal BJM	900,000 Samples You can press on to presso wcs wcs wcs	I HOLD the eye ico d HOLD the eye ico w other samples	
Varian Varian CNV Det CNV Det CNV Det CNV Cat CNV	t Gene also Transcripts CAV Classi exer transcripts A CAV Classi exer transcripts CAV Classi exer transcripts CAV	3.7 M	genic Publicati	ions Known CN	Vs Sample Va		CNV E		3 7.7 M				y2,600,001.9 dowal B.7/M PLUTERS T	900,000 Samples You can press on to press wcs wcs	(REF)	Constraints of the second seco
Varian Varian CNV Det CNV Det CNV Det CNV Brow transf: '''' ''''''''''''''''''''''''	t Gene also Transcripts CAV Classi exer Its Its Its Its Its Its Its Its Its Its	a.7 M	genic Publicati		Vs Sample Va	w CNV Plot	(CNV E		3 7.7 M		NC02778		y2,600,001.5	900,000 Samples You can press on to press wcs wcs	(REF)	Carrow
Varian Varian	t Gene also Transcripts CAV Classif	2.7 M	genic Publicati		Vs Sample Va	w CNV Plot	CNV E		3 7.7 M				y2,600,001.5	900,000 Samples You can press on to press wcs wcs	(REF)	C C C C C C C C C C C C C C C C C C C
Varian Varian	t Gene Transcripts CNV Classi rser	2.7 M	genic Publicati	ions Known CN	Vs Sample Va	w CNV Plot	CNV E		3 7.7 M REMYSHE REMYSH				y2,600,001.5	900,000 Samples Year can press on to press WCS WCS	(HOLD the eye icon wether samples [REF] (C C C C C C C C C C C C C C C C C C C
Varian Varian	t Gene Transcripts CNV Classi rser	2.7 M	genic Publicat	tons Known CN	Vs Sample Va	w CNV Plot	CNV E		3 7,7 M 1000 1000 1000 1000 1000 1000 1000 10		NC002779		y2,600,001.5	900,000 Samples Yeur can press on to press WCS WCS	I HOLD the eye icon w other samples	C C C C C C C C C C C C C C C C C C C

- CNV plot: We provide a CNV plot (static), showing how the observed read depth in the area of the CNV differs from the expected. The CNV plots are generated using a modified version of the ExomeDepth tool. You can find further information regarding this in <u>CNV Visualizations</u>.
- **Known CNVs:** We display only the relevant CNVs for the classification according to the following criteria:
 - <u>CNV deletions</u>: we retain those that fully overlap with the given CNV for gnomAD variants. For CNVs coming from clinical sources (Decipher, DBVar, ClinVar CNVs) we use the overlapping CNVs if they are benign and the contained CNVs if they are pathogenic.
 - <u>CNV duplications:</u> we keep only the CNVs encompassing the same coding genes. If the CNV is non-coding, then we retain the CNVs that have at least 85% of overlap.



Variant Gene CNN/ Detaile Transcripts	CNU Classification Bathematic Dublicatio		AN/ Dist CAD/ Browner					¢ 0
City Details Trailscripts	Civy classification Patriogenic Publicatio	anpe view C	City blowser				Description	a Facilitada
Structural Variants						y:2,600,0019,900,000		E]
strand: 1		_					show all	
				chrY: 7.2	00.775			Saved
2,7 M	3,7 M	4,7 M	5,7 M	6,7 M	7,7 M	8,7 M	9,7 M	
			1,010					112
+ Transcripts							FILTERS T	Q
CD99P1	RNF198PY	MIR9985 SERBP1P2	ENSG00000251879	TBL1Y	ZNF92P1Y	TTTY11	TSPY5P	secongs
MIR6089 ZFY-AS1		PCDH11Y	ENSG0000235094		ENSG00000236690	MTCYBP1	TTY20 RNA5SP518	
			KDT18D10					
			1					
de contracto mante la mante con							Show fewer	
OniProt Protein Regions	1 C							
+ Structural Variants							FILTERS T	
0 variants with leng	th over 100.000.000 bp 😯							
-								
								5
								Reset

Warnings are being displayed under the variant table to inform the user of the reliability of CNV calls of the sample: if (1) the correlation of the sample to its reference samples is low and (2) the number of reference samples is low.

~	varsomecinical	Samples Filter sets	Gene lists	Upload / view file	es Launch analysis	+ Support	& Manual							Tags		
	- CNV Results											c	Old layout	🕇 Filters 🛛 🛓	Analysis actions	÷
12		2										Search for gene, chr (e.g.	chrl Q	ខ • •	+ =- 4	
). Variant 🕴	Quality Score	Length 0	Call Quality	Copy Number	Туре ≬	(0	ACMG Class	CNV Rules	Number of genes 💧	Number of exons	Reads expected	Reads observed	Read ratio	Frequency	Î.
	chr16:8988292589883044	5.05	119	1.8.1	1.22651	deletion	R 🔶	Uncertain Significance	Overlap Literature	1	9	123	72	0.595	0.0001434	
C	chr1:241682871241683043	6.58	172	× × ×	1.38515	deletion	F 0	Pathogenic	Literature Overlap	1	2	666	435	0.653		
C	chr17:2942230829422408	7.82	100	× × ×	1	deletion	N	Pathogenic	Literature Overlap	2	4	156	78	0.5		
C	chr3:1018351210183892	9.66	380	× × ×	1.23634	deletion	۷ 🕘	Pathogenic	Overlap	1	3	1 012	596	0.589		Ш
	chr10:4350038043500662	5.65	282	× × ×	1.47508	deletion	R	Pathogenic	Overlap	1	2	1 739	1 208	0.695		k
C	chrX:133119282133119497	7.54	215	× × ×	1.35614	deletion	G 💿	Pathogenic	Overlap	1	4	1 034	662	0.64		
	chr8:9099673390996810	5.02	77	× × ×	1.44996	deletion	N	Pathogenic	Creelap	1	6	432	295	0.683		
C	chr9:100459383100459595	7.74	212	***	1.22897	deletion	×	Likely Pathogenic	Overlap	1	6	333	195	0.586	•	Ш
C	chr10:4357268743572800	7.55	113	× × ×	1.07313	deletion	R	Uncertain Significance		1	4	173	91	0.526		
C	chr11:29052142906740	11.5	1 526	× × ×	1.44148	deletion	c 💿	Uncertain Significance	Overlap	1	8	1 521	1 033	0.679		•
																5
Wa	ming: Low correlation.	×														
Wa	ming: Number of reference sample	es low. 🗙 💊	\mathbf{i}													
Sho	wing 1 to 10 of 15 rows 10	rows per page														

Searching through CNV results

As you inspect the CNV results of your sample, you can search by a known, or previously detected from the main analysis of the sample, SNV or small INDEL and see if it overlaps with any detected CNV.



Reads alignment visualization for CNVs

You can view the alignment of the reads in the regions of the detected CNVs on JBrowse.

NA1	A1287 ⁶ Ress Alignments on Brouse - CNV Results of NA12878_WGS_PIPE_262 (hg10)													
6	0+++++3	>									Search for ge	ne, chr (e.g. chr1 Q	1 S • 1	T- + ≣- ±
□	Variant	Length 🕴	Call Quality	Copy Number	Туре	Genes 🕴	Θ	Quality Score	ACMG Class	CNV Rules	Number of genes	Number of exons	Reads expected	Reads observed
	chrY:26000019900000	7 299 999	× x x	0	deletion	AGPAT5P1,AMELY,		2090	Pathogenic	Gere Literature Ove	132	691	58 209	1 363
	chrX:61750001154700000	92 949 999	× × ×	2.61353	duplication	AARSD1P1,ABCB7,		2390	Pathogenio	Gere Overlap	1 510	11 215	1 418 283	2 166 554
	chrX:270000152150000	49 449 999	× × ×	2.61353	duplication	ACAA2P1,ACE2,AC		1280	Pathogenio	Gere Overlap	759	6 173	776 078	1 187 039
	chrY:1390000128850000	14 949 999	× × ×	0	deletion	ACTG1P11,ACTG1		3460	Pathogenio	Gere Literature	345	1 891	86 190	1 092
	chrX:5250000158500000	5 999 999	× × ×	2.61353	duplication	ACTG1P10,ALAS2,		153	Pathogenio	Gerne Overlap	140	1 150	89 760	137 181
	chr14:106200001106750	549 999	× × ×	2.22651	duplication	ADAM6,ATP6V1G1		•	Pathogenic	Gere	96	171	5 022	5 878
	chr22:2240000123250000	849 999	× × ×	1.07587	deletion	ABHD17AP5,ASH2L		46.7	Uncertain Significance	Gene Literature	119	284	16 340	8 617
	chrY:1315000113750000	599 999	× × ×	1.42868	deletion	ACTR3BP1,DUX4L1		55.3	Uncertain Significance		13	21	21 429	14 432
	chr10:5120000151400000	199 999	× × ×	3.09085	duplication	ENSG00000174194,		8.91	Uncertain Significance		13	71	207	440
	chr14:106200001.106350	149 999	× × ×	0	deletion	ATP6V1G1P1,ENS	0	20.6	Uncertain Significance	Gere Literature	28	51	2 300	423 -

Once you have selected a variant on the Variant Table you can see the alignment of the reads by clicking on the Jbrowse icon on the top left of the screen. The CNV call region is highlighted in yellow.



Browsing through the samples of a CNV analysis

You can browse through the samples analyzed under the same CNV/SV analysis by visiting the results page of one and using the red arrows you can be directed to the next or previous sample:

	Varsomeclinical									
8	0	>	-							
	Variant 🔶	Length	Call Quality	Copy Number 🕴	т					
	chrX:67021530133119497	66 097 967	× x x	1.51197	de					
	chrX:1486166951210078	36 348 409	~ x x	1.38957	de					
	chr1:1734908317371404	22 321	× × ×	1.53804	de					
	chr8:145742414145743189	775	× × ×	2.28688	dı					
	chrY:67379336737973	40	× × ×	3.02857	dı					
	chr6:3136657531366616	41	× × ×	1.39835	de					

You also have the option to download you filtered CNV results, as it has been possible for SNP/small INDEL analyses, from the upper right corner of the Variant Table:

				Search fo	or gene, chr (e.g. chr1	Q T	រ ខ ៖	₹-	+		Ł
Ð	ACMG Class	ACMG Rules	HGVS		HGVS Protein	HGVS Co	oding	Tran	script F	Position	0
	Uncertain significance B		NM_0011	.78008	•	c.210-100	_210-99del	intron	3 of 16	positio	СВ
	Uncertain significance B	BP4	NM_0011	.78008	÷.	c.1552+91	.2dup	intron	16 of 1	.6 befor	СВ
	Likely benign	BS2 BP4	NM_0011	.78008	-	c.1467+22	24G>A	intron	15 of 1	.6 positi	СВ
	Benign	BA1 BP6 PP3	NM_0011	.78008		c.1467+12	27_1467+157d	el intron	15 of 1	.6 positi	СВ
	Benign	BA1 BP4	NM_0011	.78008		c.1467+37	7C>G	intron	15 of 1	.6 positi	СВ
	Benign	BA1 BP4	NM_0011	.78008	•	c.1467+37	'1G>A	intron	15 of 1	.6 positi	СВ
	Benign	BA1 BP4	NM_0011	.78008	•	c.1467+40	08G>C	intron	15 of 1	.6 positi	СВ
	Benign	BA1 BP4	NM_0011	.78008	-	c.1552+10)36G>A	intron	16 of 1	.6 positi	СВ
	Benign	BA1 BP4	NM_0011	.78008	-	c.1467+41	.5C>G	intron	15 of 1	.6 positi	СВ
	Benign	BA1 BP4	NM_0011	.78008	a.)	c.1467+42	1G>C	intron	15 of 1	.6 positi	СВ
	Benign	BA1 BP4	NM_0011	78008	-	c.451+29/	4>G	intron	5 of 16	positio	CB

Repeat Expansion variant table

Repeat expansion analysis variant table contains the following information:



□	0	Position v	Genes 🔶	🕚 Pathogenic 🕴	Sample repeats 🛛 🕴	MIN pathogenic repe	Normal MAX repe	Display repeat u	Zygos	Repeat u 🕴	Ref. genome repe	Filters
	••	chrX:147912050	FMR1,FMR1-AS1,	Pathogenic	881	200	55	CGG	۲	GGC	20	PASS
		chr4:3074876	HTT,HTT-AS	Pathogenic	36	36	26	CAG	•	CAG	19	PASS
		chr12:6936716	ATN1	Uncertain Significance	40	48	35	CAG	۲	CAG	19	PASS
		chrX:147912050	FMR1,FMR1-AS1,	Benign	32	200	55	CGG	۲	GGC	20	PASS
		chr12:6936716	ATN1	Benign	20	48	35	CAG	۲	CAG	19	PASS
		chr9:69037284	FXN,ENSG000002	Benign	18	51	35	GAA	۲	AAG	6	PASS
		chr9:69037284	FXN,ENSG000002	Benign	9	51	35	GAA	۲	AAG	6	PASS
		chr20:2652733	MIR1292,NOP56	Benign	6	650	14	GGCCTG	٠	GCCTGG	4	PASS
		chr14:23321472	PABPN1,BCL2L2	Benign	2	11	10	GCC	۲	GCG	6	PASS

- **Position**: chromosome and position of the repeat expansion.
- **Display repeat unit**: display repeat unit familiar to the clinician.
- **Repeat unit**: repeat unit in the reference orientation.
- Pathogenicity: repeat expansions can be classified as:
 - Pathogenic: the number of repeats detected in the sample is equal to or greater than the minimum number of pathogenic repeats.
 - Uncertain Significance: the number of repeats detected in the sample is greater than the maximum normal number of repeats but less than the minimum number of pathogenic repeats.
 - Benign: the number of repeats detected in the sample is smaller than or equal to the normal number of repeats.
- **Genes**: in which gene the repeat expansion has been found.
- **Sample repeats**: number of repeats observed in the allele.
- **MIN pathogenic repeats**: minimum number of repeats needed to consider the variant pathogenic.
- Normal MAX repeats: maximum number of repeats allowed to call the variant as normal.
- **Ref. genome repeats**: number of repeats in the reference genome.
- Zygosity
- Filters: VCF filters.

Repeat expansion cards:

- Variant: general variant information and region browser.
- **Gene:** it contains the same <u>gene cards</u> as the ones displayed under the gene level of the main analysis (small variant table).

8. Filter Sets

VarSome Clinical allows you to create 3 different filter sets:

- 1. Create new set
- 2. Create new set (SV Results): specific filter sets for CNV analysis' results.
- 3. Create new set (RE Results): specific filter sets for Repeat expansion analysis' results.



Varsomeclinical Analyses Upload / view files	Manage 👻 Launch analysis 👻 Filter sets Gene lists Help		Tags About
Filter Sets Create and manage filter sets, that you can use on multiple samples. You can dup	licate sets, reorder filters within sets or drag and drop filters from set to set.		Dashboard / Filter sets
Your Filter Sets		T Create n	ew set T Create new set (SV Results) T Create new set (RE Results)
Search filter sets			
Assigned to	≡ Zygosity • Z 🕯	Assigned to	Assigned to
HD827 Eye TUMOUR AMP.9120301883	Assigned to	You have not assigned your set to any analyses	You have not assigned your set to any analyses
	You have not assigned your set to any analyses		
Exclude Hetero		Exclude Common Variants < 🖲 🗹 🧃	Testing Gene list (SV)
Ellipse included	2 a 17 a	Ellene Induded	Ellipse included

8.1 Create a new filter set

It is possible to create a filter set directly from the analysis results or from the "Filter sets"



From the analysis results:

• Click on the **Filters** icon **T**Filters to create and manage filter sets, which you can use on multiple samples. You can duplicate sets, reorder filters within sets or drag and drop filters from set to set.

			*				
			🛓 Report generation	Analysis actions 👻			
		Search fo	r gene, chr (e.g. chr1 o Q	≣ 8 ÷ T-	+ ≣- ≛		
	HGVS	HGVS Protein	HGVS Coding	Transcript Position	Overlapping Gene:		
19	NM_004333.6:c	V600E(p.Val600	c.1799T>A	exon 15 of 18 positi			
	NM 004333.616	V600C/pVal600	C 1700TSC	avan 15 of 19 paciti	DDAE		

• Click on the **Create** icon

to create a new filter set.

Filters			×						
Select a filter set	~	Add	Create						
Total Variants: 92									
C Reset Filters	Apply F								



A pop-up window called **Filter set** will appear. This window is split into two columns: **Available Filter Sets** and **New Filter set**. The available filter sets can be selected by clicking on the green box ("Add") next to each filter's name. If you select a filter, the filter will move to the other column. By clicking on the "filter name" or on the arrow, a form is displayed where the filter can be edited. Once you have finished editing your new filter set, give it a **Name** and click on **Save filter set**.

Filter details	Save filter set
Available Filters	New filter set
✓ Call Status Add	A user-defined name A user-defined name
✓ Chromosome and Position Add	
✓ ClinVar Class Add	
✓ Custom Variant Classifications Add	
✓ Function Xed	
✓ Gene List Add	
✓ Germline Points Score Add	
✓ Germline Rules	
✓ Inheritance Add	
✓ Number of Phenotypes Add	
✓ OMIM Variants Filter Add	
✓ Pathogenicity class Add	
✓ Population Frequencies Add	
✓ Predicted Pathogenicity Scores Add	
✓ RS ID Add	
✓ Somatic Rules Add	
✓ Variant Type Add	
✓ VCF Filters	
✓ Zygosity Add	
	Sove filter set X

For example, if you want to create a filter that will give you all the variants except from the variants that are only intronic you can click on the Functions filter and then to intronic and intronic but not splicing functions as shown in the pictures below.

 Image: Second second



∧ Function	Add
 Exclude variants that match Function 3'flank 3'utr 5'flank 5'utr coding intronic splicing (show all splicing variants regardless of coding impact) non-coding exon intronic but not splicing 	3
By clicking on the Add button Add the filter	will move to column "New filter set" in which
you can give to your filter the name you want, yo	u can edit your filter and finally you can save it
by clicking on the Save filter set button	set
	Save filter set
New filter set	
A user-defined name	5
A user-defined name	
▲ Function	Remove
Exclude variants that match	
Function	
3'flank	
5'flank	
coding	
 intronic splicing (show all splicing variants regardless of coding impact) 	
non-coding exon	
Coding impact	
stoploss	
exon deletion	
☐ In frame ☐ start loss	
splice junction loss	

You can modify the name of the filter set by clicking on \square . In the following example, the filter set "Population Frequencies" will be renamed "Rare Variants":





Filter details	Save filter ×
Filter: Population Frequencies	
You may change the values in this filter set using the form below	
Name	
Population Frequencies	
Exclude matching records	
Frequency	
From	0.05
	See Ber X
Filter details	Save litter x
You may change the values in this filter set using the form below	
Name	
Rare Variants	
Exclude matching records	
Frequency	
From Intervence and genome trequencies, based on the ethnicity of the sample provided, or the general frequency if ethnicity is not provided.	0.05 In analyses done before 6 May 2017, the ExAC and 1000 genomes frequencies are used
	Save filter ×

8.2 Applying an existing filter to a variant table

Clicking on the filter icon **T** Filters it opens the **Filters** menu where filter sets can be managed. To exit this menu, simply press "**Esc**" on your keyboard.

A drop-down list will show a list of all saved filter sets. Select a filter set from the already

existing ones and add the selected filter set to the analysis with: Add . In order to apply the



selected filter set to the analysis select: **Q** Apply Filters

Filters			×
Select a filter set	~	Add	Create
Test Filter	Â		
Test Example		Apply F	
Test filter			

Each of the filters inside a filter set can be switched on and off by clicking on the green/red circles. Active filters have an associated green circle, and inactive filters have an associated red circle. Variants passing the filter are shown in blue, and variants that fail the filter are shown in green and crossed out. In the example below, all filters are active except the Function

filter.Filters can be edited 🗹 or removed 📋



The order of the filters can be changed by moving the mouse on the top of \blacksquare , then dragging the filter and dropping it at the desired position. In the example below, the "Pathogenicity class" filter has been moved to the top position:



Filters	×
Test filter	Add Create
Total Vari	ants: 2286
C Reset Filters	Q Apply Filters
🕂 Test Filter	• 2 1
≡ Pathogenicity class	<u> </u>
Class	s: 5, 4
2201	
■ Rare Variants Frequency	• 2
e equency	85
E Function Function: co	ed filter
Disabi	ed mor

Click on Apply Filters in order to filter the original results of the variant table. The number of variants that pass the filter is also shown under the variant table The number of rows of the variant table can be adjusted by selecting:

Showing 1 to 10 of 14 rows	(filtered from 19 total rows) 10	rows per page
	Filters used	
s (filtered≦	Pathogenicity class: Class: 5, 4 Rare variants: Frequency: 0.0 - 0.01	1

When downloading all the filtered variants from the Variant table in excel format there will be a separate tab called "Filters used" that mentions which filters were used.



A Please note that if existing filters with ACMG rules do not work, you will see an error message returned, like the following:



In this case you should edit (as shown above) the filter, by adding the rules you would like to have from the available options.

8.3 Shared Filters

Users of the same group can now share filters among them. To share a filter with other members of your group, click the Filters link at the navigation bar and then click the share icon on any of the filters you wish to share.

Filters that have been shared among members of the same group appear both on the Filters screen and when filtering a specific analysis. Although a shared filter can be used by anyone in the group, it can only be edited by the user that initially shared the filter. Other group members can still duplicate the filter and edit it. In the analysis results page, click on the Filters icon

TFilters , use the drop down list of the search box to select the shared filter of interest, click on

Add the shared filter set to the analysis and then click on Apply Filters

The sharing icon is a click toggle to switch between sharing status. Filter sets shared with you will show up normally on analysis side-panel alongside your owned filter sets (if they are shared, it will say by whom). Note that filter-sets sharing is only within VarSome Clinical organizations, you cannot share them, for example, with people outside your organization.

Care must be taken for shared filters because user-created filter sets are essentially composed of filter instances (e.g. a filter instance is the filtering of pathogenic variants), and filter instances can be turned on or off by the creator of the filter set. So, if the filter set is shared, other users who use this filter set may be able to apply it to their analyses, but may see that not all filter instances are enabled, and they can do nothing about it. If the original creator of the filter set chooses to, for example, disable a filter instance, then all other users who use that shared filter set will see changes in their results because of that. To avoid this you can duplicate the shared filter set as one of your own so that you have direct access to what is enabled or not.



Filter Sets

Create and manage filter sets, that you can use on multiple samples. You can duplicate sets, reorder filters within sets or drag and drop filters from set to set.

Your Filter Sets

Search filter sets								
uthogenicity Pathogenic and Likely								
Filters Included								
\equiv Pathogenicity class	• 🖸 盲							
Assigned to								
G30								

Shared Filter Sets (Saphetor)

-link	ø
Shared by:	
ilters Included	
≡ CGD Inheritance	
ssigned to	
_fastq - Analysis filtered by Intellectual disability PanelApp 20201112 as of 27-Nov-2020 (2020 Intellectual_disability.Pan)	27 Nov
fastq - Analysis filtered by Gene list from (any) phenotypes Autistic Behavior, Stereotypy, Absent Speech, Large For Gestational Age, Intellectual Disability, Severe as of 27-Nov-2020 (27 No	, w 2020

Varsomecinical Analyses Upload/viewfiles Manage - Launch analysis - Filtersets Gene lists Help Tags											About +					
Pharm_GKB.Initial (hg19) @D										generation Analysis actions +						
1	🥔 📴 🖶 🐟 🍖						e.g. chr1 c 🔍 🧃	ø	Filters	•	×					
	□	Variant 🔶	Variant Type	Gene Symbol	0	Germline Class	Germline Rules	HGVS	HGVS Protein	HGVS Coding	Transcript Position	Overlapping Genes	0 Ir			~
		:hr7:148508727 T→A 🎭	SNV	EZH2		Pothogenic	FM1 Strong FM2 Strong FM5 Strong F53 FF2 FF3 F53	NM_004456.5:c	Y646F(p.Tyr646	c.1937A>T	exon 16 of 20 positi	EZH2	AD	Add Creal		
	0	:hr21:36252972_3 insT	Insertion (1)	RUNX1		Pathagenic	PVST PM2 Streng	NM_001754.5:c	T131Hfs*7(p.Thr	c.389_390insA	exon 5 of 9 before p	RUNX1	AD	C Reset Filters	Q Apply Filters	H
	•	hr6:32551885 C→T	SNV	HLA-DRB1	•	Likely pathogenic	@ @	ENST00003600	÷	c.370+1G>A	intron 2 of 5 positio	HLA-DRB1		+	6 11	
		:hr6:32549584_5 insT	Insertion (1)	HLA-DRB1	••	Uncertain significance P	693	ENST00003600	T135Dfs*23(p.T	c.401dup	exon 3 of 6 before p	HLA-DRB1	•		ritter o 🖉	-
	0	:hr11:112832340 delC	Deletion (1)	NCAM1	٠	Uncertain significance P	(75)	ENST00005246	Q62Kfs*110(p.G	c.184del	exon 1 of 19 positio	NCAM1,LOC101928847		Description: Cancer driver genes (2 genes)	2
	0	:hr4:39064161_2 insC	Insertion (1)	KLHL5	•	Likely benign	852	ENST00003596	110Hfs*43(p.Ile1	c.27_28insC	exon 1 of 11 before	KLHL5		= Inheritance	•	2
		:hr11:89017961 G⇒A	SNV	TYR		Benigs	EA1 EFS PM1 Strong PMS Strong PP2 (PP) PPS	NM_000372.5:c	R402Q(p.Arg40	c.1205G>A	exon 4 of 5 position	TYR	AR	Description: Inheritance :Autosom	al Recessive	Γ.
	0	:hr1:115236057 G⇒A	SNV	AMPD1		Benigs	(BA1) (PPS Moderate) (PP3)	ENST000005201	Q45*(p.Gln45Ter)	c.133C>T	exon 2 of 16 positio	AMPD1	AR	Impacts: $\theta \rightarrow 0$		
		:hr16:88872145 T⇒C	SNV	CDT1		Denign	BA1 BP6 Very Streng	NM_030928.4:c	C234R(p.Cys234	c.700T>C	exon 5 of 10 positio	CDT1	AR			


9. Gene List

VarSome Clinical offers 3 ways of creating a new gene list:

- 1. Copy and modify an existing gene list
- 2. Create a new gene list
- 3. Create a gene list from phenotypes

Varsomeclinical Analyses Upload / view files Manage - Launch anal	ysis - Filter sets Gene lists Help
Gene lists Create and manage list of genes that you can use in your filters.	
Main Lists PanelApp Lists T Create new gene list T Create gene list from phenotypes	
Showing 1 to 10 of 310 rows 10 - rows per page	
aaa1 💿	aaa1 (Copy) 👔
8 1 12 1	8 8 2 2

9.1 Copy and modify an existing gene list

The options to manage the existing gene lists are the following:



Genes can be added or removed by clicking on the "Edit" button *Conce you have finished* editing the gene list, you can save the list with either the original name or with a new one.

Name
MODY genes (Copy)
List of gene names according to genenames.org standard
GCK HNF1A HNF1B HNF4A NEUROD1 PDX1
🖏 Save list
4

Finally, these gene lists can also be deleted by clicking on the "Trash" button **1**.

9.2 Create a new gene list

- Click on "Create new gene list".
- A pop up form window appears, you just need to write a name for your gene list and paste the list of genes that you want to analyze. Important note: the gene names need to follow the HGNC (Hugo Gene Nomenclature Committee) standard (http://www.genenames.org/), if the gene names do not follow this rule you will get an error message after clicking on **Save list**.

Gene list	
Name	
Fill in the name of your list	
Notes / comments	
Add any notes or comments related to this gene list	
Search for genes	
Start by typing a gene symbol and append them to the list by selecting from the res	ults
List of gene names according to genenames.org standard	
Fill in or paste a comma or space or semicolon or line by line separated list of gene	names



9.3 Create a gene list from phenotypes

- Click on "Create gene list from phenotypes"
- Start by filling in a phenotype. By selecting All, phenotype matching records from HPO, MONDO and OMIM[®] will be available

Phenotypes to gene list Create a gene list from several phenotypes.	Dashboard / gene lists / Phenotypes to gene list			
Phenotypes to gene list				
Start by filling in a phenotype. Click on phenotype from the drop down menu Select one or more phenotypes from your list Finally select one or more genes to be added	to add it to your phenotypes list. (You can right click any phe to view associated genes. to your gene list.	inotype to remove it from the list)		
Search for phenotype names from: All Only OMIM				
Start filling in a phenotype (type 3 cha	racters minimum). Search multiple by separating with semic			
Selected phenotypes ()		Associated genes	Selected genes	
🗑 Clear	₩ Select all	► Select all	Tear 2	G Save list

- Click on phenotype from the drop down menu to add it to your phenotypes list. (You can right click any phenotype to remove it from the list)
- Select one or more phenotypes from your list to view associated genes.

Phe Create	enotypes to gene list a gene list from several phenotypes.	Dashboard / gene lists / Phenotypes to ge	ne list	
Phe	notypes to gene list			
	Start by filing in a phenotype. Click on phenotype from the drop down menu to add it to your phenotypes list. (You can right click Select one or more phenotypes from your list to view associated genes. Finally select one or more genes to be added to your gene list.			
	Search for phenotype names from: All Only OMIM			
	Start filling in a phenotype (type 3 characters minimum). Search multiple by separating with	i semik		
	Selected phenotypes 1	Associated genes	Selected genes	
	Alzheimer Disease, Familial, 1 (OMIM:104300)	APP	A	
	Takenouchi-Kosaki Syndrome (OMIM:616737)	HFE		
		NOS3		
		МРО		
		CDC42	•	
			Clear B Save list	
		Select all (6)		



• Finally select one or more genes to be added to your gene list

Phe Create	enotypes to gene list a gene list from several phenotypes.			Dashboard / gene lists / Phenotypes to gene list
Pher	notypes to gene list			
	Start by filling in a phenotype. Click on phenotype from the drog down menu to add it to your phenotypes list. (You can right or Select one or more phenotypes from your list to view associated genes. Finally select one or more genes to be added to your gene list.	ck any pheno	type to remove it from the list)	
	Search for phenotype names from: All Chyl CMIM © Start filling in a phenotype (type 3 characters minimum). Search multiple by separating	with semic		
	Selected phenotypes ()		Associated genes	Selected genes
	Alzheimer Disease, Familial, 1 (OMIM:104300)		HFE	HFE 🗑 MPO 🗑 PLAU 🗑
	Takenouchi-Kosaki Syndrome (OMIM:616737)		NOS3	
			MPO	
			CDC42	
			PLAU	
			Select all (6)	Clear Save list

Note: The methodology to generate a gene list from phenotype(s) has been changed. Previously, when making a gene list from phenotypes, we would include only those genes that are directly annotated with that phenotype. We have now extended this and instead first collect all diseases linked to the phenotype and then all genes linked to those diseases, as well as any genes directly linked to the phenotype. We already worked this way when adding phenotypes to analyses, so this change ensures we are consistent and also makes sure we don't miss any genes when creating gene lists.

If you want to filter your gene list, you have to go to the Variant table page of the CNV analysis

you performed and click on the filter icon **Filters** on the left. Click on "Create" to create a new one and you will see a screen like the one below:

Filter details		Save filter set 🗙
Available Filters	New filter set	
✓ Call Status	Add A user-defined name	
✓ Chromosome and Position	Add	
✓ ClinVar Class	Add	
✓ Custom Variant Classifications	Add	
✓ Function	Add	
Gene List Exclude variants that match	3 Add	
Gene List Select Gene List	~ 2	

Click to add a Gene list filter for the specific phenotype you wish. Fill in the name of the filter set and save it in order to be applied on your analysis.

9.4 Browsing through gene lists



• Search gene lists by name:



• Toggle custom view:

Search					C	۹.	Î	្ត	0	
									Тод	gle custom view
	•	1	2	3	4	5		18	>	

• Display gene lists sorted by name or update date:

Show	ing 31 to 40 of 179 rows 10 - rows per page			x 1 2 3 4 5 18 >					
	Name	Owner 🔶	Updated at	Actions					
+	GeneSampleName3726944998	Contra in contraction	27 Sep 2021	• 0 0/ W					
+	GeneSampleName3752256782	Service and Services	23 Sep 2021	• B 12 W					
+	Gene list 23-09-2021 11:18:14	termine and the second s	23 Sep 2021	• B 27 B					
+	Gene list 23-09-2021 11:17:41	Anne commen-	23 Sep 2021	• B 12 W					
+	Gene list 09-09-2021 09:51:02	Anne contains	9 Sep 2021	• B 22 B					
+	Gene list 09-09-2021 09:50:29	the second se	9 Sep 2021	• B 2/ B					
+	GeneSampleName2977764964	territori de la constante de la	9 Sep 2021	● B 26 ■					
+	Gene list 09-09-2021 09-23:11	New Contestion	9 Sep 2021	• B 26 ·					
+	GeneSampleName9517529787	And a second sec	9 Sep 2021	b 2 2					
+	asdasd	All commencements	7 Sep 2021	• B 12 ·					
Show	Showing 31 to 4 of 179 rows (1 + 2 + 3 + 4) =								



10. **Final Report Generation**

When it comes to reporting, in addition to the options provided by the "Analysis Actions", VarSome Clinical offers users the option to generate a report of specific variants in PDF or Word format.

Variant reporting works on a sample level, i.e. you can report multiple variants belonging to the same sample, but you can't report one variant of the sample A and another of the sample B within the same report.

In the Variant Table page, you can either select one variant at a time just by clicking on it or you can select multiple variants at once (mass selection) by holding Ctrl and clicking on each variant you wish to include to your list and then click on the "Select for export" 🕮 📩 option. On the 🚣 Report generation

or

Export

Generate

Report

to see the list of the variants selected

left, click on the "Report generation"

for export, and then either choose



HD		(hg19) 🤇								Tilters Leport generation Analysis actions +
	± 0• ≧ = ●	- 🕐							Search for	gene, chr (e.g. chr1 c Q 🕤 😂 + 🔽 + ⊞ - 🛓
□	Variant	Variant Type	Gene Symbol	0	Somatic Tier 💧	Somatic Rules	Gample Metrics	O Somatic samples	Germline Class	Germline Rules
	chr3:178947865 G⇒A 🎭	SNV	PIK3CA 🎭	••	Ter	Ctd) Ores Cerm Soms (Pols) cabil (rec) (red)	\$† • ⊕ ≗ ⊕ • Ք	31	Pathogenic	PPS Very Strong (PS) (PP3 Strong (PM1) (PM2 Supporting
	chr7:140453136 A⇒T 🌏	SNV	BRAF 🌯	••	Tier	Ctd) Gray Crrm) Sons) Febs) Fabil Free Preds	1 • • 🔒 🖞 🗢	5.565	Pathogenic	P53 Very Streng (PMI Streng) (PMS Streng) (PP5 Streng) (PP3 Maderate) (PM2 Supporting) (853 Support
	chr3:178936091 G⇒A 🌏	SNV	PIK3CA 🎭	••	Tier		ቱ ₺ ≗ ା ∘ ሇ	4.521	Pothogenic	PS3 Very Strens (PF5 Very Strens) (PM1 Strens) (PM5 Strens) (PF5 Moderate (PM2 Supporting)
	chr3:178952085 A⇒G 🍋	SNV	PIK3CA 🌯	••	111		* 5 🛎 💿 이 사	4.572	Pathogenic	P53 Very Streng PP5 Very Streng PM1 Streng PM5 Streng PM2 Supporting PP1
	chr10:43615641 A⇒C	SNV	RET	•	Tierl	Cray (and) (and) (and)	\$* ≜ ≗ ⊜ ০ ∿	3	Likely pathogenic	PM3 PP3 Mederate PM1 Supporting PM2 Supporting PP5
0	chr10:43604493 C→T	SNV	RET	••	Tierl	Cres] Cons) rath (real)	\$† ≜ ≗ ⊜ օ Դ	10	Likely pathogenic	773 Strong (FNS 7M2 Supporting 773 773
	chr17:37881082 G⇒A	SNV	ERBB2		Tier	Oreg] Germil Path Free] Pred	\$† ≜ ≗ ⊜ օ Դ։	1	Likely pathogenic	PP3 Stroog PM1 PM2 Supporting
	chr17:41234451 G⇒A 🌏	SNV	BRCA1 🎭	••	Tier II		\$ <u>b</u> ≜ ⊕ o ⊕	8	Pathogenic	PVS1 PP5 Very Stross PM2 Supporting
	chr17:29553478_9 insC	Insertion (1)	NF1	••	TierII	(ISN (ISN (ISN (ISN) (ISN)	\$•≜≜⊜∘∿	132	Pathogenic	PVS1 PP5 Very Strong PM2 Supporting
	chr1:115256530 G⇒T 🌯	SNV	NRAS	•••	TierII		\$* ₺ ≛ 🔍 ० ₺	1.136	Pathogenic	PS3 PM1 Strong PMS Strong PP3 Moderate PM2 Supporting PPS
4		_								,
Showin	• 1 2 3 4 5 6 7 8 9 10 rowsperpage • 1 rowsperpage • 1 rowsperpage • 1 rowsperpage • 1									

		-									
HD	Se	elect for export	00	2							TFilters & Report generation Analysis actions -
9	Search for gene, dr (e.g. dr 1 ⊂ Q I Z + T+ II+ ±										
□.	Variant 0	Variant Type	Gene Symbol	0	Somatic Tier	Somatic Rules	Sample	Metrics	O Somatic samples	Germline Class	Germline Rules
	chr3:178947865 G⇒A 🌯	SNV	PIK3CA 🎭	••	Tier)	Crtd) Grug) Germ) Soma) Pubs) Pathil Freq) Predil	* <u>►</u> = :	• ሌ	31	Pathogenic	(PPS Very Strong (PS) (PP3 Strong (PM1) (PM2 Supporting
	chr7:140453136 A⇒T 🍋	SNV	BRAF 🌯	••	Tier	Crtd Greg Germ Gona Puba Path Freq Pred	* 5 🕹	የጉየ	5.565	Pathogenic	PS3 Very Strong PM1 Strong PMS Strong PPS Strong PP3 Moderate PM2 Supporting BS3 Suppo
	chr3:178936091 G⇒A 🎭	SNV	РІКЗСА 🌏	••	Tier	Crtdl Oregi Cermi Somal Publi Pathil Freq Pred I	* 5 🛔	የጉየ	4.521	Pathogenic	P33 Very Strong PP5 Very Strong PM1 Strong PM5 Strong PF3 Moderate PM2 Supporting
	chr3:178952085 A→G 🎭	SNV	PIK3CA 🍤	• •	Tier	Crtd) Oreal Germ) Somal Publi Pethil Frea	* 5 🛔	• ሇ	4.572	Pathogenic	PS3 Very Strong PPS Very Strong PM1 Strong PMS Strong PM2 Supporting PP3
	chr10:43615641 A⇒C	SNV	RET	٠	Tier		≉ <u>b</u> ≞ ∈	• ሌ	3	Likely pathogenic	PMS PP3 Maderate PM1 Supporting PM2 Supporting PP5
	chr10:43604493 C⇒T	SNV	RET	••	Tier		≉ <u>b</u> ≞ (የሌ	10	Likely pathogenic	PP3 Strong PMS PM2 Supporting PP2 PP3
	chr17:37881082 G⇒A	SNV	ERBB2		Tier	Croy Cornel Fath Freq Fred I	⊅ <u>b</u> ≞ :	• ሌ	1	Likely pathogenic	PP3 Strong (PMT) (PM2 Supporting
	chr17:41234451 G→A 🎭	SNV	BRCA1 🎭	••	Tier II	Citil Orași Cormi Sonai (Visi (Rih) (Tecil (recil	⊅ <u>b</u> ≟ :	• • •	8	Pathogenic	PVS1 PPS Very Strong (PMZ Supporting
	chr17:29553478_9 insC	Insertion (1)	NF1	••	Tier II	Cited Down Corrent Const Path (Spee) Free)	₽ <u>b</u> ≞ (• ሌ	132	Pathogenic	PVS1 PPS Very Strong (PM2 Supporting)
0	chr1:115256530 G⇒T 🎭	SNV	NRAS	••	Tier II	Crtd) Oray II Corm Coma Publa Path II Freq Pred II	* 5 4	• ሌ	1.136	Pethogenic	PS) PM1 Strong PM5 Strong PP3 Maderate PM2 Supporting PP5
•	с										
Showi	Showing 1 to 10 of \$126 rows 10 + + + + + + + + + + + + + + + + + +										



Varsomeclinical										
Variants to	export									
chr3:17893609	1 G⇒A	Ť								
chr3:17895208	5 A⇒G	Ē								
chr3:17894786	5 G⇒A	Ť								
chr7:14045313	6 A⇒T	۲								
Export list	Remove all	Generate Report								
Showing 1 to 4 of 4 rows										

Selecting

Export

will export the variants in an .xlsx file as shown below:

	A	в		U U	E	F		H			K		M			I P	Q	R				V .	
1	Variant	Chromosof	Position	RS ID	Ref seq	Var seq	Туре	HGVS	Genes	Phenotype	Number of	Transcript	OMIM pher	OMIM inhe	Function	Functions	Coding im	Inheritanc	ClinVar cla	ClinVar dis	Cosmic pri ^a	Allelic bala	Freq
2	chr3:1789	chr3	178936091	rs10488600	G	Α	SNV	NM_006218	PIK3CA			View	Gastric cano	er;Nonsmall	•coding	coding	missense	AD	Pathogenic		Breast	0.0732600#	4,03:
3	chr3:1789	chr3	178952085	rs12191321	A	G	SNV	NM_006218	PIK3CA			View	Gastric cand	er;Nonsmall	*coding	coding	missense	AD	Pathogenic		Breast	0.1323	4,02
4	chr3:1789	chr3	178947865	rs58777693	G	A	SNV	NM 006218	PIK3CA			View	Gastric cano	er;Nonsmall	coding	coding	missense	AD	Pathogenic		Large Intest	0.2897	
5	chr7:1404	∳chr7	140453136	rs11348802	A	т	SNV	NM_004333	BRAF			View	Nonsmall ce	Autosomal	coding,non-	coding	missense	AD	Pathogenic		Thyroid	0.1013	3,975
6																							



direct you to the following screen:



short2_L001_R2_001 clinical	report (hg19)			Dashboard / Samples / Analysis clinical re
Clinical report				▲ Download as PDF Export as docx Re
	Saphetor Your organization's address Tet: Your organization's phone		~	varsomeclinical
				••
	Sample:	short2_L001_R2_001	Predicted gender:	
	Phenotypes		Diseases	
		Sample short	2_L001_R2_001 Report	0 +
	Primary finding			# +
	Secondary finding			.
	Drug related			

The form we provide as a template to create your report is completely customizable, where all section tables can be edited according to the user's needs. The final report layout you choose to generate will be saved for future use.

The report template consists of table sections, where all fields can be changed, along with the colors of both text and background and even the logo can be replaced with the logo of your organization.



In each section table, there is the option of changing the background color of a cell by right-clicking on it and then going to Cell >> Cell properties as shown below:



D D X B I U	5	© © © ×, ×' ≮ 1	•	→ Q 均 甲 = == ※ 本 が	· · · · · · · · · · · · · · · · · · ·		Varsomecinica
Styles •	1.0	format - Fo	nt	Size short2_L001_R2_0	· ▲· Ø· ₪ ?	Predicted gender:	
Phenotypes		Cell Row Column		Insert Cell Before Insert Cell After Delete Cells		Diseases	
		Delete Table Table Properties	Me Me	Merge Cells Merge Right Merge Down	mple short2_L001_R2_00	•4	
Primary finding				Split Cell Horizontally Split Cell Vertically Cell Properties			• 4

In the Cell properties menu you can choose to change the background color among other available styling options:

Cell Properties		×
Width	Cell Type	
pixel •	Header •	
Height pixels	Rows Span	
Word Wrap Yes 🔻	Columns Span	
Horizontal Alignment	Background Color	
<not set=""> •</not>	rgb(204, 51, 102)	Choose
Vertical Alignment	Border Color	
SHOL SELS		Choose
	ок	Cancel

What is more, by clicking on the *icon*, a Report widgets menu will be shown, with the following options:





- Analysis information: General information including sample name, allelic balance, predicted gender, phenotypes, and diseases.
- **Content area:** Field to write content-related information of the analysis.
- All variants' information: Table with all selected variant's information.
- Variant information: Mutation identification, followed by its classification and a short description of the variant. More information is included containing HGVS notation, related gene, exon, variant type, rs ID, zygosity, frequency, and coverage.
- Variant references: List of scientific papers referring to a specific variant.



- **Gene information:** Disease-related information from CGD, about the gene containing a specific variant.
- Related drugs (if applicable): Information regarding the drug-gene interactions from DGIdb, and other data resources.
 For example:

Pharmacogenomics Knowledge Base (PharmGKB)

PharmGKB is a pharmacogenomics knowledge resource including clinical information such as clinical guidelines, drug labels, potentially clinically actionable gene-drug associations, and genotype-phenotype relationships. Information on the impact of human genetic variation on drug responses is aggregated through various approaches which include, among others: literature review for the annotation of genetic variants and gene-drug-disease relationships, curation of FDA labels, and drug dosing guidelines in a pharmacogenomic approach.

VarSome integrates information from PharmGKB that may concern:

• The relationship between a variant selected from the table and corresponding medications along with related supporting publications. This information, if available, is located in tabular format in the "PharmGKB" Tab under the Variants Table. An example is shown in the following screenshot.

PharmGKB Version: 21-Mar-20	hamGKE Version: 21.Mar.2023											
Chemical Relations												
Drug Association Significant		Drug Variant Relation	Pharmacodynamic	Pharmacokinetic	Publications	Annotation ID	Curator notes					
Cetuximab	primab Not associated No Efficacy		Efficacy		-	22734028	1447247696	Expand notes				
Disease Relations												
Disease			Association		Pharmacodynamic		Pharmacokin	etic				
Progression-free survival			Not associated		No		No					
Colorectal neoplasms			Not associated		No		No					

This Table holds the Clinical Annotations of the variant which can be also accessed at the PharmGKB site through the respective ID. The "Annotation" column refers to manually curated genotype-based summaries describing the phenotypic impact of the variant. The user can also see an overview of the corresponding Publication. If the corresponding data are available, the user may also see a table for chemical relations or disease relations.

• This section concerns the relations of the *gene* containing the variant with drug substances. All the information can be found under the "Gene basic info" tab on the right menu of the Variant Table page. If the corresponding information is available the user may see a table for Guidelines, Chemical relations, Drug Label Annotations and Disease Relations. The user can also see an overview of the corresponding Publications.



Variant 💿 Gene KR	AS IN	-					Current Annotation
Gene basic Info Gene fund	tion Region Browse	er Structural Variants Transcript	Publications Known gene variant	s Gene Expression JAX CKB CIV	iC PMKB dbNSFP GnomAD Genes EB	I Gene2Phenotype GenCC ClinGen Dise	ase Validity DOMINO PanelApp gene lists
Clinical Genomic Database	Human Phenotype Or	ntology Human Protein Atlas Fus	ion GDB Community Contributions	Cancer Gene Census Pharm GKB F	DA DGI CPIC AACT Clinical Trials Prot	ein Viewer OMIM ® LOVD	
Gene basic info							
	Vieter	Description			Synonyms	c	Cytobands
	Kirsten	rat sarcoma virai oncogene nomolog			KRAS1		12012.1
	DECIPHER	Clinical resource	es Reference ONIM MalaCards		HONC GENAT	Other database links	GG Pathway
PharmGKB Version: 21-Mar	-2023						
Chemical Relations							
Drug	Association	Pharmacodynamic	Pharmacokinetic		Publicati	ions	
Encorafenib	Associated	No	No				
Cetuximab	Ambiguous	Yes	No	20978259, 225	37608, 22734028, 23071293, 23090619, 23324806,	, 24727325, 25183481, 25210463, 26162609	Show 5 more
Sotorasib	Associated	No	No				
Ramucirumab	Associated	No	No				
Capecitabine	Associated	No	No		261626	09	
Oxaliplatin	Associated	No	No		22734028, 20	6162609	
Trametinib	Associated	No	No				
Vemurafenib	Associated	No	No				
Fluorouracil	Ambiguous	No	No		22734028, 23	3324806	
Panitumumab	Ambiguous	Yes	No		23090619, 23324806, 24727325, 25183481, 252104	63, 26162609, 26438111, 26812186, 2789726	1
 Snow 5 more 							
rug Label Annotations							
Drug		Title				PGX Level	FDA Biomarker List
		Annotation of ema label for encoral	enib and hras, kras, nras			Actionable PGx	
Encorafenib		Annotation of hcsc label for encora	fenib and hras, kras, nras			Actionable PGx	
		Annotation of fda label for encorafe	nib and hras, kras, nras			Actionable PGx	On FDA Biomarker List
		Annotation of ema label for cetuxin	ab and egfr, kras, nras			Testing required	-
		Annotation of hcsc label for cetuxin	nab and egfr, kras			Testing required	
Cetuximab		Annotation of fda label for cetuxima	ib and egfr, kras, nras			Testing required	On FDA Biomarker List
		Annotation of pmda label for cetuxi	mab and egfr, kras			Testing required	-
		Annotation of hcsc label for sotoras	ib and kras			Testing required	
isease Relations							
Disease				Association	Pharmacodynamic		Pharmacokinetic
Event-free survival				Not associated	No		No
Rectal neoplasms				Associated	No		No
Neoplasms				Associated	Yes		No
Progression-free survival				Ambiguous	No		No
Overall survival				Ambiguous	No		No
Adenocarcinoma				Associated	No		No

The gene-related tables are also available for the user during the report generation process which is described in more detail in the <u>Final Report Generation</u> above. The addition of the PharmGKB section in the report will generate a table similar to the following example:



BRAF related drugs from PharmaGKB Chemical Reactions

Drug ID	Drug Name	Association	Pharmacodynamic	Pharmacokinetic	PUBMED References	Drug Label Annotations
PA166179872	encorafenib	associated	False	False		['name': 'Annotation of EMA Label for encorafenib and BRAF', 'testing_levef': 'Testing required', 'id': 'PA166182948') ['id': 'PA166179878', 'testing_level': 'Testing required', 'biomarker_flag': 'On', 'name': 'Annotation of FDA Label for encorafenib and BRAF'}
PA166129522	nivolumab	associated	False	False		[id': 'PA166160121', 'testing_level': 'Informative PGx', 'biomarker_flag': 'On', 'name': 'Annotation of FDA Label for nivolumab and ALK,BRAF,CD274,EGFR']
PA165946873	vemurafenib	associated	False	False		['id: 'PA166114482', 'testing_level': 'Testing required', 'biomarker_flag: 'Off (Never On)', 'name': 'Annotation of EMA Label for vemuratenib and BRAF'} ['id': 'PA166104858', 'testing_level': 'Testing required', 'biomarker_flag: 'On', 'name': 'Annotation of FDA Label for vemuratenib and BRAF', 'name': 'Annotation of PDMA Label for vemuratenib and BRAF', 'testing_level': 'Testing required', 'id': 'PA166160851'} ('name': 'Annotation of HCSC Label for vemuratenib and BRAF', 'testing_level': 'Testing required', 'id': 'PA166127728'}
PA166114911	dabrafenib	associated	True	False		 (id: 'PA166119826', 'testing_level': 'Testing required', 'biomarker_flag: 'Off (Never On)', 'name': 'Annotation of EMA Label for dabrafenib and BRAF', ('name': 'Annotation of HCSC Label for dabrafenib and BRAF', 'testing_level': 'Testing required', 'id: 'PA166127666') ('id: 'PA166114912', 'testing_level': 'Testing required', 'biomarker_flag': 'On', 'name': 'Annotation of FDA Label for dabrafenib and BRAF')
PA166115364	trametinib	associated	False	False		('id': 'PA166115365', 'testing_level': 'Testing required', 'biomarker_flag': 'On', 'name': 'Annotation of FDA Label for trametinib and BRAF'} ('name': 'Annotation of HCSC Label for trametinib and BRAF', 'testing_level': 'Testing required', 'id': 'PA166127722')

- Precision Medicine Knowledgebase (PMKB)

PMKB is a knowledge database that presents clinical interpretations of cancer variants in a structured way. It allows user contributions in terms of browsing, submitting, and editing existing entries. In addition, all changes are reviewed by molecular pathologists and oncologists.

VarSome presents the information retrieved from PMKB in the "Genes" Tab on the right menu of the Variant Table page:

PMKB	MKB Version: 21-Sep-2022								
Tier	Tumor	Tissue	Definition	Interpretation	PUBMED references				
1	Adenocarcinoma	Breast Ovary Prostate	BRCA1 any mutation	Inactivating mutations in BRCA1 may be associated with sensitivity to PARP inhibitors. Drug Rucaparib Niraparib Olaparib	-				

The aforementioned information is also available for the user during the report generation process which is described in more detail in section 15. <u>Final Report Generation</u>. The addition of the PMKB section in the Report will generate a table similar to the following example:



ARID1A any mutation

Tumor

This gene is a known cancer gene. ARID1A/BAF250A subunit of the SWI/SNF (BAF) chromatin remodeling complex has emerged as recurrently mutated in a broad array of tumor types and a potential tumor suppressor. There is evidence indicating that ARID1A-mutated cancers may be subjected to therapeutic intervention.

cinar Cell Carcinoma, Acinic Cell Carcinoma, Acute Myeloid Leukemia, Adenocarcinoma, Adenoid Cystic Carcinoma, Adenosarcoma, Ameloblastic Tumor, napiastic Large Cell Lymphoma, Angioimmunoblastic T-Cell Lymphoma, Angiomatoid Fibrous Histiocytoma, Angiomatosis, Angiomyolipoma, Angiosarcoma, storcytoma, Anaplastic, Atypical Chronic Myeloid Leukemia, B Lymphoblastic Leukemia/Lymphoma, Basal Cell Carcinoma, Burkitt Lymphoma, Carcinoid Tumor, carcinoma, Carcinosarcoma, Cholangiocarcinoma, Chondrosarcoma, Chordona, Choroicoarcinoma, Chromophobe Renal Cell Carcinoma, Chronic Lymphocytic eukemia, Chronic Myeloid Leukemia, Chronic Myelomonocytic Leukemia, Chronic Neutrophilic Leukemia, Classical Hodgkin Lymphoma, Carer Cell Carcinoma, Jear Cell Renal Cell Carcinoma, Craniopharyngioma, Dermatofibrosarcoma, Desmoplastic Small Round Cell Tumor, Diffuse Large B Cell Lymphoma, Ductal arcinoma, Ependymoma, Essential Thrombocythemia, Eving Sarcoma, Fibromatosis, Follicular Carcinoma, Gastrointestinal Stromal Tumor, Gearm Cell Tumor, Giant Cell Tumor, Globlastoma, Glomus Tumor, Granular Cell Tumor, Hairy Cell Leukemia, Hemangioendothelioma, Hepatocellular Carcinoma, Jisosarcoma, Lobular Carcinoma, Lymphoplasmacytic Lymphoma, Maignant Mullerian Mixed Tumor, Mante Cell Lymphoma, Marginal Zone B Cell Lymphoma, Marginat Mullerian Mixed Tumor, Marginal Zone B Cell Lymphoma, Marginal Zone B Cell Lymphoma, Marginal Zone B Cell Lymphoma, Margina Zone B Cell Lymphoma, Margina Zone B Cell Lymphora, Neuroendocrine Neoplasm, NK Cell Lymphory, Mucinous Tumors of Ovary, Muccepidermoid Carcinoma, Melodysplastic Syndrome, Myeloproliferative Disorder, NLPHL, Non-Small Cell angoharynegal Carcinoma, Reurobalstoma, Abaidomyosarcoma, Sacrona, Sc

Adrenal Gland, Anus, Ampulla (Pancreaticobiliary Duct) Appendix, Bladder, Blood, Bone, Bone Marrow, Brain, Breast, Spinal Cord, Cervix, Chest Wall, Colon, Endometrium Esophagus, Eye, Fallopian Tube, Fibroadipose Tissue, Gall Bladder, Kidney, Larynx, Liver, Lung, Lymph Node, Nasal Cavity, Oral Cavity, Ovary Pancreas, Parathyroid, Penis Peripheral Nervous Syster Peritoneum, Pharynx, Pituitary, Placenta, Pleura, Prostate, Retroperitoneum, Salivary Gland, Seminal Vesicle, Skeletal Muscle, Skin, Small Intestine Soft Tissue, Spleen, Stomach, Testis, Thymus, Thyroid, Tonsil, Ureter, Uterus, Vagina, Rectum Cartilage, Blood Vessel, Buccal Swab, Heart, Trachea, Salivary Duct, Spermatic Cord, Vulva, Infratentorial, Supratentorial, Gastroesophageal Junction, Sellar, Suprasellar

Tissue

***** 4

- Drug-Gene Interaction Database (DGIdb)

DGIdb provides information about known or potential drug-gene interactions using expert curation and text-mining methods. The presented drug-gene interactions have been mined from DrugBank, therapeutic target database (TTD), PharmGKB, a list of targeted agents in lung cancer, and ClinicalTrials.gov. In addition, retrieved data from Gene Ontology, dGene, and "druggable genome" lists from Hopkins and Groom (2002) and Russ and Lampel (2005) are utilized for the categorization of the genes as potentially druggable.

VarSome presents the information retrieved from DGIdb at the bottom of the "Genes" Tab on the right menu of the Variant Table page:

DGI Version: 21-Feb-2023										
Drug name primary	Disease	FDA Approved	Immunotherapy	Antineoplastic	Details					
2-Phenylquinoline (CHEMBL89786)	-	No	No	No	More details					
2X-121 (CHEMBL3644587)	Advanced Solid Tumor, Triple-receptor negative breast cancer	No	No	Yes	More details					
6-Hydroxyflavone (CHEMBL138649)	-	No	No	No	More details					
6-Hydroxyflavone (CHEMBL138649)		No	No	No	More details					
7-Hydroxy Isoflavone (CHEMBL491981)		No	No	No	More details					
7-Hydroxy Isoflavone (CHEMBL491981)	-	No	No	No	More details					
Adavosertib (CHEMBL1976040)	Ovarian cancer	No	No	Yes	More details					
Apigenin (CHEMBL28)		No	No	No	More details					
Aurintricarboxylic Acid (CHEMBL275938)	-	No	No	No	More details					
Berzosertib (CHEMBL3989870)	Ovarian cancer, Triple-receptor negative breast cancer	No	No	No	More details					
✓ Show 90 more										

VarSome Clinical retrieves information from DGIdb and makes it available in the Report. By



adding the DGIdb section to your Report, the information will be shown in tabular form, as below. This table holds aggregated information from DGIdb on the interaction of the respective gene with different drug substances and the various attributes of the interaction.

PAH related drugs		∎ ⊕
Drug name	Attributes	Disease
Chembi117168 FDA Approved: No		
Sapropterin Interaction type: activator, cofactor FDA Approved: Yes	Details Of The Assay For Interaction: Binding affinity for human PAH. Specific Action Of The Ligand: Activation Endogenous Drug: False Direct Interaction: True Trial Name: - Novel Drug Target: Established target Trial Name: Sapropterin	
Sapropterin Dihydrochloride Interaction type: activator FDA Approved: Yes	Mechanism Of Interaction: Phenylalanine-4-hydroxylase activator Direct Interaction: yes	
Fencionine Interaction type: inhibitor FDA Approved: No	Specific Action Of The Ligand: Inhibition Endogenous Drug: False	
Norepinephrine Interaction type: inhibitor FDA Approved: Yes		
Droxidopa Interaction type: inhibitor FDA Approved: Yes		
Norleucine FDA Approved: No		
L-Phenylalanine FDA Approved: No		්

- **PMKB info (if applicable):** Information about the gene containing a specific variant, retrieved from PMKB database.
- **CIVIC info (if applicable):** Information about mutations shown to predict response to targeted therapy from CIViC database.
- **GHR info (if applicable):** gene information from GHR.
- Clinical Trials info (if applicable): Information about related clinical studies from AACT database. The AACT Database is a publicly available relational database and its purpose is to include all information (protocol and result data elements) about every study registered in ClinicalTrials.gov. VarSome Clinical retrieves information from AACT concerning clinical trials related to the gene that contains the selected variant. The information, if available, is displayed under the "Gene" option on the bottom of the variant table in the Variant Table results page, as shown in the screenshot below.

The card:

*								
Variant O Gene BRCA1	\$ 0							
Gene basic Info Gene function Region Browser Structural Variants Transcripts Publications Known gene variants Gene Expression JAX CKB CIVIC PMKB dbNSFP								
GnomAD Genes EBI Gene2Phenotype GenCC ClinGen Disease Validity DOMINO PanelApp gene lists Clinical Generation Disease Validity Human Protein Atlas								
Fusion GDB Community Contributions Cancer Gene Census Pharm GKB FDA DGI CPIC AACT Clinical Trials Protein Viewer OMIM ® LOVD								
AACT Clinical Trials								

opens a table that holds an overview (protocol title, trial status, research sites, supporting



references, etc.) of the clinical trials related to the gene, providing to the user direct and complete insight into the interest around the gene in the clinical research field.

AACT Clinical Trials	Version: 10-Mar-2023								
SEARCH FOR STATUS	ARCH FOR STATUS, DATES, CONDITIONS, PHENOTYPES, INTERVENTIONS OR LOCATIONS: Islata, data, conditions, phenotypes, interventions or locations								
Start Date ASC	Completion ASC Date ^	Status ASC	Source ASC	Title ASC	Conditions / Phenotypes ASC	Interventions ASC	Locations ASC	Publications ASC	
18-May-2023	18-May-2025	Not yet recruiting	City of Hope Medical Center	Phase III Trial of Ninaparib/Selenium Combination Treatment in Patients With BRCAU2-Wid Type Recurrent Platinum-Resistant Ovarian Cancer	Conditions Adnexal Diseases Carcinoma Carcinoma, Ovarian Epithelial Disease Attributes Endocrine Gland Neoplasms Show 20 more Phenotypes	Biopsy Procedure Undergo needle or core biopsy Biospectimen Collection Procedure Undergo biodx simple collection V Show 3 more	United States, Duarte, California City of Hope Medical Center		
01-Apr-2023	31-Dec-2027	Not yet recruiting	University of California, San Francisco	Combination Therapy of Niraparib and Irinotecan In Cancers With Mutations in DNA Repair Genes	Conditions ATM Gene Mutation BRCA1 Mutation BRCA2 Mutation Metastatic Solid Tumor PALB2 Gene Mutation Phenotypes	Irinotecan Drug Given intravenously (IV) Niraparib Drug Given orally	United States, San Francisco, California University of California, San Francisco	-	
28-Feb-2023	01-May-2027	Not yet recruiting	Dana-Farber Cancer Institute	Pilot Trial of Claparib in Patients With Unresectable or Metastatic Melanoma With Mutations in BRCA1/2 Genes	Conditions Cutaneous Melanoma Melanoma Mucosal Melanoma Neoplasms by Histologic Type V Show 7 more Phenotypes	Olaparib Drug Oral, twice a day, dosage per protocol, per 28 day cycle	United States, Boston, Massachusetts Brigham and Women's Hospital United States, Boston, Massachusetts Dana Farber Cancer Institute		

This information is also available for integration into the Report.

	clinical report	g19)			Dast
+ Contract Area	arganization'h address bur organization'h phone			Varsomecinica	
				4	
Accumulated variant information	role	VP-2_510_1001_www.waraca	Predicted ponder:	formatio(p=5.412a-58)	
• All variants' information	00		Diseases		
Specific Variant Information					
NM_004304.51:3645+27>6	4			47	22
+ Variant information					P
+ Vanant references		Sample VP-3_S10	Loc mem.varscan.snp Repo	ort	
ALK Gene information					Ð.
+ ALK related thogs	nary finding				
A ALC PART OF					
AM 12000E No. 2001ColGin Defilition	condary finding				1
He rough a constrainty of the	ondary mong				
Wanard erformalien					
• Variant references					÷
and the second second second	g related				



11. Visualization

11.1 Browsers

VarSome Clinical offers different browsers to visualize the data. These are the Region Browser, CNV Browser and Sample View Browser. Each browser provides different information. The region browser displays the variant in the context of its genomic "neighborhood", including transcripts and other, nearby variants and their pathogenicity as well as frequencies from GnomAD.

The CNV Browser provides an interactive graph to visualize the CNV call region in all samples of the cohort.

Finally, the Sample View presents a genome region browser, but one that is focused on displaying the variants (SNPs, indels, CNVs) identified in the current sample. This provides a global overview of the variants identified in the current sample.

11.1.1 Browser interface settings

There are four icons on the right hand side which can be used to maximize the browser, save the current position on the browser to use in any of the 3 browsers, and change various display settings as shown in the image below. The reset button will bring the browser back to the original position in case the browser has been moved.

General settings:





the system theme, the theme is adjusted to the one you have selected from the setting background option on your browser.

Gesture control options:

- Hold ctrl and scroll to zoom : Use CTRL + SCROLL to zoom in/out
- Click to activate scroll zoom : Use SCROLL to zoom in/out

Full screen mode: You have the option to select the size of the genome browser window.

- Native full screen window
- Expand to current window size

<u>Zoom speed:</u> The option to adjust the speed while zooming in/out inside the genome browser window (default = 1x).

Color scales:

You can choose the colors/color scales for every track in the different browsers, e.g. Conservation and Pathogenicity on the genome region browser and sample view. Click on the drop-down menu of the track that you want to change, and select one of the predefined palette options.

Genome Browser color scales:

General Color	Scales	×
Conservation	Red / Green	•••••
Pathogenicity	Turbo	
Frequency DBs	Cividis	••••••
Germline DBs	Cube Helix	•••••
Somatic DRs	Spectral	•••••
Somatic DBS	Inferno	••••••
Other DBs	Viridis	••••••
	Cool Rainbow	••••••
Restore default color sc	Warm Rainbow	••••••
	Gray Tone	•••••
	Orange / Red	•••••
	 Red / Green 	••••••

CNV Browser color scales:



General	Color Scales	×
Reads	Red / Blue	••••••
Coverage	Red / Blue	••••••
CNV Calls	Red / Blue	••••••

Restore default color scales

Sample View color scales:

General	Color Scales	\times
Conservation	Red / Green	
Pathogenicity	Red / Green	

Restore default color scales

Search for a position or a gene

You can use the search box on the top right of the browser to search for a position in the genome (e.g. chr1) or for a specific gene (e.g. BRCA1). To search for a gene, you need to delete the genomic location displayed and type in the name of your gene of interest.



The saved icon ^{Saved} on the right can be used to bookmark the current position, providing a label that can be used later.



+ Boo	kmark Current Position
Position	Label
chr13:32,889,64532,974,	no label
Type new label for chr13:32,88 BRCA2	9,64532,974,404: Cancel SAVE
chr17:41,196,31241,277,	BRCA1
chr17:7,571,7397,590,807	TP53

Chromosome track

In the chromosome track, the user can switch strands by clicking on the "STRAND" button. STRAND 1 (default) refers to the positive strand (5'-3' direction), while STRAND -1 to the negative one (3'-5' direction). Additionally, the browser offers the option to expand the chromosome-level visualization from the current chromosome to all chromosomes (and *vice versa*), by clicking on the "show all" button on the right-hand side of the bar. Each chromosome can be selected for visualization to inspect the identified variants.



Zoom in and out area

It is possible to zoom in and out on the different tracks of the browser. If the area is marked with a green rectangle this means the focus is on that area and it is possible to use the mouse scroll VarSome Clinical User Manual Version: 11.9.1 - 19th December 2023 Page 201 of 254



to zoom in and out. Otherwise no green rectangle is displayed and there is a need to use Ctrl + scroll to zoom in and out.



Green rectangle

Without green rectangle



11.2 CNV Visualizations

11.2.1 CNV Browser

The CNV Browser can be found under the variant table. It provides an interactive graph toVarSome Clinical User ManualVersion: 11.9.1 - 19th December 2023Page 202 of 254



visualize the CNV call region in all samples of the cohort.

8	0-+	>									s	earch for gene, chr (e.g. ch	n q 🔋 S	÷ ⊺ -	+ ≣- ≛
0	Variant 🔶	Length 👌	Call Quality	Copy Number 👙	Туре	(0	Quality Score	ACMG Class	CNV Rules	Number of genes 👙	Number of exons	Reads expected	Reads observed	Read ratio	Frequency
	chrX:4931966949342980	23.311	× x x	0.959028	deletion	G 🔹	33.8	Pathogenic	Gene	15	30	527	256	0.486	
	chr2:178657652178662391	4.739	× × ×	2.31034	duplication	т	8.32	Pathogenic	Overlap	2	12	512	636	1.24	
	chrX:2018626020186379	119	111	2.50589	duplication	R	6.79	Pathogenic	Overlap	1	1	268	380	1.42	
	chr8:77702467938800	168.554	* * *	2.69599	duplication	D	70.5	Likely Pathogenic	Gene	16	52	1.369	2.221	1.62	•
	chr21:4306775443095782	28.028	× × ×	0	deletion	с	47.7	Likely Pathogenic	Overlap	4	33	242	21	0.0868	•
	chr8:74627277541839	79.112	x x x	2.54597	duplication	D	13.5	Uncertain Significance	-	8	26	289	421	1.46	
	chr10:2949493129495248	317	111	1.44784	deletion	E	5.81	Uncertain Significance	-	2	3	359	245	0.682	
	chr11:2225024622250389	143	111	1.35163	deletion	A	5.8	Uncertain Significance		1	2	199	127	0.638	•
	chr6:123375560123375679	119	111	0.703101	deletion	т	8.28	Uncertain Significance	-	1	1	86	35	0.407	0.00009219
	chr10:4746887147720091	251.220	× × ×	2.55582	duplication	Α	107	Uncertain Significance		17	90	3.759	5.538	1.47	
Showin	ng 1 to 10 of 404 rows 10 -	• rows per p	age									< <u>1</u> 2	3 4 5 6 7	8 9 10) 41 >
🔵 Var	riant Gene							÷							¢ 6
CNV	Details Transcripts CNV	Classification	Pathogenic Pu	blications Known C	NVs Samp	le View	CNV Plot CNV Browse	er							

The position of the CNV call in the reference genome is marked at the top of the CNV Browser and the chromosome name is shown on the left. The user can change the size of the region shown by dragging:

CNV Browser	_		x:49,319,66949,34	2,980 - + < >
strand: 1			show all	>
CNV Browser				
hg19 STRAND: 1				EXPAND TO ALL CHRS
17,350 kb	17,355 kb	17,360 kb	17,365 kb	17,370 kb
+ CNV calls		22.3 kb		
+ Read Ratio				
14				
1				
0.6	* *	•		-
0.4				
+ Coverage				FILTERS T
1K				
100				
1				
0 + Genes		CDMB (introd)		
	\leftrightarrow		~	

The CNV calls track shows if it is a deletion or a duplication. Deletions are represented by an empty rectangle and duplications by a full one. The position in the reference genome and the length of the CNV are shown when the user hovers over the browser with the mouse on the rectangle.

Deletion



CNV Browser								C	-+ +>	* 2
hg19 STRAND: 1		_		1					Ð	XPAND TO ALL CHRS
							chr8: 145,743,514			
145,742.1 kb	145,742.3 kb	145,742.5 kb	145,742.7 kb	145,742.9 kb	145,743.1 kb	145,743.3 kb	145,743.5 kb	145,743.7 kb	145,743.9 kb	145,744.1 k
					1.8 kb				>	
+ CNV calls										
+ Read Ratio										
2										

Duplication

CNV Browser						+ - C	
hg19 STRAND: 1			chr8:145742414-145743189 Type: duplication				EXPAND TO ALL CHRS
145,739.3 kb	145,740.3 kb	145,741.3 kb	Copy Number: 2.29 1 Start: chr8: 145,742,414 End: chr8: 145,743,190	145,743.3 kb	145,744.3 kb	145,745.3 kb	145,746.3 kb
+ CNV calls			Length: 776				>
+ Read Ratio							
2							
1 • • •				·			
0.5							
0							

The region of the CNV call is highlighted in orange. The purple area shows the 95% confidence interval of the read ratio (observed/expected reads). Each colored rectangle represents one of the target regions of the assay used to sequence the sample, and its position on the vertical axis indicates the read ratio for this region. Target regions whose read ratio falls within the confidence interval will be colored blue, while those whose ratio falls outside the expected range will be red. The observed and expected read ratio is shown when the user hovers the mouse over the target region.



Below the read ratio is the coverage track which shows the depth of coverage across both the test sample (the sample currently under analysis) and the control samples used in this analysis. The test sample's coverage is shown in red and the others are in blue. The user can choose to view the coverage on a logarithmic or linear scale under the Filters panel.



CNV Browser							- +	
hg19 STRAND: 1		_	1			10 C 10 C 10 C		EXPAND TO ALL CHRS
				chr8: 14	5,738,078			
45,730.3 kb	145,732.3 kb	145,734.3 kb	145,736.3 kb		145,738.3 kb	145,740.3 kb	145,742.3 kb	145,744.3 kb
				12.4 kb				>
+ CNV calls								
+ Read Ratio								
3								
2.5								
2								
1.5							· · · ·	
1								
0.5								
0								
+ Coverage					COVERAGE: 2125 SAMPLE: 1	2958_17358		FILTERS ^
1K								 Current Sample 2958_17358
100							1	 Reference Samples 2971_17371
10								 Other Samples 2924 17312
1								2927_17315
0 + Genes								2956_17356
								2970_17370
					RECQL4 (exor: 17)			2990_17390 V Axis
								Logarithmic

The user can hover along the CNV and see the coverage for the selected sample, the transcript and its exons under the read ratio track.





Clicking on the Samples icon on the right shows a new window where the user can choose to visualize the coverage of the other samples of the cohort by clicking and holding on the eye icon.



hg19		chr8:145,742	2,658145,743,171 Q	-+ < > []
strand: 1		show all	Samples	×
145.7 M	145.7 M	145.7 M	You can press and HOLD to preview other se	saved started
+ CNV calls	513	,		Setting
			2924_17312	8
Read Ratio			2940_17328 [REF]	0
.4			2970 17370 [REF]	() ()
1			2027 17315 (DEE)	Sample
.8			2321_11313 [REF]	
			2979_17379 [REF]	•
.2			2956_17356	•
0			2958_17358	۲
► Coverage		FILTERS T	2971_17371	۲
			2990_17390	۲
100				
10				
1				
0				

The filters panel allows the user to visualize the CNV call region in the rest of the samples of the CNV analysis by checking the box of the sample name. Once checked, hover over the sample names on the filters and only these will be shown in the Coverage track.





Finally, the exonic structure of any genes the CNV overlaps with is shown in the Genes track under the coverage.





11.2.2 CNV Plot

For CNV cohort analyses, VarSome Clinical provides a CNV plot, showing how the observed read depth in the area of the CNV differs from the expected. The plot can be found on its own tab, under the Variant Table.

9	0- • • * <	>									S	earch for gene, chr (e.g. chr	ବ 👔 ଅ	• 🕶	+ ≡- ≵
□	Variant 👙	Length 🖕	Call Quality	Copy Number 👙	Туре 🖕	C 🕒	Quality Score	ACMG Class	CNV Rules	Number of genes 👙	Number of exons	Reads expected	Reads observed 👙	Read ratio	Frequency
	chrX:4931966949342980	23.311	× × ×	0.959028	deletion	G 🔹	33.8	Pathogenic	Gene	15	30	527	256	0.486	•
	chr2:178657652178662391	4.739	× × ×	2.31034	duplication	т	8.32	Pathogenic	Overlap	2	12	512	636	1.24	-
	chrX:2018626020186379	119	111	2.50589	duplication	R	6.79	Pathogenic	Overlap	1	1	268	380	1.42	-
	chr8:77702467938800	168.554	* * *	2.69599	duplication	D	70.5	Likely Pathogenic	Gene	16	52	1.369	2.221	1.62	•
	chr21:4306775443095782	28.028	× × ×	0	deletion	с	47.7	Likely Pathogenic	Overlap	4	33	242	21	0.0868	-
	chr8:74627277541839	79.112	× × ×	2.54597	duplication	D	13.5	Uncertain Significance	•	8	26	289	421	1.46	-
	chr10:2949493129495248	317	111	1.44784	deletion	E	5.81	Uncertain Significance	-	2	3	359	245	0.682	
	chr11:2225024622250389	143	111	1.35163	deletion	A	5.8	Uncertain Significance	•	1	2	199	127	0.638	-
	chr6:123375560123375679	119	111	0.703101	deletion	т	8.28	Uncertain Significance	-	1	1	86	35	0.407	0.00009219
	chr10:4746887147720091	251.220	× × ×	2.55582	duplication	A	107	Uncertain Significance		17	90	3.759	5.538	1.47	
1 2 3 4 5 6 7 8 9 10 41 >															
© Variant Gene ¢ @															
CNV	Details Transcripts CNV	Classification Pa	athogenic Pu	blications Known C	NVs Samp	e View	CNV Plot CNV Browse	н							

The CNV plots are generated using a modified version of the ExomeDepth tool.

- The gray area indicates the 95% confidence interval of the observed/expected read ratio and the red crosses are specific read depth values at those positions. The genomic location of the CNV is given by the vertical dotted lines.
- The left Y-axis shows the "Observed vs expected read ratio" and the right Y-axis represents the "Copy number". The X-axis shows the chromosome coordinates.
- For short CNVs that encompass a few exons, the plot is displayed in a gene-centered view where the exon numbers and their position along the gene are represented in a horizontal axis above the gene name.
- For large CNVs, A please note that, since the CNV region might encompass several genes, making it impossible to plot all of them, we only show the position of the canonical transcript for each gene. Therefore, if a gene has no canonical transcript (e.g. annotated pseudogene) or if its canonical transcript doesn't overlap with the CNV, you may see genes listed in the Variant Table that are not shown in the CNV plot since their canonical transcript isn't the one that overlaps with the CNV.





Figure: In this example, the observed to expected read depth ratio in the region of the CNV is lower than 1 and below the 95% confidence interval (grey area). These results support the hypothesis of a deletion in the exon 20 of the BRCA1 gene.

11.2.3 Known CNVs

We display only the relevant CNVs for the classification according to the following criteria:

- CNV deletions: we retain those that fully overlap with the given CNV for gnomAD variants. For CNVs coming from clinical sources (Decipher, DBVar, ClinVar CNVs) we use the overlapping CNVs if they are benign and the contained CNVs if they are pathogenic.
- CNV duplications: we keep only the CNVs encompassing the same coding genes. If the CNV is non-coding, then we retain the CNVs that have at least 85% of overlap.



 Variar 	iant Gene			¢ 0
CNV De	Details Transcripts CNV Classification Pathogenic Publications Known CNVs Sample View CNV Plot CNV Browser			
O Back to	to legacy view			Report Issues & Feedback
Structur. hg38	ural Variants		x:49,319,66949,342,980	-+ • •
strand: 1	· · · · · · · · · · · · · · · · · · ·			show all
			chrX: 49.337.793	Saved
К	49.323,5 K 49.328,5 K	49.333,5 K	49.338,5 K	49.343,5 K
	¢ 23,3 K	ь ———		
+ Transc	xorps →>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	■→ <u></u> →→→→	·) · · · · · · · · · · · · · · · · · · ·	FILTERS T Settings
+ UniPro	Prot Protein Regions			
+ Structi	In the second			FILTERS T
	Collapse 138 variants with length over 100.000 bp 😧			

11.2.4 Read alignment visualization in JBrowse

	0- + + 4	>									Search for gene, chr	(e.g. chr1 Q 👕	<i>ଅ</i> ÷ T-	+ ≣• ≛
□	Variant 🔶	Length 🝦	Call Quality	Copy Number 👙	Туре 🔶	Genes 🖕	ACMG Class	CNV Rules	0	🚯 Quality Score 🛛 👙	Number of genes	Number of exons 🛛 🔻	Reads expected 👙	Reads observed
	chr1:724863224202605	223 477 7	x	-1	duplication	A3GALT2,	Pathogenic	Gene Overlap		•	4 905	26 849	0	
0	chr1:22904550226836993	203 932 4	×	-1	deletion	A3GALT2,	Pathogenic	Gene Literature Ove	••		4 282	22 983	0	
	chr2:33141309243062004	209 920 6	x	-1	duplication	AAK1,AAM	Pathogenic	Gene Overlap			3 608	19 354	0	
0	chr2:33141309230579301	197 437 9	×	-1	duplication	AAK1,AAM	Pathogenic	Gene Overlap		•	3 287	17 452	0	
	chr2:18127714181886342	163 758 6	×	-1	duplication	AAK1,ABC	Pathogenic	Gene Overlap			2 812	14 943	0	
	chr1:91164795225090472	133 925 6	x	-1	deletion	ABCA4,AB	Pathogenic	Gene Literature Ove		-	2 718	14 603	0	

Once you have selected a variant in the variant table, you can click on the JBrowse icon to view the alignment of the reads in the regions of the detected CNVs. The CNV called region is highlighted in yellow. The gene and the transcripts are represented above the aligned reads.

11.2.5 Browsing through the samples of a CNV analysis

You can search through the samples analyzed under the same CNV/SV analysis by visiting the results page of one and using the arrows to move to the next or previous sample:

	0 • • A 🦘 K	>	_								Search for gene, chr	(e.g. chr1 Q 👕	ଟ ÷ ⊺ -	+ ≣• ≵
	. Variant 🔶	Length 👙	Call Quality	Copy Number 👙	Туре 🝦	Genes 🝦	ACMG Class	CNV Rules	0	Quality Score	Number of genes 👙	Number of exons 💡	Reads expected 👙	Reads observed
	chr1:724863224202605	223 477 7	x	-1	duplication	A3GALT2,	Pathogenic	Gene Overlap			4 905	26 849	0	
0	chr1:22904550226836993	203 932 4	x	-1	deletion	A3GALT2,	Pathogenic	Gene Literature Ov	••		4 282	22 983	0	
	chr2:33141309243062004	209 920 6	x	-1	duplication	AAK1,AAM	Pathogenic	Gene Overlap		•	3 608	19 354	0	

11.3 Sample View

A new "Sample View" card has been added to VarSome Clinical. This presents a genome region browser, but one that is focused on displaying the variants (SNPs, indels, CNVs) identified in the current sample.



Variant Gene	
General Information SNV KRAS(NM_004985.5):c.35G>A p.(Gly12Asp)	PharmGKB Drugs: 1
Region Browser ≘	Publications
Sample View 🚖	Transcripts NM_004985.5 - missense MANE Select
Community Contributions	Expression Data Top: cells_ebv_transformed_lymphoc Tissues: 54

This sample browser incorporates the following features:

- Conservation scores per position
- A track showing the transcript that was used for the Germline Variant Classification of the selected variant
- More filtering options
- Ability to switch strand and expand the visualization to all chromosomes to see all analyses' data if the analyses are small enough (less than 10k variants).
- Touch and gesture support
- Customisable color palette
- 6 track levels: Chromosome Conservation scores Transcripts Runs of Homozygosity Variants in the sample (Single or Multi) CNVs in the sample.



							1:1249,250,621	(+)(
1 2		5 C 6 7			0 0 0 0 0 0			
	chr1: 50 M	chr1: 100 M		hr1: 150 M		chr1: 21	00 M	chr1: 25
4			240.2 Mb					
acrinta								
of Homozygosity in oslo son 1								
1							-	
11.								
ants in oslo_son_1 - [Affected]								
s in oslo_son_1 - [Arretted]								FILTER
// 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1								
	· · · · ·							
	111			=			1.1	
su	b_add_acmg_annotation			-	i i i			
			1					
of Homozygosity in oslo_mother_1								
// I							1 M 1 M 1 M 1 M 1 M 1 M 1 M 1 M 1 M 1 M	
ints in oslo_mother_1								
s in oslo_mother_1								FILTER
77								10 0.010 10 0.001//
// i					1 1 1 01 1		1.1	
// i i i i i i i i i i i i i i i i i i			_	5 i i		1111		
					1.1			
and the second s								
						_	_	
nts in oslo_father_deleted								FILTER
nts in oslo_father_deleted								
nts in oslo_father_deleted in oslo_father_deleted							11	
nts in oslo_father_deleted in oslo_father_deleted								10 10 10 10 10 10 10 10 1
ants in oslo_father_deleted s in oslo_father_deleted								10 1010 101 001
ants in osio_father_deleted								10 1010 101 001
ants in osio_father_deleted								10 0.010 10.1 000
ants in osio_father_deleted					1 1 1 01 1			10 11 01 10 1 10 1 00 11
atcl in osio_father_deleted								10 0 0 0 0 0 1 0 1 000
								10 0 0 0 0 0 1 0 1 000
					1 1 1 DI I			

Conservation scores



The conservation score per chromosome position is displayed with different colors related to pathogenicity, and its value is presented as a bar. To inspect the conservation scores for a smaller region, the user can zoom in from the corresponding button on the upper right of the browser. When hovering over the conservation bars, the chromosome position is displayed inside a black box alongside the conservation score, provided by the PhyloP software.



Chr7: 140,453,136 140,453,130 Chr7:140453136 Plylop 100 way: 9.201 CTED ACMG MANE CANONICAL NM_004333.6 BRAF EXON: 15 CODON: 600, Valine Ser Lys Val Thr Ala

Transcripts



The Transcripts track shows the MANE RefSeq transcript and the transcript selected for the Germline Variant Classification by default. Hovering over the transcript track, the user is able to see useful transcript-related information such as the source, the transcript reference ID, gene name, exon number, number of codons and amino acids.

						SEL	ECTED	ACMG MA	NM_004333.	6 BRAF E	XON: 2	CODON: 67, Lysine			
← • • • • • • • • • • • • • • • • • • •	(←←	· · · ·	~	 («	*			 (<←			 	 	

A user can select which transcripts of interest should be displayed by using the filtering options on the top right of the track. Moreover, a transcript of interest can be selected by clicking on the displayed option of the tooltip, and it will be visible regardless of the applied filters.

+ Transcripts	FILTERS ^
SELECTED ACIDS MARE CANSINGLY, IML CONSIGNATION (SEE CONSIGNATION OF CONSIGNATI OF CONSIGNATIO	Only MANE select 11 Only Canonical 17
NM_0013546012	■ DBs 76 ✓ RefSeq 24
NM_001374244.1 [ERA]	Ensembl [TSL: 1] 11 Ensembl [TSL: not 1] 41
	*

	≯→ →
$ \qquad \qquad$	*
	· · · · ·
NM_001376477.1.1 @72	
NM_001376472.1 10AA	
	(
	>>
	····



Runs of Homozygosity

The Runs of Homozygosity (ROH) are displayed as bars across the genome. Hovering over the bar will display the start, end and length of the ROH.



Variants in sample

The SNVs and indels of the sample are displayed as a lollipop graph with different colors related to pathogenicity. The zygosity state of each variant is presented at the edge of the lollipop display (homozygous variants as filled disks, heterozygous as half-filled), and it corresponds to the one presented in the results table.



The Sample Browser can be used to visualize variants across all component samples of multi-sample analyses (up to 10 samples).



Sample View				C' ()	-+ <> # 2
hg19 STRAND: 1					EXPAND TO ALL CHRS
147,523.3 kb	147,533.3 kb	147,543.3 kb	147,553.3 kb	147,563.3 kb	147,573.3 kb
		51.9 kb			>
+ Variants in oslo_mother_1					FILTERS T
p					
LP					
VUS					
в		ÓÓ	•		
+ CNVs in oslo mother 1					
					The fellow i
+ Variants in oslo, father, 1					
					FILTERO
P					
LP					
vus					
LB					

Hovering over a variant, a tooltip is presented containing information about the Germline Variant Classification class, the triggered rules, coding impact, zygosity, allelic balance etc.

,	chr19:2227078	3 G⇒C			
SELECTED A	ACMG class: ACMG Rules:	e Lik PVS1	ely Pathogen PM2 Supportin	ic g	FILTERS T
MANE CA	Genes: Zygosity:		ZNF257 Heterozygou	IS	
	Allelic Balance: Coverage: Frequency:		0.10344 5 0.0000055438	18 58 35	FILTERS T
	Sample:		20	01	
		0			

By clicking on the variant, a box is presented providing some additional options, such as a link to open the analysis in which the variant has been identified, a zoom-in option and one to highlight this variant. It will also display a table with the sample, coverage, allelic balance and zygosity information.


Single sample:

Germline Rules:	PVS1 PP5	Very Strong PS3	PM2 Supporting	13,03
Coding Impact:			Nonsense	
Genes:			РКР2	
Frequency:		0.00003978	3999999999999995	
Sample	Coverage	Allelic Balance	Zygosity	
	25	0.29	Heterozygous	

Multi sample:

			9 C⇒T	chr5:147555029
	 Benigr 		ication:	Germline Classif
!	BA1 BP4 Strong	C		Germline Rules:
				Coding Impact:
4	SPINK14			Genes:
	700000000000	0.557		Frequency:
	Zygosity	Allelic Balance	Coverage	Sample
		1	67	oslo_son_1
	Homozygous			
	Homozygous Heterozygous	0.326	92	oslo_mother_1

This track can be filtered based on pathogenicity class, zygosity, coding impact as well as CNVs



that are present in any other sample or are present in all other samples or are not present in the current sample and coverage. To see the available options, click on the filters button on the top right of the track.

	FILTERS
✓ □ CNV Overlap In Any Sample	
Oslo_son_1	46
Oslo_mother_1	46
Oslo_father_1	46
Variant In Any Other Sample	
Oslo_mother_1	33
Oslo_father_1	47
✓	
Oslo_mother_1	33
Oslo_father_1	47
✓ □ Variant Not In Sample	
Oslo_mother_1	21
Oslo_father_1	7
✓ ■ Pathogenicity	54
Benign	54
✓ ☐ Zygosity	54
Heterozygous	33
Homozygous	21
✓ □ Coding Impact	7
Null Variant	0
Inframe Indel	0
Missense	3
Synonymous	4
Non-Coding	0
coverage 0 app	y

CNVs in sample

If a CNV sub-analysis has been launched for a sample, an additional track will be displayed for the identified CNVs. Each CNV is coloured based on the pathogenicity classification and either filled (insertions) or empty (deletions).



Hovering over a CNV, a tooltip is displayed containing information about the Germline Variant Classification class, zygosity, type etc.



chr7:100554905	:L51150:DEL	
ACMG class: ACMG Rules: Zygosity: Type: Copy Number: Sample: Passes all quality Fails all quality ch In CNV Analyses: 12/5/2022, 8:51:	Benign Literature B Heterozygous deletion 0 201 checks: false necks: false 10 AM	

By clicking on the CNV, a box is displayed providing a zoom option for it.

chr7:100554905:L51150:DEL Q
ACMG class: • Benign
ACMG Rules: Literature B
Zygosity: Heterozygous
Type: deletion
Copy Number: 0
Sample: 201
Passes all quality checks: false
Fails all quality checks: false
In CNV Analyses:
12/5/2022, 8:51:10 AM

Additionally, this track can be filtered based on pathogenicity class, type and analyses in which the CNVs are identified. To see the available options, click on the filters button on the top right of the track.



FILTERS	^
Pathogenicity	692
Pathogenic	23
Likely Pathogenic	0
	135
 Likely Benign	0
 Benign	534
🗌 Туре	692
Deletion	607
Duplication	85
Analyses	692
12/5/2022, 8:51:10 AM	688
12/5/2022, 12:05:33 PM	4

12. VarSome Clinical Frequently Asked Questions (FAQs)

12.1 3D protein viewer

How to access the protein viewer tool

This tool is accessible from the "Variant/Gene" table at the top of the window. Click on the "Protein Viewer" button to open a new window with the 3D Protein Viewer tool.

From the Variant table

Variant Gene							
General Information Deletion	Region Browser	Sample View	Community Contributions	Publications	Transcripts	Ріатока	erm ⁱ ine Cla
In-Silico Predictors OMIM ®	Deafness Variation	n Database Cl	inGen Expression Data	GWAS Struc	tural Variants	Protein Viewer	Nearby
From the Gene table							
				Ĵ			

Let's use the variant TP53:R175L as an example on the 3D Protein Viewer.

Go to the top right corner of the window to maximize the protein viewer page (1) or to close the window (2) and keep navigating through the VarSome website.



3d Protein Viewer 📀		Documentation
Gene: TP53 View on UniProt View on SwissModel Transcript: ENST000002693054	Select protein structure SwissModel:8xre 12-351 (number of _ 🗸	1 2
		Show All Residues Variants Pahogenic Dickety Pathogenic Likety Pathogenic Dickety Benign Benign C Benign C Current Variant

Please note that if there is no available structure, the "Protein Viewer" icon will be grayed out:

Variant	Gene												
General I	nformation SNV	Region Browser	Sample View	Community Contributions	Public	ations	Transcripts	s Pharn	nGKB	German	Classi	ification Uncert	ain Significance
In-Silico P	Predictors BP4: Be	enign Moderate	OMIM ® Deafn	ess Variation Database	ClinGen	Expres	sion Data	GWAS	Structur	ral Varian	s	Protein Viewer	Nearby Variants

How to display variants on the protein structure

Variants reported from ClinVar, Uniprot and the VarSome Community are mapped onto the protein structure and are colored according to the Germline Variant Classification. Additionally, you will be able to see the variant of interest (the variant whose VarSome page you are currently visiting) in pink. You can decide which variants are shown in the "Show" menu on the top right corner.

• Click on "Variants" to deselect all variants and display the protein structure without highlighting any residue position.





• Click on "current variant" to show the variant you have searched on VarSome.com.

3d Protein Viewer 📀		Documentation		\times
Gene: TP53 View on UniProt View on SwissModel Transcript: ENST00000269305.4	Select protein structure SwissModel:@xre 12-351 (number of _			
	En Alexan	Show All Residue Variants Pathogen Likely Pat Benign C Likely Ber C Benign C Current V	s ic hogenic i Significance nign ariant	D.

• Click on the other variant options to start adding variants to the protein structure. For example, you can select "current variant" and "pathogenic" variants to check whether your variant clusters with the reported pathogenic variants in that gene.





• Click on a residue of the protein structure to show a list of reported variants at that position below the protein structure.

Selected residue: [ARG]175 Chain: M						
ACMG Classification	Codon	Variant	Transcript	HGVS	Review Stars	Source
Pathogenic	175	chr17:7578406 C→T	ENST00000269305.4	R175H		Saphetor PubMedUserEntry
Pathogenic	175	chr17:7578406 C⇒T	ENST00000269305.4	R175H		UniProt
Pathogenic	175	chr17:7578405 C→T	ENST00000269305.4	R175H		Saphetor VarSome Comment
Pathogenic	175	chr17:7578406 C⇒T	ENST00000269305.4	R175H	黄黄立立	ClinVar
Pathogenic	175	chr17:7578406 C⇒A	ENST00000269305.4	R175L	***	ClinVar

How to select a different protein structure

We import protein structures from <u>Swiss-Model</u>. The protein structure shown by default will be:

- 1. The structure that contains the variant of interest, if available.
- 2. The structure with the highest number of pathogenic variants.

Click the arrow next to the protein structure's name to select a different protein structure from the drop-down menu.



3d Protein Viewer 😣		Documentation	3 ×
Gene 7P53 View on UniProt View on SwissModel Transcript, ENST000002603054	Select protein structure BuissModel.tore 12-35: Inumber of		
State State	SwiskAddag5g5_13a0 (number of parbogene vers 50 SwiskAddag5g5_53a (number of parbogene vers 50 SwiskAddag5g5_94-326 (number of parbogene vers 50 SwiskAddag5g5_94-326 (number of parbogene vers 50) SwiskAddag5g5_94-326 (number of parbogene vers 50)	Show All Restore Show Show	enic nificance it

Sometimes the selected protein structure does not contain any data related to the variant you have searched. If this is the case, a message below the structure will be shown to notify the user.

3d Protein Viewer 📀	
Gene: ER882 View on UniProt View on SwissModel Transcript: ENST000002695715	Select protein structure SwissModel7mn8 1-599 (number of 4
This protein structure does not contain any data related to the variant chr17-37876073-C-T.	×

Please note that if you experience issues when using the 3D protein, please try to clearing your <u>browser data</u> (e.g. cookies) before continue navigating. Please, contact us if the issue persists.

12.2 ClinVar

12.2.1 How can I submit to ClinVar?

Steps for the submission to ClinVar

Step 1

Select from the Dashboard or the Analyses page the analysis that you want.

Step 2

Select the variant of interest from the results table and click on the "Germline Classification" component that is available on the Varsome Clinical results analysis page.



Then click on the "Submit to ClinVar" button, which is available inside the Germline Classification component (screenshot below).

×М	ulti sample analy	sis from				(hg19) 🕑 Multi sample analysi	s from 🧿		. © F	+2 m	ore OD	🕇 Filters 🛓 Reg	oort generation	Analysis act	lons 👻
	0- 🖻 = 🦦 🕈	1								Search	for gene, chr (e.g. chr1 oi	ຊ 👔 ຊ	÷ T+	+ ∷.	*
	Variant	Variant Type	Gene Symbol	0	Germline Class	Germline Rules	HGVS	HGVS Protein	HGVS Coding	Transcript Position	Overlapping Genes	Inheritance	In Function	Zygosity	() Ph
	chr1:116247826 T→C 🍋	SNV	CASQ2	•••	Likely pathogenic	PP3 Strong PM2 Supporting PP3	NM_001232.4:c	D309G(p.(Asp30	c.926A>G	exon 9 of 11 positio	CASQ2	AD/AR	8 🗢	00	•
	chr1:25880262 delAAAAA	Deletion (5)	LDLRAP1		Likely benign	•	NM_015627.3:c	p.?	c.89-137_89-133del	intron 1 of 8 positio	LDLRAP1	AR	0 0	00	
	chr1:156109059 T→A	SNV	LMNA		Likely benign	BP4 Moderate PM2 Supporting	NM_170707.4:c	p.?	c.*162T>A	exon 12 of 12 (3'UT	LMNA	AD/AR	8) 80 🖨 63	00	
	chr1:237656125 delAAGG	Deletion (4)	RYR2		Likely benign	BP4 Moderate PM2 Supporting	NM_001035.3:c	p.?	c.1828-86_1828-83del	intron 18 of 104 pos	RYR2	AD	٥	00	
	chr14:19089600_1 insACG	Insertion (3)	-		Likely benign	BP4 Moderate PM2 Supporting								00	
	chr21:44488824 delAA 🎭	Deletion (2)	CBS 🎭	••	Likely benign	•	ENST00000398	p.?	c.210-100_210-99del	intron 3 of 16 positi	CBS,CBSL	AR	<u>6</u>) (a)	00	
	chr1:6819612 delAAAAA	Deletion (5)			Likely benign	RP4 Moderate								00	•
0	chr1:142803275_6 insAAC	Insertion (3)	-		Likely benign	BP4 Moderate				227 bp before trans	ANKRD20A14P,ENSG00		SI 60	00	-
	chr21:9831558_9 insC	Insertion (1)			Likely benign	BP4 Moderate			·	•		•		00	
	chr21:19885165 delAAC	Deletion (3)	+	••	Likely benign	BP4 Moderate	•		÷		·	÷		00	
4 Showin	g 1 to 10 of 245 rows 10	rows per pa	ge			2					(1 2 3	4 5 6 7	8 9 1	0 25	>
💿 Va	iant 🔿 Gene												Current	Annotation	¢ 0
Gene	ral Information SNV Regi	on Browser Sa	Imple View Comm	unity Con	tributions Publication	ns Transcripts PharmGKB	Germline Classifica	tion Likely Pathogen	ic Uniprot Variants	ClinVar Uncertain Signif	icance 🛣 🏠 LOVD	Frequencies Mitol	dap Conservati	on Scores	
In-Si	co Predictors PP3: Pathoge	nic Strong OM			tabase ClinGen t	Expression Data GWAS Str	uctural Variants Pr	otein Viewer Neart	oy Variants (Sample)	inical Population Frequ	encies #Samples Mult	i			
Gen	nline Variant Classification	Version: 11.9.1											Terms of use De	ocumentation	
							Likely Pathogenic Using transcript NM_0	6 points = 6 P	D-0B €		3-		-a Subr Save as manual	nit to ClinVar classification	

Step 3

After Submitting to ClinVar, you will be able to see all submissions you have made, track their status or tackle any potential submission errors.

		Tags About -
٦	Filters	🛓 Profile
01	Q 📋	··· Preferences
÷	🖲 Inheriti	ClinVar Submissions
	AD/AR	Illumina BaseSpace
	AR	Claim Assay Token
	AD/AR	Lock Session
	AD	C Logout
	10	



VarSome ClinVar Submissions									
Submission	Status	Description of the variant	Gene	SCV accesion number	Created	Updated	Actions		
SUB10328618	X Error - 1 issue	13-38182641-T-C	MYD88		6 September 2021 - 23:12:20	6 September 2021 - 23:24:51	× 🕜 🍵		
SUB10326945	X Error - 1 issue	ENST00000396334.3:c.794T>C	MYD88		6 September 2021 - 12:07:11	6 September 2021 - 23:12:20	Q 📋		
SUB10302555	Processed (not published)	NM_004972.4:c.1849G>T	JAK2	SCV001821200	1 September 2021 - 19:37:45	1 September 2021 - 19:55:11	2		
SUB10297032	X Error - 1 issue	NM_004972.4:c.1849G>T	JAK2		1 September 2021 - 0:12:57	1 September 2021 - 19:37:45	Q 📋		
SUB10297062	Processed (not published)	3-178952085-A-T	PIK3CA	SCV001815864	1 September 2021 - 0:32:51	1 September 2021 - 1:25:13	2		
SUB10296832	Processed (not published)	12-11803166-C-T	ETV8	SCV001815863	31 August 2021 - 23:03:43	1 September 2021 - 1:25:12	2		
SUB10296842	Processed (not published)	NM_001113378.2:c.2511_2512insA	FANCI	SCV001815862	31 August 2021 - 23:10:02	1 September 2021 - 1:15:12	2		
SUB10296826	X Error - 1 issue	NM_000051.4:c.3901G>T	ATM		31 August 2021 - 22:58:53	31 August 2021 - 23:05:13	2		
SUB10265109	Processing	NM_000222.3:c.2326G>A			26 August 2021 - 14:01:28	27 August 2021 - 6:00:08	Q 盲		

abmission	Status	Submission errors			×	Updated	Action
	N Erst.Linse	19.00			12.20	8 September 2021 - 23-24-51	2f 1
	M Brox. Linner	Errors			17.15	6 September 2021 - 23 12 20	Q I
	Processed (not published)	Reference sequence 'T' is on the negative	strand. It is expected to be the positive	strand.	37.45	1 September 2021 - 19 55 11	27 B
	M Erec.Linne	10.0			2.57	1 September 2021 - 10-37-45	Q.
101007062	Processed (not published)		PRICA		1 September 2021 - 0.32 51	1 September 2021 - 1.25.13	12° 1
1011100012	Processed (not published)		6718		31 August 2021 - 22 02 43	1 September 2021 - 1.25.52	12f 1
101000442	🔗 Processed (not published)		FRED		31 August 2021 - 23 32-02	1 September 2021 - 1.15.52	12f 1
	M Error-Linson		A756		31 August 2021 - 22 58:53	31 August 2021 - 23 05 13	12f 1
-	D Processing				28 August 2021 - 14 01 28	27 August 2021 - 8:00:08	Q.

inVar Submissions	ClinVar Submission			How to submit on ClinVar P Update API key
	Fill out the form in order to submit to ClinVar			
Some ClinVar Subm	ClinVar processed your submission and Reference sequence 'T' is on the negative	reported the folk strand. It is expe	owing errors. Please fix them and try to submit the for cted to be the positive strand.	m again.
demission 3				
and the second se	Assertion Criteria			
	Assertion criteria refers to documentation of the crite	eria that your orgar	nization uses to classify variants. It can be provided as a datat	base identifier, like a PubMed ID, or a file that is submitted to ClinVar, but not
	both.			
	Method *			
	ACMG Guidelines, 2015			
	A name for your assertion criteria is required.			
	Assertion Method Citation			
	Database *		Citation Id *	Citation Url *
	PubMed	•	24553177	
			Somatic mutations in MYD88 and CXCR4 are determinants of clinical preser overall survival in Waldenstrom macroglobulinemia. PubMed 24553177	The URL for a file that you have already submitted to ClinVar as assertion criteria. 📀
			Treon, S. P. et al. (2014) Blood, volume:123, issue:1	

12.2.2 What is ClinVar class?

Clinvar class is the clinical significance value given by ClinVar database: http://www.ncbi.nlm.nih.gov/clinvar/docs/clinsig/

Guidance for use in ClinVar SCV records

- Benign As recommended by ACMG for variants interpreted for Mendelian disorders.
- Likely benign As recommended by ACMG for variants interpreted for Mendelian disorders.
- Uncertain significance As recommended by ACMG for variants interpreted for Mendelian



disorders.

- Likely pathogenic As recommended by ACMG for variants interpreted for Mendelian disorders.
- **Pathogenic** As recommended by ACMG for variants interpreted for Mendelian disorders.
- **drug response** A general term for a variant that affects a drug response, not a disease. We anticipate adding more specific drug response terms based on a recommendation by CPIC.
- **association** For variants identified in a GWAS study and further interpreted for their clinical significance.
- **risk factor** For variants that are interpreted not to cause a disorder but to increase the risk.
- protective For variants that decrease the risk of a disorder, including infections.
- affects For variants that cause a non-disease phenotype, such as lactose intolerance.
- conflicting data from submitters Only for submissions from a consortium, where groups within the consortium have conflicting interpretations of a variant but provide a single submission to ClinVar.
- **other** If ClinVar does not have the appropriate term for your submission, we ask that you submit "other" as clinical significance and contact us to discuss if there are other terms we should add.
- **not provided** For submissions without an interpretation of clinical significance. The primary goal of ClinVar is to archive reports of clinical significance of variants. Therefore submissions with a clinical significance of "not provided" should be limited to:
 - "literature only" submissions that report a publication about the variant, without interpreting the clinical significance
 - "research" submissions that provide functional significance (e.g. undetectable protein level) but no interpretation of clinical significance
 - "clinical testing" submissions from clinics or physicians that provide additional information about individuals with the variant, such as observed phenotypes, but do not interpret the clinical significance

12.3 Explanation of VCF attributes

A <u>VCF</u> (Variant Calling Format) is a text file format that stores information about genetic variation. Each variant shown in the variant table corresponds to one line in the VCF file. Besides the variant position, reference and alternative alleles, each line contains additional information (attributes) such as different quality measures that can be accessed from VarSome Clinical. To see these, you need to select the variant in the table, then you can click on the VCF icon to display the VCF attributes as shown in the picture below.

The germline and somatic variant calling pipelines use different algorithms for variant calling and, therefore, the VCFs obtained contain different attributes. In the following lines we will describe the VCF attributes for germline and somatic VCFs.





VCF attributes for Chromosome: chr1, Position: 115705205

VCFA	Attribut	es																	
QUAL	FILTER	AC	AF	AN	BaseQRankSum	ClippingRankSum	DP	ExcessHet	FS	MLEAC	MLEAF	MQ	MQRankSum	QD	ReadPosRankSum	SOR	GT	AB	AD
340.77	PASS	1	0.5	2	-1.152	0	51	3.0103	0	1	0.5	60	0	6.82	0.304	0.592	0/1	0.42	29,2
4																			•

The germline and somatic variant calling pipelines use different algorithms for variant calling and, therefore, the VCFs obtained contain different attributes. In the following lines we will describe the VCF attributes for germline and somatic VCFs.

Germline VCF

- <u>QUAL</u>: it is the phred-scaled probability that the site has no variant. This quality value is used in the pre-processing step (before the annotation) to decide which variants have a PASS call status and which have a FAIL.
- <u>FILTER:</u> filters that have been applied to the variant.
- <u>AC:</u> allele count in genotypes, for each ALT (alternative) allele, in the same order as listed
- <u>AF:</u> allele frequency for each ALT allele in the same order as listed.
- <u>AN:</u> total number of alleles in called genotypes.
- <u>BaseQRankSum:</u> a z-score for base qualities of reference and alternative alleles. For example, a BaseQRankSum close to 0 means the reference and alternative alleles have the same base qualities and a BaseQRankSum around 2 means they differ by 2 SDs. A positive value of 2 means that ALT alleles have higher qualities than REF (reference).
- <u>ClippingRankSum</u>: Z-score From Wilcoxon rank sum test of ALT vs. REF number of hard clipped bases.
- <u>DP:</u> approximate read depth; some reads may have been filtered.
- <u>ExcessHet:</u> phred-scaled p-value for exact test of excess heterozygosity.
- <u>FS:</u> Fisher strand. It is a measure of sequencing bias. This measures if one strand is preferred than the other when sequencing. Larger values means larger bias.
- MLEAC: maximum likelihood expectation of AC (Allele counts).
- <u>MLEAF:</u> maximum likelihood expectation of AF (Allele Frequency).
- <u>MQ:</u> mapping quality. Comparison quality value.
- <u>MQRankSum</u>: this is the u-based z-approximation from the Rank Sum Test for mapping qualities. It compares the mapping qualities of the reads supporting the reference allele and the alternate allele.
- <u>QD:</u> QUAL normalized by read-depth (QUAL/DP).
- <u>ReadPosRankSum</u>: this is the u-based z-approximation from the Rank Sum Test for site position within reads. It compares whether the positions of the REF and ALT alleles are different within the reads.
- <u>SOR (StrandOddsRatio)</u>: this is another way to estimate strand bias using a test similar to the symmetric odds ratio test. FS tends to penalize variants that occur at the ends of exons and SOR does not. Reads at the ends of exons tend to only be covered by reads in one direction and FS gives those variants a bad score. SOR will take into account the ratios of reads that cover both alleles.
- <u>GT:</u> genotype. It is encoded as allele values separated by either of / (not phased) or | (phased). The allele values are 0 for the reference allele (what is in the REF field), 1 for the first allele listed in ALT, 2 for the second allele list in ALT and so on.

×



- <u>AD:</u> allelic depths for the REF and ALT alleles in the order listed.
- <u>GQ:</u> phred-scaled probability that the call is incorrect.
- <u>PGT:</u> physical phasing haplotype information, describing how the alternate alleles are phased in relation to one another.
- <u>PID:</u> physical phasing ID information, where each unique ID within a given sample (but not across samples) connects records within a phasing group.
- <u>PL:</u> normalized, phred-scaled likelihoods for genotypes as defined in the VCF specification.
- <u>SAC:</u> number of reads on the forward and reverse strand supporting each allele (including reference).
- <u>MIN_COVERAGE</u>: minimum coverage threshold considered to give the variant a status of PASS.
- <u>MIN QUALITY INDELS</u>: minimum QUAL threshold considered to give the INDEL variant a status of PASS.
- <u>MIN_QUALITY_SNV</u>: minimum QUAL threshold considered to give the variant a status of PASS.

Somatic VCF

- <u>FILTER:</u> filters that have been applied to the variant.
- <u>AS_FilterStatus:</u> filter status for each allele, as assessed by ApplyRecalibration. Note that the VCF filter field will reflect the most lenient/sensitive status across all alleles.
- <u>AS_SB_TABLE</u>: allele-specific forward/reverse read counts for strand bias tests. Includes the reference and alleles separated by a '|'.
- <u>DP:</u> approximate read depth; some reads may have been filtered.
- <u>ECNT</u>: number of events in this haplotype.
- <u>GERMQ</u>: phred-scaled quality that ALT alleles are not germline variants
- <u>MBQ</u>: median base quality.
- <u>MFRL:</u> median fragment length.
- <u>MMQ:</u> median mapping quality.
- <u>MPOS:</u> median distance from end of read.
- <u>POPAF</u>: negative log 10 population allele frequencies of ALT alleles.
- <u>TLOD:</u> log 10 likelihood ratio score of variant existing versus not existing.
- <u>GT:</u> genotype, encoded as allele values separated by either of / (not phased) or | (phased). The allele values are 0 for the reference allele (what is in the REF field), 1 for the first allele listed in ALT, 2 for the second allele list in ALT and so on.
- <u>AD:</u> allelic depths for the REF and ALT alleles in the order listed.
- <u>AF:</u> allele fractions of alternate alleles in the tumor.
- <u>F1R2:</u> count of reads in F1R2 pair orientation supporting each allele.
- <u>F2R1:</u> count of reads in F2R1 pair orientation supporting each allele.
- <u>SB:</u> per-sample component statistics which comprise the Fisher's Exact Test to detect strand bias.
- <u>MIN_COVERAGE</u>: minimum coverage threshold considered to give the variant a status of PASS.
- <u>MIN_QUALITY_INDELS</u>: minimum QUAL threshold considered to give the INDEL variant a status of PASS.
- <u>MIN_QUALITY_SNV</u>: minimum QUAL threshold considered to give the variant a status of PASS.



12.4 How can I filter out artifacts from my samples?

Please, follow the next steps in order to tag and filter out artifacts from the variant table.

1. Tag a variant as a "Common artifact"

First of all you need to identify all variants that are potential artifacts. To do this you can:

Add a custom variant classification: click on the
 "Common artifact":



• If you have a pre-defined list of common artifacts, you can send us your list and we will label these variants as "Common artifacts". Please, find more information on section <u>Custom Classifications and Comments</u>.

Please note that custom classifications and comments are associated to the variant. You only need to classify the variant once and this classification will be kept and shown in case the same variant is found in any other analysis of your VarSome Clinical account.

2. Filter out "Common artifacts"

Click on the **Line** icon above the variant table to select which variants would you like to show:

Varsomeclinical

	+1 more	ØD	T Filters		🛓 Repo	rt gene	ration
Search for gene, chr (e.g. ch	nr1 o	۹ 📋	ខ	•	۲-	+	≣ •
All manually classified				Ĥ	IGVS		н
) All except common artefact) All manually classified exce	pt commor	n artefacts	5	N	M_0049	85.5:c	G1
Clear filter				N	M_0005	46.6:c	R2
Global classifications	Υοι	ur classific	ations	N	M_0005	46.6:c	R1
Benign Likely Benign	🔵 Artil 🛑 Auto	fact omatically	Tagged Lik	N	M_0178	82.3:c	E2
Uncertain Significance	ely Art	efact					
Likely Pathogenic	🔵 blue	ered				(1))
Pathogenic	Cust	tom_Class	ification_1			\smile	
Common Artefact	Cust	tom_Class	ification_2	*			

Click on show "All except common artifacts". This option will exclude the common artifacts from the variant table.

12.5 Why are variants identified in a single strand considered as artifacts, in particular when using amplicon kits?

Please see the following Technical Note from Illumina:



Figure 1: Detection and Differentiation of DNA Damage from Mutation—In panel A (left), cytosine deamination results in a nucleotide change in one strand of a DNA molecule, but does not alter the complementary nucleotide on the opposite strand. Sequencing each strand independently will yield base calls that differ between the 2 strands. In panel B (right), a true DNA mutation results in a nucleotide change in both strands of a DNA molecule. Sequencing each strand independently will yield the same variant call for both strands. The Amplicon - DS somatic variant caller filters false positives and reports confirmed variants that were identified using TruSight Turnor 26.

Each amplicon is the PCR product from one DNA strand, but since DNA is double stranded, a real variant is supposed to be found in both stands:



12.6 Pseudogenes

The list of pseudogenes is obtained from <u>https://www.genenames.org</u>.

For filtering with a pseudogene list you can use <u>dynamic filters</u> (the funnel icon on the left) and select a Gene List filter.

Or you can perform a Gene List analysis, which creates a sub-analysis like an algorithmic filter.

Where to find the gene list for pseudogenes?

e lists and manage list of genes that you can use in your filters.	
Main Lists PanelApp Lists T Create new gene list T Create gene list from phenotypes	2 pseudogene
Pseudogenes 5.2018 (1998)	
pseudopense according to Actemal.data/cene/pno20180518/hgnc_complete_set.20180518.bxt A2MP1 A3CALT2 AACSP1 AADACP1 AARS1P1 ABCA17P ABCB10P1 ABCB10P3 ABC	
• b c •	

Can VarSome Clinical recognize if a variant was identified on the true transcript or the



pseudogene? How can I be sure that the variants called in pseudogene rich regions do not derive from the pseudogene?

That's one of the limitations of NGS sequencing. Pseudogenes are real sequences and are present in the genome. When a read can be perfectly aligned to multiple regions of the genome, as can be the case if a region of a pseudogene is an exact copy of a region of a functioning gene, then what usually happens is that the read is either discarded or marked with a very low score since the aligner cannot know which of the two (or more) matching regions it should be aligned against.

And this isn't a problem that is specific to pseudogenes. There are multiple, real and functioning genes that are very similar to one another. SMN1 and SMN2, for example. One way of dealing with such duplicate sequences is to mask all such regions and only keep one of them unmasked. Please see <u>Ebbert et al</u> for more details. There is no good solution to this problem, it isn't a limitation of VarSome Clinical, but a direct result of how NGS sequencing works (see, for instance the following two links: <u>https://www.fulgentgenetics.com/Neurofibromatosis</u> and <u>https://blueprintgenetics.com/pseudogene/</u>).

12.7 Variant Sorting

We sort by pathogenicity, and the most pathogenic variants are usually coding. In the rare cases where that isn't the case, then an intronic variant that is more pathogenic than a coding one will be sorted first. The simplest way around this is to use a soft filter and select only coding variants. That doesn't affect the results of the analysis, the user can simply activate and deactivate the filter with a single click and can choose to see only coding or everything each time.

12.8 Tumor Mutational Burden (TMB)

The Tumor Mutational Burden (TMB) or tumor mutation load is considered an effective and clinically verifiable biomarker in multiple cancer types. Unfortunately, because there are many different ways employed by different clinics and labs to measure TMB, and there is no clear standard, VarSome Clinical currently does not offer a calculation of TMB in somatic samples as it would be impossible to provide a TMB metric that could satisfy all of our customers. However, VarSome Clinical offers the necessary tools to get the approximate number of germline and somatic variants and so calculate the TMB using your own definition.

First, you need to know the total number of variants found in your sample. This metric can be found, for example, in the Analyses menu, under the column Variants.

	Analysis	State	Phenotypes	Туре	Genome	Date	Variants	Б
≡ >	Description:	() New		s	hg19	28 Nov 2023	5.126	0

If you are starting your somatic analysis from FASTQ, it is important to bear in mind thatVarSome Clinical reports all variants found in the sample no matter their quality. Therefore, as aVarSome Clinical User ManualVersion: 11.9.1 - 19th December 2023Page 233 of 254



first step to calculate the TMB, we highly recommend you to get the total number of variants found with a Call status of PASS, meaning that you will be counting only those variants passing the quality filters. To do this, please go to "Filters" > "Create" > "Call status", and create a filter to keep only the variants that have passed the quality filters.

v filter set 1 r-defined name	2
ser-defined name	
Call Status Exclude variants that match	Remov
Call status	
Pass for multi-sample analyses, this applies to the first sample	~
Allelic balance	
From To	
for multi-sample analyses, this applies to the first sample	
Coverage	
From	

Once the filter is applied to your sample, you will get the total number of variants found with a PASS call status:



Then, you can filter by allelic balance as an approach to differentiate between germline and somatic variants in your sample. This can be assessed through the "Call Status" filter too. This filter contains the allelic balance field where you can include a specific range of allelic balance to retain and count the number of potential germline variants in your somatic sample.



To do this, please, create a new Call Status filter, selecting the variants with a PASS call status and the allelic balance range for germline variants.

filter set				
defined name				
nline variants				
Call Status				Remov
Exclude variants that match				
Call status				
Pass				~
for multi-sample analyses, this applies to the	first sample			
Allelic balance				
0.4		1.0	2	
for multi-sample analyses, this applies to the	first sample			
Coverage				
From		То		
	C			

Please note that the allelic balance range shown in the screenshot above might not be appropriate in your case. Please provide your own allelic balance range.

Finally, once you have obtained the number of germline variants, you can use this number and the total number of variants previously obtained to calculate the number of somatic variants found in your sample. These metrics can be used to calculate the TMB using your own definition.

12.9 Variant not found

The most common problems related to missing variants are usually twofold. First, at positions with high coverage (e.g. >3000), the Variant Caller will "down-sample" and select a random group of reads to perform the calling on. It is therefore possible that the selection doesn't accurately represent the real data and the variant might be missed. However, in cases where a good percentage (50% or more) of the reads support the variant, most likely the actual problem is coming from the noise in the region. As shown in the IGV screenshot example below, there are multiple mismatches in the region around the variant (green line).



Every color indicates a mismatch and colored reads indicate paired reads whose pair was aligned to a different chromosome. Such "noisy" regions confuse the Variant Callers because when calling, they don't only look at the target position but attempt to call by region haplotype. So they use the region around the variant and build a haplotype based on the region and can miss the variant of interest. Another such example could be the presence of segmental duplications, which makes each read align almost perfectly to separate locations in the genome, leading to low alignment scores.

12.10 Investigating exon coverage

There's no direct way of seeing gaps in an exon's coverage, but you can get part way there by using the Coding coverage report as opposed to the <u>Region list coverage report</u>.



Clicking on that option will bring up a popup where you can choose a gene list and then download a coverage report for the genes in that list. Please note that you will need to create the gene list before clicking on this option. To do so, click on Gene Lists in the top bar:



Then, click on <u>Create new gene list</u> and enter your genes of interest and save it with a name of your choosing. You will now be able to generate coding coverage reports using that list. This file has a breakdown of coverage by exon. Although it won't directly pinpoint gaps, it will at least show the minimum coverage per exon, so if that is 0 you know there's a gap somewhere and can then use the Gene Coverage function to locate the target exon:



Subsequently, click on the exon and so open the JBrowse window that shows the actual reads at that position:



80,000,000	100,000,000	120,000,000	140,000,000	160,000,000	180,000,000
€ →	Q Q Q	chr1 🗸 chr1:944	186372_94487290 (920 b)	100	
25	94,486,750		94.486.875	94.4	87,000
1.1.1.1					1 a a a a a a a a a a a a a a a a a a a
				_	
		exam 35			
		800 3			
		600 - 500 -			
		400			
		100 -			
		i			
	- Auto-				
		4			
		· ·			
	a statistics				

12.11 Publications Timeline Visualization

This interactive viewer allows you to quickly see the number of publications by year, broken down by gene, disease, variants etc. with the ability to then further drill-down and identify specific publications related to the topic you are researching. It is possible to drill-down into sub-categories of terms, for example to see what diseases are associated with a particular gene:

To use the publications Tag Timeline. Click on the link on the top left hand side of the Publications section.

Publications	
Show Timeline	

This opens the new Publications Timeline feature.

Publications							How to link publication	vns?
Show Timeline			 13 publications related to this 	variant.				
			835 publications related to ge	me F11				
Tag Categories:	nna Moda 🔿 Direastar & Phanotomat 🔿	Dava Ograans & Tissuas O Species	Callina Codina Impact C Clinical	Significance O Eurotional Study O	Chamical Compound	Show tags individe	lually Color code by Ta	ag Type
	Diseases & Priendypes	Showing 81 out of 81 un	ique tags. (113 out of 113 data points)	Significance O Punctional Study O	chemical compound			
							filter	
						ble	eding	(8)
						fac	tor XI deficiency	(6)
diseases & phenotypes		•	* :					(6)
						F1	1	(3)
gene	• •	• •	s :		*	FX	I deficiency	(3)
						ble	eding disorder	(3)
clinical significance				•	***	thr	ombosi	(3)
variant						pla	telet disorder	(3)
						au	losomal recessive	(2)
coding impact	•		•	÷		no	nsense	(2)
inheritance mode			:			au	osomal bleeding disorder	(2)
						fac	tor XI	(2)
species				50 C	•	thr	ombotic	(2)
Chemical Compound	•	•	•	•		liko	ly pathogenic	(2)
Eurotional Study						pa	hogenic	(2)
ranoaona otaay						va	rant of uncertain significance	(2)
drug		•		•			certain	(1)
organs & tissues	:					he	mostatic disease	(1)
						un lite	certain significance	(1)
cellline		•				ab	normalities of blood vessel	(1)
	1990	2000		2010	2020	IBI	n	(1)

On the top bar there are a number of tools you can use to amend the data presentation of the viewer.

Tag Categories:							
● All ○ Variant ○ Gene ○ Inheritance Mo	de 🔿 Diseases & Phenotypes 🔿 Drug	⊖ Organs & Tissues ⊖) Species 🔿 Cellline	O Coding Impact	O Clinical Significance	O Functional Study	Chemical Compound

The viewer defaults to showing all tags. It is possible to select only required tag categories by selecting the relevant tag radio buttons.

You can also select to show tags individually. In this way you can clearly visualize the number of publications associated with each tag.



Show Timeline				13 publications related to this v	ariant.			
			\bigcirc	835 publications related to gen	e F11			
Tag Categories:							Show tags individual	y Olor code by Tag 1
All OVariant OGene OInheritance Mod	e 🔿 Diseases & Phenotypes 🔇	Organs & Tissue	s 🔿 Species 🔿 Cellline 📿	Coding Impact O Clinical S	gnificance O Functional Stu	dy 🔿 Chemical Compound		
			Showing 81 of	out of 81 unique tags. (113 out	of 113 data points)			
factor XI deficiency	•		•	•	•	•		
bleeding	• •	•		•		• •		
Finant or uncertain sign							•	
FXI deficiency								
autosomal recessive							•	
nonsense				•		•		
hlanding disorder			•	•	•	•		
somal bleeding disord	-							
thrombosi								
factor XI			•	•				
platelet disorder								
thrombotic likely pethogenic								
pathogenic								
functional assay							•	
uncertain								
hemostatic disease								
iormalities of blood ve								
IBD							· · · · · ·	
erited bleeding disord							•	
disease							1	
geneac aisoraer								
P2RY12								
SLFN14							•	
GFI1B							•	
EZ EZ								
F2								
F8								
MYH9							•	
F13A1								
Mendelian inheritance								
atelet function disorde								
hemorrhagic							•	
thrombocstopenia	1990		2000)		2010	2020	

Hovering over a bubble tells you how many publications are linked



Clicking on a "bubble" will then display the list of associated publications.

Article	Content	Display Options
The two common mutations causing factor XI deficiency in Jews stem from distinct founders: one of ancient Middle Eastern origin and another of more recent European origin.	Previous studies showed that factor XI (FXI) deficiency commonly observed in Ashkenazi J similarly frequent mutations, type [] (Glu117stop) and type [I] (Phe283Leu) with allele freq 0.0254, respectively. In Iraqi Jews, who represent the ancient gene pool of Jews, only the	lews is caused by two uencies of 0.0217 and e type II mutation was
Peretz H et al. 01-Oct-1997 Journal: Blood 42 citations	observed with an allele frequency of 0.0167. In this study we sought founder effects examination of four FXI gene polymorphisms enabling haplotype analysis in affected Jewish Iraai, and other origins and in Anb patients. Initial noculation surveys of 387 Middle Eastern	for each mutation by patients of Ashkenazi, Jews (excluding Iragi
Source: ClinVar	factor XI deficiency II	2 🍆

You can use the "color code by tag type" in the following way:

In the initial setting the "bubbles" reflect the same color as the tags in the publications component. For example a green Disease tag in a publication will be represented by a green disease "bubble" and it is clear which disease is mentioned most often.

You can also switch to using a color scale

In this case the color of the "bubbles" represents a scale of the number of publications that mention a tag overall and is depicted at the bottom of the screen.



The size of the "bubbles" represents the number of publications mentioning a certain tag per year. So the larger the bubble the more publications there are in a given year mentioning the specific tag.

12.12 Joint calling

Implementation

When running a multi-sample analysis in VarSome Clinical, we use joint calling for variants. The variant caller will perform a local realignment of all of the samples' reads around the target region being investigated and then will try to identify the haplotypes supported by the resulting pileup. This way the caller uses the information of all samples to boost the evidence of each variant site.

Use an existing analysis for joint calling

When selecting the option "use an existing analysis", if all of the component analyses were initially launched from FASTQ, joint calling is performed as if re-using the FASTQ files from scratch. The only case where joint calling cannot be implemented is if any of the component analyses were launched starting from a VCF file, for which we only merge the VCF results.

12.13 Variant calling and quality filters

Variant calling is the process by which a software program (the variant caller) identifies variants from sequence data. For single and multi germline samples sequenced using hybridization-based capture kits, we use <u>Sentieon's DNA Scope</u> variant caller and for single and tumor/normal somatic samples, we use the <u>Sentieon TNhaplotyper2</u> algorithm. On the other hand, for amplicon kits on both germline and somatic samples VarSome Clinical uses <u>VarDict</u>.

Depending on the type of assay used to sequence the sample and on the sample itself, VarSome Clinical uses different variant callers.

<u>VarDict</u> is a sensitive variant caller and is especially well-suited to amplicon samples. VarDict implements several novel features such as strand bias aware variant calling from targeted sequencing experiments.

Default workflow

- 1. Adapter trimming is normally done before the FASTQ file is generated. The FASTQ files VarSome gets are therefore already trimmed.
- 2. Read mapping
- 3. Primer clipping (amplicon-based samples)
- 4. Deduplication
- 5. Indel realignment
- 6. Base-quality recalibration
- 7. Variant calling



Call status

There are different quality metrics associated with each variant which can be used in subsequent steps of the pipeline to assign it a call status. The call status of a variant can be:

- PASS: all the quality metrics are above the thresholds (i.e. the variant has passed all quality filters).
- FAIL: the variant has not passed all quality filters.

Quality filters

The quality filters used for germline and somatic analyses are different since we use different variant callers, which, in turn, use different parametrization. The parameters used in Sentieon have been optimized for the detection of variants from the GIAB set. In VarDict we currently use a set of minimally changed default <u>parameters</u>, adjusted through exchange with VarSome Clinical users.

Quality filters for capture kit samples

- Germline analyses

Sentieon's DNAScope

This caller is used for germline capture kit samples and performs an improved version of GATK Haplotype variant calling. We apply the following quality filters after the variant calling step:

- **Coverage**: number of reads aligned against the variant position. The minimum coverage for capture kit samples is 8; all variants with coverage lower than 8 reads will be considered as FAIL.
- **Quality**: the quality score is an internal score calculated by the variant caller algorithm. It can be used to estimate how confident we are that the variant caller has correctly identified a variation in a given genomic position.
- **Single sample analyses**: we assign a FAIL call status to the variants having a QUAL lower than 100 in single sample analyses. The QUAL is the Phred-scaled probability that a REF/ALT polymorphism exists at this site given the sequencing data.
- **Multisample analyses** (couple, family trio or generic multisample): we use the GQ (genotype quality) which represents the Phred-scaled confidence that the genotype assignment (GT) is correct. All variants with a GQ lower than 20 will be marked as FAIL. Please bear in mind that the GQ is associated with each sample. For example, a variant called in a trio analysis will have three different GQs, one per each sample. The variant might have a GQ below the threshold in one of the samples while having a GQ above of it in the other samples. In that case, the variant will be marked as "Failed/Not genotyped" in the sample where it had a low GQ and PASS in the others.

- Somatic analyses

Sentieon's Tnhaplotyper2

Tnhaplotyper2, which is used for somatic capture kit samples, is designed to behave like GATK'sMutect2. Tnhaplotyper2, like mutect2, has associated filtering tools which are applied to theVarSome Clinical User ManualVersion: 11.9.1 - 19th December 2023Page 242 of 254



variants found by the caller. These filters can then be used to decide whether a variant should be marked as PASS or FAIL. If a variant fails any of the filters present in the "FAIL" column of the table below, it will be marked as FAIL. Failing to pass a filter in the "PASS" column will not cause the variant to be marked as FAIL.

PASS	FAIL		
clustered_events	map_qual		
duplicate	base_qual		
fragment	contamination		
multiallelic	weak_evidence		
n_ratio	low_allele_frac		
orientation	normal_artifact		
position	panel_of_normals		
slippage	strand_bias		
haplotype			
germline			
strict_strand			

Somatic VCF filters that do not mark a variant as FAIL:

- clustered_events: multiple events are present on the same haplotype as the variant which is indicative of a false-positive call.
- duplicate: the alternate allele is overrepresented by apparent sequencing duplicates.
- fragment: a large difference is observed in the median fragment length for reads supporting the reference and alternate alleles.
- multiallelic: the mutation occurs at a multiallelic site.
- n_ratio: too many 'N' bases at the site.
- orientation: the variant is likely an artifact due to orientation bias.
- position: the allele is close to the ends of the reads.
- slippage: the variant is likely an artifact due to polymerase slippage.
- haplotype: variant is on the same haplotype as other filtered variants.
- germline: there is evidence that the variant is germline.
- strict_strand: evidence for the alternate allele is not significant in both directions.

Somatic VCF filters that mark a variant as FAIL:



- base_qual: the median base quality of bases supporting the alternate allele is too low.
- contamination: the alternate allele is present due to contamination.
- weak_evidence: the mutation does not have significant support above noise.
- low_allele_frac: the variant allele fraction is below the threshold.
- normal_artifact: the variant is likely an artifact in the normal sample.
- panel_of_normals: the site is present in the panel of normals.
- strand_bias: evidence for the alternate allele comes from only one read direction.

Quality filters for amplicon kit samples

Both somatic and germline amplicon kit samples are analyzed using VarDict with different quality thresholds.

- Germline analyses

The Allelic Balance (AB) cutoff is set to 0.2. However this rule applies only for positions covered by more than 100 reads. Otherwise the variant is reported only if the AB is < 20/coverage depth. (I.e. the call will not be made if the variant is supported by less than 20 reads.)

- Somatic analyses

The Allelic Balance cutoff is 0.005. However this rule applies only for positions covered by more than 400 reads. Otherwise the variant is reported only if the AB is < 20/coverage depth. (I.e. the call will not be made if the variant is supported by less than 20 reads.)

Call status variant filtering

When launching a germline/somatic analysis from FASTQ, the user will have two options:

- **All variants**: the variant table will contain all variants called by the variant caller including both variants with PASS and FAIL call status.

- Variants that pass the quality filters: the variant table will contain only variants having a call status of PASS.

Keep variants (j		
All variants	×	~
All variants		
Variants that pass the quality filters		

Variants can be filtered by its call status using the <u>dynamic filters</u> feature. The "Call Status" filter allows the user to filter variants based on the following criteria:

- Call Status: PASS, FAIL or anything.
- Allelic balance: proportion of reads supporting the alternative allele.
- Coverage: number of aligned reads against the variant position.



Exclude variants that match		
Call status		
Pass		~
for multi-sample analyses, this applies to the first sample		
Allelic balance		
From	То	
for multi-sample analyses, this applies to the first sample		
-		
Coverage		
From	То	
for multi-sample analyses, this applies to the first sample		

12.13.1 Variant calling local reassembly

Sometimes, the variants reported in the variant table are not visible in the alignment displayed through IGV/JBrowse. The reason lies in an intermediate step of the variant calling process called local reassembly.

Variant calling algorithms

As we mentioned above, we use Sentieon variant caller algorithms to perform the variant calling procedure. Sentieon is a toolkit analogous to GATK (The Genome Analysis Toolkit) but built on a highly optimized backend (Kendig et al., 2019). For calling germline variants we use Sentieon DNAscope and for calling somatic variants we use Sentieon TNHaplotyper2.

Sentieon & GATK local reassembly

Sentieon and GATK use a procedure to re-assemble read data and determine candidate haplotypes as a prelude to variant calling. This procedure, named local reassembly, is a middle step where the program first builds an intermediate alignment and the reads are locally aligned. The algorithm calls variants during this middle step and assigns the variants found values such as allele frequency or read-depth.

Please note that sometimes, for some particular genomic regions, the whole sequence alignment does not match with the intermediate alignment constructed during the local reassembly. This is the reason why some variants reported in the variant table are not visible when opening the alignment using IGV or JBrowse.

Purpose of performing local reassembly during the variant calling

Local reassembly based methods are less dependent on prior mapping of sequence reads for variant calling and as a result have higher sensitivity and specificity in indel calling (Rimmer et al., 2014).



Can I access the intermediate alignment?

No, the intermediate alignment file is not available for download or visualization. This file is temporarily created during the variant calling process by the algorithm. This is the reason why it is not displayed together with the other alignment in the results.

Should I consider the variant a true positive or an artifact?

A variant found under these circumstances should be treated like any other variant called by the variant calling pipeline. All of them are candidates of being true positives and other factors should be taken into consideration to evaluate whether a variant is real (e.g. quality of the call, read depth, allele frequency...). Our validated variant calling pipeline has demonstrated to have a high sensitivity and specificity, however, as any other NGS pipeline, there might be false positives. Even if the variant is not a variant calling or mapping artifact, there might be other type of artifacts (sequencing errors) that should not be overlooked.

12.14 CNVs

12.14.1 Length of CNVs

There is no specific minimum size of CNV that can be detected by VarSome Clinical. The length of CNV calls depends on the length of the target genomic intervals of the assay used to sequence the sample. Such intervals are encoded in the form of a .<u>bed</u> file and are assay specific. In whole-exome sequencing (WES) assays, typically each interval is an exon, however assays may also target additional, much shorter regions that are a few bases long. In gene panel assays, the length of regions also varies.

VarSome Clinical's CNV calling pipeline is read depth based, so higher read counts generate more robust CNV signals. On the other hand, low read counts may result in untrustworthy calls. VarSome Clinical checks the coverage and informs the user if it is too low in the call region (Call Quality column). User can also visually inspect the relevant region using the provided links to IGV/JBrowse. Detailed sample coverage analysis is available in the Coverage Report (available in the Analysis Actions menu).

12.14.2 Gene list analysis for CNV calling

In gene list analyzes you can only access by default the results for the targeted genes and not the full parent analysis that is needed for the cohort of a CNV/SV analysis.

The primary reason we don't show the full results when a user has run a gene list analysis and we only show the results of the gene list, is to ensure we don't show incidental findings. In a clinical setting, it is sometimes required that there be no way of seeing any information other than what the target regions include. So, reporting a variant that falls outside the target genes might reveal something that wasn't intended to be known for legal and/or ethical reasons. This is why we don't allow such analyses to be included in CNV calling since that might find CNVs outside the region of interest and can cause issues.



If you know this is not a problem, you can choose to show the variants of the gene list's parent analysis:



Once you have done that, the analyses will be available for CNV calling. To avoid this, you can choose not to run a gene list analysis in the beginning and, instead, run a full analysis and then, once that is finished, run a gene list analysis separately. This will keep both analyses (full and gene list) visible to the user, making the full analysis available for CNV calling by default.

12.14.3 Can I reuse a standard set of control samples to call CNVs?

It is recommended that all samples processed in VarSome Clinical CNV analyses have been sequenced by the same instrument, experimental methodology, in the same batch. However, for practical and cost-related reasons, it may not be possible to repeatedly sequence control samples in every batch, or each time one or more test samples require sequencing. In that case, it is possible to re-use a set of control samples that were sequenced previously. The resulting CNV calls are likely to be less accurate.

In-house benchmarking studies have attempted to quantify this loss in performance using a dataset of 96 samples (66 samples with CNV calls, 30 control). Comparison of a complete CNV run to multiple runs using a fixed limited set of 10 control samples as a pool to select references, showed a single digit drop in sensitivity and a small number of changes in CNV limits/breakpoints.

It is important to always check which samples have been selected as reference in the analysis QC report, and their count, available from Analysis Actions Menu. This is because samples of the reference set should not contain the same CNV as the queried test sample, otherwise such CNV will not be called. Ideally, there should be 5-10 reference samples of high correlation (>0.97 for panels, >0.98 for exomes) for each test sample analyzed. For further detail in assessing CNV run and call quality, please see the sections <u>CNV variant table</u> and <u>Quality Control report for CNV analyses</u>.



12.14.4 How is frequency calculated for CNV calls?

We use the gnomAD database to get the population frequencies for a given CNV. Depending on the type of variant, the frequencies are calculated as follows:

- Deletions: we use gnomAD variants if they fully overlap with the given variants.
- Duplications in coding regions: we compare at the gene level and we use those gnomAD variants that encompass the same coding genes as the given variant.
- Duplications in non-coding regions: we use gnomAD variants if they cover at least 85% of the variant region.

Why are frequencies calculated differently for gains vs losses?

GnomAD reports common structural variants; currently, we retrieve information only for deletions and duplications.

- For deletions, we consider the gnomAD CNVs when they fully encompass a sample's annotated CNV. Even if the gnomAD reported deletion is larger than the sample's CNV, we can assume that the sample's CNV is contained in the gnomAD population. This way, if the gnomAD CNV is reported as a benign loss, then the sample's CNV will be most likely benign as it is contained in that region. Conversely, if the sample's CNV overlaps with a gnomAD pathogenic CNV, then the sample's CNV will be most likely pathogenic as it contains a region whose loss results in pathogenicity.
- For duplications, we differentiate our frequency calculation approach based on the genomic location of a CNV. In case of duplications found at coding regions, we compare the sample's CNVs and gnomAD CNVs at gene level. We will consider that both CNVs may have an equivalent effect only if they encompass the same coding genes.

12.14.5 CNV calling with WES or targeted panel data

VarSome Clinical currently offers Copy Number Variation (CNV) analysis for both Whole Genome Sequencing (WGS) and Whole Exome Sequencing (WES) or targeted panel data. For the non-WGS analyses we use <u>ExomeDepth</u>, a CNV caller based on a read depth approach. To accurately detect CNVs, ExomeDepth requires at least five samples (ideally around ten) that will be run as a cohort with each sample analyzed, using the rest as a pool to select reference samples.

Why are reference samples necessary in CNV calling for non-WGS samples?

The read depth approaches used for CNV calling in WGS usually assume that reads distribute in a more or less uniform way across the genome and, therefore, the differences in read depth are used to identify CNVs. However, this assumption fails in the context of WES and targeted sequencing. One of the main reasons is that the probes used for capturing the different targeted regions have variable specificity and efficiency depending on the region. This fact introduces strong biases in the number of mapped reads per region that hamper the CNV detection. ExomeDepth requires multiple samples because it uses them to control the biases given by the extensive variability in capture efficiency across exons and/or target areas.



What characteristics must the reference samples have?

For optimal results, the reference set of samples must have the following characteristics in common with the test sample (sample of interest):

- Samples should be prepared with the same library protocol and sequenced by the same sequencing platform.
- All samples (the test one and the reference) should all have been generated as part of the same sequencing batch. It is possible to use samples generated in different batches but the resulting CNV calls are likely to be much less accurate.
- Samples should originate from individuals unrelated to each other. For example, if samples come from the same family, related individuals should be excluded.
- For CNV calls in sex chromosomes, all samples should be of the same sex (either all male or all female). If they are not all of the same sex, calls on those chromosomes will not be reliable.

How are reference samples used by ExomeDepth

Each sample given as input for ExomeDepth analysis will be taken to call CNVs on it by using a selection of the remaining as reference samples. This means that, when running a CNV analysis with ExomeDepth, you will get calls for all of the input samples, no matter if you consider the sample as a test or reference.

Another important key to bear in mind is that every input sample might not be compared to all other samples. Each input sample is compared against an optimized set of reference samples that are well correlated with it. The first step of the CNV calling process is to construct the reference set of samples. To do this, ExomeDepth takes one of the input samples and ranks the remaining by order of coverage correlation with the first sample. Then, the remaining samples are sequentially added to the reference set. After the addition of one sample to the reference set, a statistical calculation is performed to see how good the current reference set is to predict CNVs on the test sample. The addition of samples to the reference set stops when it is unable to improve the reference set power to predict CNVs. Therefore, using a high number of samples for CNV analysis does not necessarily increase the accuracy of the results because:

- Not all available samples are included in the reference set, which means that not all the samples are used as reference for calling CNVs in the test sample.
- Some of the CNVs present in the test sample can be missed if they are shared between the test and the reference samples.

The reference selection process is automated in the analytical pipeline implemented in VarSome Clinical and does not require any additional steps by the user. VarSome Clinical SV pipeline will analyze all samples of the cohort successively and generate CNV calls for all.

Please note that ExomeDepth authors estimate that the optimum size of the reference set is ~10. Adding further samples in the reference set actually might decrease the power.

12.15 Why does the position of some indels reported not match the position of the indels reported by ExAC/Gnomad?



The same sequence can be the result of multiple insertions or deletions. In cases of such equivalent indels, one version may have been called by the variant caller but a different one found ExAC or other databases.

When we inspect the read alignments however, the insertion may not be seen in the same place as shown on our system. Therefore, our algorithm carefully and correctly matches indels to the frequency of their equivalent found in Gnomad and other databases.

12.16 Filter by Frequency or Pathogenicity

After an analysis is complete, your results are available at the Dashboard. By clicking on the name of your analysis, you will see the Variant Table:

(P athoge	mic Ont D D TFilters	🛓 Report gen	eration Analysis	actions 🗸
:h for ge	ene, chr (e.g. chr1 c 🔍	1 8 ÷	▼- + ≡	- 🛓
ition	Overlapping Genes 🛛 👙	Inheritance	Function	Zygosit
ositi	RB1	AD	6 9	000
tion	IGF2,INS-IGF2	AD	C 💀 📶 📭	000
itio	SNHG14,UBE3A	AD	C in n: nm ns	000
oosi	ANO10	AR	in sp	000
itio	SNHG14,UBE3A	AD	C in n: nmd fs	000
tion	KCNK9	AD		000
tion	MKRN3,ENSG00000260	AD	37 C 🗓 🚥	000
tion	CDKN1C	AD	C nc ms sy	000
tion	KCNK9	AD	C nc ms	•00

Once you click on the Filters red button, you will get a screen with <u>dynamic filters</u>:



her details			
vailable Filters		New filter set	
✓ ACMG Rules	- A41	Name	
✓ Able Frequencies			
♥ Predicted Pathogenicity Scores	-		
Y ClinVar Class	- MA		
✓ Chromosome and Position	AM		
✓ Pathogenicity class	AM		
✓ Zygosity	AM		
¥ Function	244		
♥ Gene List	200		
¥ Call Status	AM		
¥ Variant Type	AM		
¥ Strand Bias	AM.		
✓Number of Samples	-		
♥R\$ID	-		
✓ COD Inheritance	244		

In order to filter for specific categories (e.g. Allele Frequency and Pathogenicity), you could click on the filter icon and then "Create" and a pop-up screen will appear.

qunecy & Pathogenicity		
✓ Pathogenicity class		Remove
Exclude variants that match		
lass		
5 - Pathogenic		
4 - Likely pathogenic		
3 - Uncertain significance		
2 - Likely benign		
) 1 - Benign		
✓ Allele Frequencies		
		POINTOW
nomAD frequency		
Exclude variants that match		
equency		
Franc	To	

After you have completed a name for the Filter Set you have just created, clicking on "save" will apply the filter set to your results.

12.17 Hemizygous Variants

Hemizygousvariants are the variants that fall within regions without allelic counterparts andVarSome Clinical User ManualVersion: 11.9.1 - 19th December 2023Page 251 of 254



pertaining to a diploid cell with only one copy of a gene instead of the usual two copies.

Are those that are X-linked and are not in the pseudoautosomal regions according to this ref <u>https://en.wikipedia.org/wiki/Pseudoautosomal_region</u>:

• The pseudoautosomal regions get their name because any genes within them (so far at least 29 have been found) [ref] are inherited just like any autosomal genes. PAR1 comprises 2.6 Mbp of the short-arm tips of both X and Y chromosomes in humans and great apes (X and Y are 155 Mbp and 59 Mbp in total). PAR2 is at the tips of the long arms, spanning 320 kbp [ref].

An example of an X-linked variant within a pseudoautosomal region can be found on <u>https://varsome.com/variant/hg19/chrX%3A38145358%3AT%3A</u> (gene RPGR). VarSome Clinical would call the variant above as heterozygous. In this case, VarSome Clinical's Sentieon/GATK variant caller does not distinguish between heterozygous and hemizygous terms. However, since the difference between the terms hemizygous and heterozygous is largely semantic, we feel the variant is still called correctly.

For more Frequently Asked Questions about the platform please see also the following doc file https://docs.varsome.com/en/faq.


13. Troubleshooting

13.1 Clear VarSome Cookies from the browser

The procedure for Firefox is the same as described below. For Microsoft Edge, please see <u>here</u> and for Safari, please see <u>here</u>.

Sometimes there is a need to clear the cookies and the cache from the browser in order to visualize the variant table.

Google Chrome

On the VarSome Clinical webpage, there is a *lock icon* on the left of the URL. Clicking on it displays a drop-down menu. After clicking on Cookies, a new window will be displayed.



It is possible to select which cookies to remove from VarSome.com and/or from VarSome Clinical.

We recommend removing both, and then click done.





Once finished, there is the need to refresh the page which was not possible to visualize the data properly.