

GETTING STARTED WITH SEAVE

This will get you started using Seave, a comprehensive variant filtration platform for clinical genomics, developed at the KCCG.

For a more detailed introduction, you can watch a demonstration video [here](#).

Every case is different, so experiment with different selections and see how they change your results.

If you have any questions or feedback, please contact us at kccgeducation@garvan.org.au.

1. Login

Navigate to Seave at <https://seave.bio/>

Click "Log In" on the top right of the page and enter your username and password (login details for the case studies are in the downloadable instructions). There are also publically available datasets that don't require a login.

2. Go to the data

Click 'Take me to the data' in the middle of the page.

3. Select a database

Scroll to the bottom of the page to see the databases that you have access to.

Click on the relevant database.

4. Select a family

Databases can include multiple individuals, grouped into multiple families.

Select the relevant family and review their affected or unaffected status.

5. Select an analysis type

Select a relevant analysis type.

Multiple analysis types might fit the family's pedigree, meaning you may need to filter the data multiple times.

Connect with KCCG:

Visit us <http://www.garvan.org.au/kccg>

Follow our newsfeed on Twitter @kinghorngenomes

Email us at kccgenquiries@garvan.org.au

The screenshot shows the Seave web application interface. At the top, there are navigation tabs: 'Databases', 'Familial Filters', 'SEAVE research' (selected), 'Data Sources', and 'Log in'. Below the tabs, the main content area is divided into two columns. The left column contains text instructions and a form for selecting a database and family. The right column contains pedigree diagrams and text explaining the filtering mechanisms for different inheritance patterns.

Your database contains **families**.

You can choose to use familial information to conduct variant filtration on members of a single family using predefined analysis methods. Alternatively, you can choose to analyse the entire dataset.

Familial filtering

Click each of the headings below if you would like more information regarding the filtration mechanism and for different example familial scenarios.

Heterozygous dominant

All affected individuals have a heterozygous genotype and all unaffected individuals do not have a heterozygous genotype. Equivalent to autosomal dominant in the autosome and X-linked dominant in the X chromosome.

Homozygous/hemizygous recessive

All affected individuals have a homozygous alternate genotype and all unaffected individuals do not have a homozygous alternate genotype. Equivalent to autosomal recessive in the autosome and X-linked recessive in the X chromosome.

Database selected
NA12878tio.hc.vqsr.decomposed.normalised.vcf.db

Select a family to analyse

Family information
NA12878 (Female) - Affected
NA12891 (Male) - Unaffected
NA12892 (Female) - Unaffected
Please ensure this information is correct before proceeding.

Select an analysis type

Proceed to query options

To start with, try homozygous recessive, compound heterozygous and de novo dominant. If multiple generations are affected, replace de novo dominant with heterozygous dominant.

You can read about inheritance patterns on the right of the page, or by clicking Familial Filters at the top.

Click 'Proceed to query options'.

6. Set up your query

These steps will determine which variants are filtered out, and which are kept for you to manually review.

Remember, you can experiment with these settings.

Genomic locations (optional)

Inclusion: You can enter a location to only return variants in that region, e.g. 'chr1:1-100000' for chromosome 1, bases 1 to 100000, or 'MT' for the mitochondrial genome.

Exclusion: This will exclude all variants in this region, which is a useful way of reducing incidental findings.

Gene lists (optional)

Gene lists are compiled by researchers, clinicians and commercial providers.

Select multiple gene lists by holding command (mac) or control (pc).

You can also enter your own custom gene list.

Impact

Leave 'Impact' at high & medium.

Some variants will have little to no impact, e.g if they code for the same amino acid. Others are likely to be harmful, e.g if they disrupt the transcription of a gene.

Seave follows the Ensembl variant impact prediction, which you can learn more about [here](#).

CADD score

Set 'CADD score' at 10.

The CADD (combined annotation dependent depletion) score predicts a variant's functional significance. Read more about CADD scores [here](#).

Prevalence

Set 'Prevalence' at 1%.

If a variant is found frequently in control databases, it is less likely to be damaging.

Quality

Leave 'Minimum sequencing depth' at 0.

Set 'Minimum variant quality' at 200.

Leave 'Exclude failed variants' selected.

These filters allow you to exclude variants that were not well sequenced. Note: when looking at a family, these filters may exclude variants that were poorly sequenced in only one sample.

Variant type(s)

Leave 'Variant type(s)' at both to include single nucleotide polymorphisms (SNPs) and insertions and deletions (INDELS).

Maximum variants to return

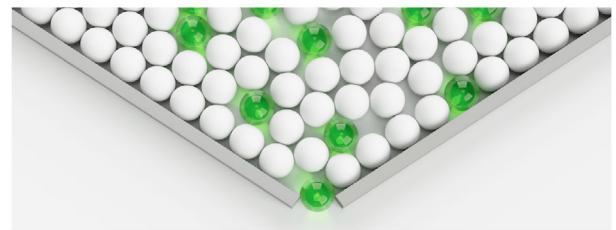
Leave 'Maximum variants to return' at 200.

This gives you a manageable list of variants to review. The maximum may need to be higher if you chose no analysis type in step 5.



OK, you have some data. Now filter it.

Select from the filtration options below.



Database selected NA128781rio_hc.vcpr.decomposed.normalised.vcf.db	Family selected NA128781rio
Inclusion genomic location(s) Search region(s) e.g. chr2:15483-25583;chr1:37211-67824;chr5,MT <small>Separate multiple regions to search with a semicolon. To search all regions, leave this box blank. Any genes specified will be restricted to these coordinates.</small>	Exclusion genomic location(s) Exclude region(s) e.g. chr2:15483-25583;chr1:37211-67824;chr5,MT <small>Separate multiple regions to exclude with a semicolon. To search all regions, leave this box blank. Any genes specified will be restricted to these coordinates.</small>
Search gene list(s) ACMG 56 genes (56) ACMG cancer genes (AD only) (22) ACMG cancer genes (AR + AD) (23) ACMG cancer genes (AR only) (1) Arrhythmic_Syndromes_Aug_2015_Fatkin (4) <input type="button" value="Clear"/>	Exclude gene list(s) ACMG 56 genes (56) ACMG cancer genes (AD only) (22) ACMG cancer genes (AR + AD) (23) ACMG cancer genes (AR only) (1) Arrhythmic_Syndromes_Aug_2015_Fatkin (4) <input type="button" value="Clear"/>
Search custom gene list e.g. BRCA1;PIK3CA;TP53 <small>Separate multiple genes with a semicolon, comma or space. To search all genes, leave this box blank.</small>	Exclude custom gene list e.g. BRCA1;PIK3CA;TP53 <small>Separate multiple genes with a semicolon, comma or space. To not exclude any genes, leave this box blank.</small>
Impact Restrict variants by impact <input type="checkbox"/> Loss of Function <input type="checkbox"/> High impact <input checked="" type="checkbox"/> High & Medium impact <input type="checkbox"/> Coding <input type="checkbox"/> Missense Minimum scaled CADD score <input type="text" value="10"/> <small>All variants without CADD scores are returned. For no minimum scaled CADD score, set this value to 0.</small>	Prevalence Frequency in control databases 1000 Genomes <input type="text" value="1%"/> ESP <input type="text" value="1%"/> EXAC <input type="text" value="1%"/> <small>Variants will be returned that are either below the allele frequency set or not present in the database. For no minimum allele frequency, set the value to 0%.</small> <input type="button" value="Exclude dbSNP Common"/> <input type="button" value="Exclude dbSNP Flagged"/>
Quality Minimum sequencing depth in all samples selected <input type="text" value="0"/> <small>For no minimum sequencing depth, set this value to 0.</small> Minimum variant quality <input type="text" value="200"/> <small>For no minimum variant quality, set this value to 0.</small> <input checked="" type="button" value="Exclude Failed Contigs"/>	Variant type(s) <input type="checkbox"/> SNPs <input type="checkbox"/> INDELS <input checked="" type="checkbox"/> Both Maximum number of variants to return <input type="text" value="200"/>

Connect with KCCG:

Visit us <http://www.garvan.org.au/kccg>

Follow our newsfeed on Twitter @kinghorngenomes

Email us at kccgenquiries@garvan.org.au

KINGHORN
CENTRE
FOR CLINICAL
GENOMICS



7. Get some results

On this screen you will see a table with the filtered variants, ready for you to investigate more closely. There are many approaches to variant analysis, but to get started you could try the following steps.

Sort by the 'Impact Summary' column to see which variants 'light up' with red, meaning multiple tools predict they are damaging.

Scroll down to the 'Functional prediction and conservation scores' menu and **select** CADD Scores (look for scores over 10), SIFT (look for deleterious), and PolyPhen (look for possibly and probably damaging)

Scroll down to the 'Disease phenotypes' menu to see if any databases have disorders associated with the filtered variants. **Select** OMIM to start with, clicking on the OMIM number to get more information.

Scroll down to 'Allele frequencies' to see a variant's frequency in unaffected populations. **Select** EXAC and 1000 Genomes to start with.

The results table

Click on the headings to sort the table by any column.

Variant: location of the variant and details of the change

Quality: quality of the data showing the variant.

Gene: gene the variant is within or affects.

Type: type of variant (insertion, deletion, SNP).

Impact: impact of the variant.

MGRB: presence of the variant in the Medical Genome Reference Bank control cohort.

Impact summary: A summary of the evidence in key databases, hover over a square for more details.

Red = evidence for pathogenicity

Grey = no evidence/entry

White = evidence for non-pathogenicity.

Searching

Search by any of the terms in the table, including gene, genomic location and syndrome.

Additional columns

Scroll down to see further options and click on any of these to bring up additional columns in the table.

For example, selecting OMIM will bring up four columns showing the variant's number, title, status and associated disorders in that database.

Download the data

Click 'Download query results' to download all of the data for closer investigation. This TSV file can be opened in Excel.

8. Edit your search

Click 'Back to query options' to edit your selections if you can't find what you are looking for, or would like to see how different selections change your results.



Great. It's time for some results.

The table below displays your variants. *Click any row* to fetch all GEMINI information for that variant in a separate table.

Variant	Quality	Gene	Type	Impact	KCCG Exomes AF	KCCG Genomes AF	Impact Summary
chr1.g.8666991C>G	548.13004883	CBP4	SNP	missense_variant	0	0	■■■■■
chr1.g.231181171T>A	322.95956518	FRS3B	insertion	splice_region_variant	0	0	■■■■■
chr3.g.18780962C>G	700.13004883	DNAH3	SNP	missense_variant	0	0	■■■■■
chr6.g.110229480C>T	444.13004883	PRKCG3	SNP	missense_variant	0	0	■■■■■
chr5.g.27028484C>A	333.14001468	CD49	Deletion	splice_region_variant	0	0.16 (25/160)	■■■■■
chr6.g.271131242C>A	646.13004883	HCT1132H	Deletion	inframe_deletion	0	0	■■■■■
chr6.g.20859390C>G	3246.85990234	HLA-A	insertion	splice_region_variant	0	0.15 (23/152)	■■■■■
chr6.g.128028848C>T	538.13004883	TTDOK	SNP	missense_variant	0	0	■■■■■
chr6.g.14440282C>G	409.80991509	RP11-42N13.7	insertion	splice_region_variant	0	0	■■■■■
chr6.g.2091769A>T	744.13004883	JAMBCA2	SNP	missense_variant	0	0	■■■■■

Showing 1 to 10 of 33 entries

Previous 1 2 3 4 Next

Download query results (tsv format)

Show or hide specific columns

Click on one or more buttons in each section to dynamically show or hide the columns in the results table above.

Variant and gene information

Variant & Type	Gene & Impact	Variant Quality	Genotypes	Genome Block Score	Variant Allele Frequency
Genotype Quality	HQMS	Transcript Impact	Protein Impact	dbSNP	UniProt
Genomic Location	RefSeq	Navigate in IGV			

Allele frequencies

KCCG Allele Frequencies	1000 Genomes	ESP	ExAC	MITOMAP

Disease phenotypes

Impact Summary	OMIM	Orphanet	ChEMBL	COSMIC	COSMIC Census
MITOMAP					

Functional prediction and conservation scores

CADD Scores	IKMS Percentile	FATHMM	MetaLR	MetaSVM	PolyPhen
GERP++	SIFT	PolyPhen2			

Back to query options

Back to family and analysis selection

Start over

Connect with KCCG:

Visit us <http://www.garvan.org.au/kccg>

Follow our newsfeed on Twitter @kinghorngenomes

Email us at kccgenquiries@garvan.org.au

KINGHORN
CENTRE
FOR CLINICAL
GENOMICS

