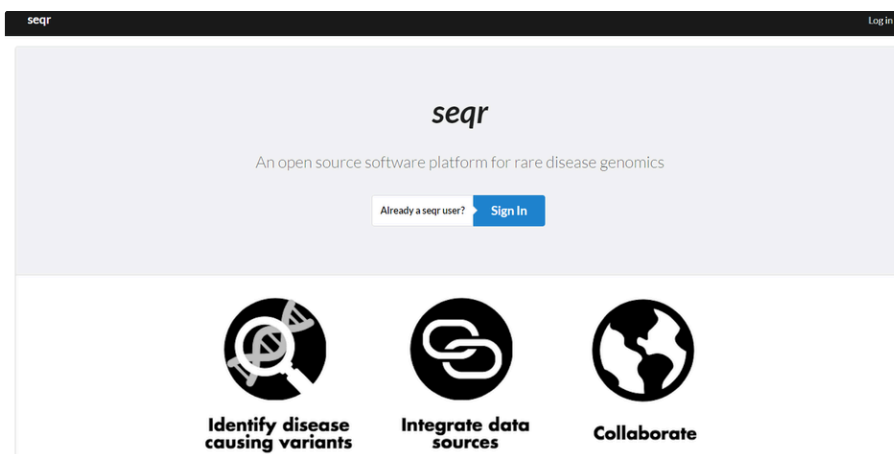


seqr documentation

Seqr is an open-source genomic data interpretation platform. It provides a user-friendly web interface for rare disease diagnosis and gene discovery. It allows uploading VCF (variant call format) files and annotates them with relevant information for Mendelian analysis. The platform supports variant filtering and searching in gene lists to prioritise variants meeting certain criteria.



Uploading VCF file

```
1 #Before executing the commands, ensure that the pipeline-runner container is active
2 #For more information on setting up an on-premises seqr instance, refer to
3 #https://github.com/broadinstitute/seqr/blob/master/deploy/LOCAL_INSTALL.md
4
5 BUILD_VERSION=38           # can be 37 or 38
6 SAMPLE_TYPE=WES           # can be WES or WGS
7 INDEX_NAME=your-dataset-name # the desired index name to output.
8
9 INPUT_FILE_PATH=${FILE_PATH}/${FILENAME}
10
11 docker-compose exec pipeline-runner load_data.sh $BUILD_VERSION $SAMPLE_TYPE $INDEX_NAME $INPUT_FILE_PATH
```

⚠ Make sure your VCF is annotated with fields AC, AN and AF, if not, run bcftools +fill-tags:

```
1 bcftools +fill-tags file.vcf -- -t AC,AN,AF
```

Adding loaded data to a project

You can find loaded data indices under the 'Elasticsearch Status' tab in the 'Data Management' section. Admin rights are required to load data into projects.

The screenshot shows the 'Data Management' section of the seqr interface. The 'Elasticsearch Status' tab is active. It displays two tables: 'Disk Status' and 'Node Status'. Below these is a 'Loaded Indices' table.

Node name	Shards	Disk available	Disk used	Disk %	Heap %
84eb7bb50188	786	6.3tb	1.6tb	20	77

Node name	Heap %
-----------	--------

Index	Project(s)	Caller Type	Data Type	Genome Version	Created Date	# Records	Size	File Path
na12878	Genomikas plelietojums medicfnā	SNV	WES	38	2024-02-29	569118	192.8mb	/input_vcf/wes/k

To add data to specific project, on the start page click on the project you want to load data into:

Name	Last Accessed	Created	Fam.	Indiv.	Samples
Genomikas pielietojums medicīnā seqr-DEMO	12 seconds ago	1 month ago	2	2	2 WES

Click on "Edit Families & Individuals":

Project » Genomikas pielietojums medicīnā

seqr DEMO [Edit Project](#)

Analysis Groups

[Create New Analysis Group](#)

Gene Lists

[Add Gene List](#)

Overview

2 Families, 2 Individuals
2 families with 1 individual

Genome Version
GRCh38

Exome Datasets
29/02/2024 - 1 samples
15/03/2024 - 1 samples

Matchmaker Submissions
No Submissions

Variant Tags
No Saved Variants

[Edit Families & Individuals](#)

[Edit Datasets](#)

[View All](#)

Then, under "Bulk edit Individuals" upload a .csv file that is formatted as in the instructions and press "Submit":

Edit Families & Individuals

[Edit Families](#)
[Edit Individuals](#)
[Bulk Edit Families](#)
[Bulk Edit Individuals](#)
[Bulk Edit Individual Metadata](#)

To bulk-add or edit individuals, upload a table in one of these formats:

Excel (.xlsx) [download blank template](#) or [current individuals](#)
Text (.tsv / .csv / .fam) [download blank template](#) or [current individuals](#)

The table must have a header row with the following column names:

Required Columns:
Family ID
Individual ID

Optional Columns:
Paternal ID Individual ID of the father
Maternal ID Individual ID of the mother
Sex Male or Female, leave blank if unknown
Affected Status Affected or Unaffected, leave blank if unknown
Notes free-text notes related to this individual

If the Family ID and Individual ID in the table match those of an existing individual in the project, the matching individual's data will be updated with values from the table. Otherwise, a new individual will be created. To edit an existing individual's ID include a **Previous Individual ID** column.

Click here to upload a table, or drag-drop it into this box.

[Cancel](#) [Submit](#)

⚠ Make sure that "Family ID" is the same as the index specified when loading the data and can be seen under Data Management tab.

"Individual ID" should be the same as the sample ID specified in the VCF file.

Then, click on "Edit Datasets", enter the index to add data to the project and press "Submit":

Project » Genomikas pielietojums medicīnā

seqr DEMO [Edit Project](#)

Analysis Groups

[Create New Analysis Group](#)

Gene Lists

[Add Gene List](#)

Overview

2 Families, 2 Individuals
2 families with 1 individual

Genome Version
GRCh38

Exome Datasets
29/02/2024 - 1 samples
15/03/2024 - 1 samples

Matchmaker Submissions
No Submissions

Variant Tags
No Saved Variants

[Edit Families & Individuals](#)

[Edit Datasets](#)

[View All](#)

Datasets

Upload New Callset Add IGV Paths

Elasticsearch Index*

Caller Type*

ID Mapping File Path*

ignore extra samples in callset

Adding patient information

On the project page, scroll down to "Families" and find sample by typing index into the search bar:

Families [Download Table](#)

Showing 1 out of 2 families Search Sort By Family Name

Analysis Status	Analysed By	Family Description	Saved Variants
● WES		na12878 (1)	No Saved Variants

Click on the index name and open the sample page. Here, you can add information about the patient's phenotype and associate data analysts.

na12878 [Edit Pedigree Image](#)

■ **DR#s**
ADDED 13/03/2024

- Analysis Groups
- Family Description
- Analysis Status: ● analysis in progress
- Assigned Analyst
- Analysed By: WES/WGS: +, gCNV/SV: +, RHseq: +, Mitochondrial: +, STR: +
- Case Notes [Add Note](#)
- Analysis Notes [Add Note](#)
- Matchmaker Notes [Add Note](#)
- Coded Phenotype
- Post-discovery OMIM #
- Age: Unknown
- Age of Onset
- Individual Notes
- Consanguinity: Unknown
- Other Affected Relatives: Unknown
- Expected Mode of Inheritance
- Assisted Reproduction
- Maternal Ancestry
- Paternal Ancestry
- Features
- Pre-discovery OMIM disorders
- Previously Tested Genes
- Candidate Genes

No Saved Variants

Variant Search

Match-Maker Exchange

● WES ADDED 13/03/2024

Variant search & filtering

To start filtering variants click on "Variant Search" on patient's page:

na12878 [Edit Pedigree Image](#)

■ **DR#s**
ADDED 13/03/2024

- Analysis Groups
- Family Description
- Analysis Status: ● analysis in progress
- Assigned Analyst
- Analysed By: WES/WGS: +, gCNV/SV: +, RHseq: +, Mitochondrial: +, STR: +

No Saved Variants

Variant Search

Match-Maker Exchange

Then, adjust the filters and gene lists before pressing 'Search'.

Project: Genomikas pielietojums medicīnā

Include All Families Families: Analysis Groups:

Add Project: Add Project Group:

Select a Saved Search (Recommended)

Customize Search: Expand All | Collapse All

- Inheritance:
- Pathogenicity:
- Annotations:
- In Silico Filters
- Frequency: AF: AC: HMI: Callset AF: Callset AC:
- Location: Gene List:
- Call Quality:

[Save Search](#) [Search](#)


More on how to apply filters can be found in this video:

[seqr: Variant search](#)

Created by Broad Institute • Updated on Apr 21, 2021

seqr: Variant search In genomic data analysis, search parameters or filters allow users to prioritize the review of variants meeting certain criteria (i.e. variants that are extremely rare in the general population). This tutorial will provide an overview of the different filters that can b...

[YouTube](#) [Open preview](#)



After applying filters you will see the filtered variants:

Showing 1-100 of 315 variants Sort By: Position Download [View Gene Breakdown](#)

Family na12878 Tags: Review Excluded Edit Tag

Notes: [Add Note](#)

NOC2L Decipher | [seqr](#) | [Gene Search](#)

missense variant
HGVS: C_1528A>C
HGVS: P_GAIP510His
1:948519 T > G
hg19:1883899
[seqr](#) | [Gene2HP](#) | [Inframe](#) | [-](#)

- Callset: 25
- Revel: 0.28
- PrimateAI: 0.28
- SpliceAI: 0.0
- Eigen: 4.6
- Polyphen: POSSIBLY DAMAGING
- Sim: TOLERATED
- Mut Taster: DISEASE CAUSING
- Pathway: TOLERATED

This Callset: 0.50 Hom=0 AC=1 out of 2
ExAC: 0.000030 Hom=0
gnomAD-v2 exomes: 0.000036 Hom=0
gnomAD-v3 exomes: 0.0000241 Hom=0
TopMed: 0.00000378 Hom=0

seqr
T/G
17,051

Family na12878 Tags: Review Excluded Edit Tag

Notes: [Add Note](#)

MXRAB Decipher | [seqr](#) | [Gene Search](#)

synonymous variant
HGVS: C_735G>C
HGVS: P_GV91245<*
1:1354896 C > G
hg19:11290276
[seqr](#) | [Gene2HP](#) | [Inframe](#) | [-](#)

- Callset: 65
- SpliceAI: 0.0
- Eigen: 15

This Callset: 0.50 Hom=0 AC=1 out of 2
ExAC: 0.000074 Hom=0
gnomAD-v2 exomes: 0.000083 Hom=0
gnomAD-v3 exomes: 0.000414 Hom=0
TopMed: 0.000155 Hom=0

seqr
C/G
16,856


More on information that can be seen about each variant can be found in this video:

[seqr: Variant information](#)

Created by Broad Institute • Updated on Apr 21, 2021

seqr: Variant information To facilitate efficient evaluation of variant impact, seqr annotates variants called in exome and genome sequencing data with many types of useful information. This tutorial will provide an overview of the information available for each variant returned fro...

[YouTube](#) [Open preview](#)



Adding gene lists

To add a gene list go to Summary Data → Gene Lists → Create New Gene List:

seqr Summary Data Logged in as: test test [Log out](#)

Summary Data Pages: [Gene Info](#) [Gene Lists](#) [Saved Variants](#) [Matchmaker](#)

My Gene Lists Filter:

List Name	Entries	Description	Created Date	Last Updated	Projects Public List
Create New Gene List					

Add genes or intervals, name, description and the genome version:

Create a New Gene List

List Name

Description

Public List Yes No

Genes/Intervals

Genome Version GRCh37 GRCh38

Ignore invalid genes and intervals

More information about seqr:

Broad Institute's official seqr YouTube playlist: [seqr: An open source software platform for rare disease genomics](#)

seqr GitHub page: [GitHub - broadinstitute/seqr: web-based analysis tool for rare disease genomics](#)