

MSeqDR Hands-on Tutorial

Welcome! Today we are going to walk you through a brief hands-on tutorial for MSeqDR. MSeqDR is a centralized genomic resource designed specifically for the mitochondrial disease community. MSeqDR can be utilized as tools for researchers and clinicians with real time queries, data organization, and more... all for FREE to academic users! After this tutorial you will be able to login, navigate within the database, submit, explore, and share data.

MSeqDR has been described in more details in several publications, including a general overview for clinicians published in the March edition of *Mol Genet Metab*: Falk MJ, *et al*. Mitochondrial Disease Sequence Data Resource (MSeqDR): A global grass-roots consortium to facilitate deposition, curation, annotation, and integrated analysis of genomic data for the mitochondrial disease clinical and research communities.

MSeqDR WEB PORTAL: <https://mseqdr.org>

PERKS OF SIGNING UP!

ACCESS ON HOME PAGE; CLICK ON LARGE BOX ICONS

The diagram features a vertical blue bar on the left with the text "USER ACCOUNTS" written vertically. A large blue arrow curves from the top of this bar to the right, pointing towards a stack of four grey boxes. Each box contains a small screenshot of a web interface and a list of features:

- MSeqDR GBrowse**
 - Visualization of variants in both nuclear & mitochondrial (mtDNA) genomes
 - Hosts custom tracks for mitochondrial disease community
- MSeqDR LSDB**
 - Locus specific database for all mitochondrial disease genes and all genes that encode mitochondrial proteins
 - Curates gene, transcript, variant, and disease data relevant to mitochondria
- MSeqDR-GEM.app**
 - Web-based repository and tool to readily enable analysis of sequence data from gene panels, exomes, genomes, and mtDNA genomes
 - Supports analysis of data from individuals, families, or cohorts
- MSeqDR Tools**
 - Centralized host and link to public and custom tools that enable users to perform dataset and variant level analyses in both nuclear & mtDNA genomes
 - Provides support to phenome and ontology tools for mitochondrial disease

I. HOME PAGE

LOGIN: UMDF15

PASSWORD: Mito15

Login <https://mseqdr.org/index.php> or [Register!](#)

***Exercise 1:**

1. Click on “About” - Place cursor over to hover on specific items in tool bar located on the top of the screen AND/OR over the large boxes on the bottom.

2. **Look up data:** click on “TOOLS” menu at the top of the home page or in the large inset box icon. A wide variety of bioinformatics tools and resources are accessible from here: Phy-Mer, mvTool, MToolBox, Exome Prioritization, etc

***Exercise 2:**

1. Enter *SURF1* in the top search box

-How many entries are listed? Scroll down and see what else is there.

2. Enter *MT-ND5* in the top search box

-How many entries are listed? Scroll down and see what else is there.

II. General Purpose Tool - mvTool

Mitochondrial variant annotation engine with inputs using any HGVS nomenclature

<https://mseqdr.org/mv.php>

Quick single variant annotation, naming conversion combining mutalyzer and Ensembl tools (VEP, REST). It utilizes Ensembl VEP, Mutalyzer, supported by our data collection especially population allele frequencies for multiple populations, and lot of variant impact scores, our existing LSDB entry and full ClinVar annotation).

Question/Hint: do you know that the vertebrate mitochondrial code is not the same as the standard code, so that a generic annotation tool may not even be able to annotate mitochondrial variants correctly?

***Exercise 3:** <https://mseqdr.org/mv.php>

1. Enter the full list of same mtDNA variant in multiple existing formats (nuclear-encoded) and click ‘Convert and Annotate’

#Sample input of a mixture of multiple mtDNA variant nomenclatures:

NC_012920.1:m.8993T>G

MT:g.8993T>G

m.8993T>G

8993G

T8993G

8993d

2. Appreciate that the literature uses multiple naming conventions or nomenclatures for mtDNA variants.

3. Check the population frequencies collected by MSeqDR.

4. Is there a dbSNP ID associated with the variant at this position?

5. *Click on NCBI and Ensembl links under ‘Disease and phenotype in-house data, ClinVar and more’*
6. [mvTool for Haplogroup https://mseqdr.org/mv4phyloree.php](https://mseqdr.org/mv4phyloree.php): Select any haplogroup to view the full SNP list and diagnostic SNP list. Quickly run mvTool annotation on them.

***Exercise 4:** <https://mseqdr.org/variant.php>

1. Enter 1:g.215821999G>A (nuclear-encoded) and click ‘ANNOTATE’
2. What is the population frequency?
3. Is there a dbSNP allele associated with this position?
4. *Click on NCBI and Ensembl links under ‘Disease and phenotype in-house data, ClinVar and more’*

III. MSeqDR PhenoTips

Search and explore diseases and human phenotype ontology and associations
Click on Upper right hand corner MSeqDR PhenoTips

Login UMDF15
Password Mito15

Exercise 6

1. Browse all data on previously entered patients for the demo
2. Click ‘create new patient record’
3. Create an identifier
4. Scroll down and click clinical symptoms and physical findings
 - a. Click ‘MSeqDR CHOP-NAMDC core mitochondrial disease symptoms’
 - b. Use the following symptoms to determine a diagnosis for a patient: *sideroblastic anemia, mild global developmental delay, and exocrine pancreatic insufficiency*
5. Click on “you may want to investigate”
6. Click on the tab “Diagnosis”

VI. Phenotype-Guided Exome Variant Prioritization with HPO terms and Exomiser

<https://mseqdr.org/exomiser.php>

*** Exercise 5**

1. Click on files available drop down (Figure 3a) and click FAM_Q1423
2. Input clinical symptoms and diagnosis description as free text, or as HPO terms.
One phenotype per line

3. Click the “HPO Annotator” button twice to find likely matching HPO IDs for the input phenotypes. Check the page bottom to review the returned HPO terms that are ranked according to semantic similarity to HPO dictionary for each input term. Use check box to select most specific matches.
4. The “HPO ID” text box is the final phenotype list.
5. Alternatively, directly type HPO IDs in the “HPO ID” text box, separated by commas.
6. Fill in email address to receive result report and notice.
7. Click the “Annotate” button to submit your job.
8. This page runs Exomiser in batch mode by running all 6 possible combinations (Prioritization method x Inheritance model).
9. The run may need about 6-15 minutes for input with 1500 variants. Use email function to receive result.
 - a. Please limit input size to avoid time-out. Used pre-filtered vcf files with at most ~2000 variants.
10. Review the result from report page. Each combination has 7 links to different result file:
 - a. Original outputs from Exomiser
 - b. Result: HTML, Genes, Variants, Enhanced Annotation, VCF, Log
11. Click on “Enhanced Annotation” link to get the full annotations (genomic, external link, population frequency, dbNSFP) for top 25 candidate variants, provided by MSeqDR VariantOneStop variant annotation tool for both nuclear and mitochondrial variants.

Figure 1a. Phenotype-Guided Exome Variant Annotation and Prioritization

Phenotype-Guided Exome Variant Annotation and Prioritization with HPO and Exomiser

Input clinical symptoms and diagnosis description

Dysphagia
Aminoglycoside-induced hearing loss
Hepatopathy and steatosis
Sensorineural hearing loss

HPO Annotator **HPO ID:** HP:0002015,HP:0008542,HP:0000799,HP:0000 * review at bottom "HPO Annotator"

Explore with alternative: Monarch Text Annotator

Email send me report when finished

Upload Variant File (VCF v4, the sample column is required):

Select Files

Drag & Drop vcf File Here, VCF v4, the sample genotype column is required...

Files available:

1. OGI793_001540.vcf - 18,514,912 KB - 2016-02-19 07:20:22

* This page runs Exomiser in batch mode by running all 6 possible combinations (Prioritise method x Inheritance model). No setting on them needed below.

* The run may need about 15 minutes for input with 1500 variants. Use email function to receive result.

* Please limit input size to avoid time-out. Used pre-filtered vcf files with at most ~2000 variants.

****To keep your data private from public users, please [register](#) and [login](#)**

The Exomiser Settings:

Associated Mendelian disease OMIM ID
e.g. OMIM:252010

PED file for multi-sample VCF files
 No file chosen

Inheritance model

Prioritise genes using:

Minimum variant call quality e.g. 30.0 Maximum minor allele frequency (%)

Contact and feedback:

Interested in joining MSeqDR consortium and collaboration?

Please contact Dr. Marni Falk at falkm@email.chop.edu, Dr. Xiaowu Gai at xgai@chla.usc.edu

Technical questions about the tutorial, data, tools and website?

Please contact Dr. Lishuang Shen lishen@chla.usc.edu , or [feedback](#).

