MSeqDR Hands-on Tutorial

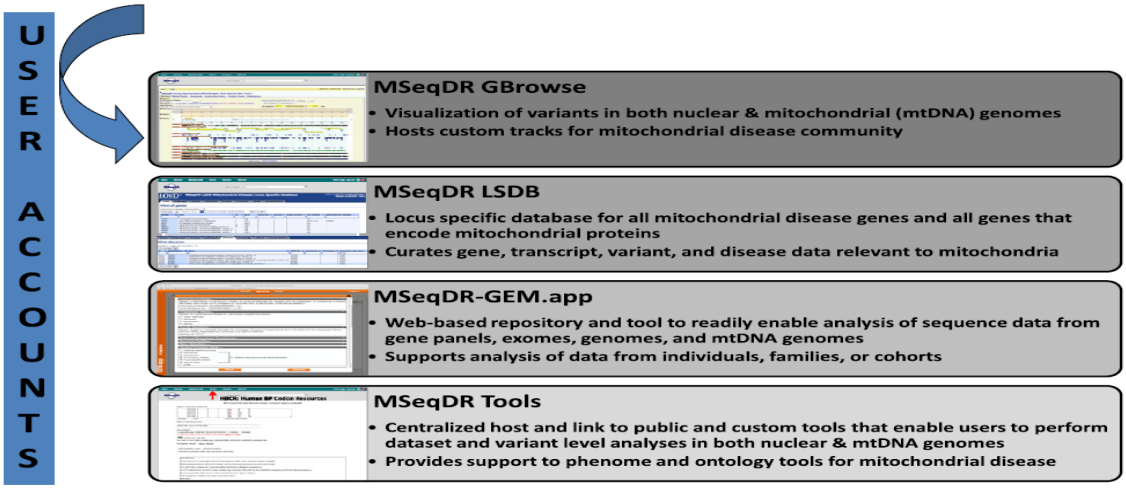
Presentation: Dr. Marni Falk, MSeqDR
Date & Time: Friday March 10th

Welcome! Today we are going to walk you through a brief hands-on tutorial for getting started with using MSeqDR. This database 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 is described in more detail 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



I. HOME PAGE

LOGIN: UMDF
PASSWORD: umdf@mseqdr

Login https://mseqdr.org/index.php or Register**!**

***Exercise 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.

II. LOOK UP DATA

1. Click on "TOOLS" on the home page header or at the large box icon. A wide variety of distinct resources are accessible from here: MitoMap, *POLG* database, MT.AT, HBCR, Transcriptome data, 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.

III. MSeqDR-LSDB, the Locus-Specific Database for Mitochondrial Diseases

a) Our central system to utilize our publicly curated core database that focus on linking diseases, and ultimately will include deidentified patient-level information to causative variants

b) Extensively personalized customization of LOVD (Leiden) system for MSeqDR.

c) Top Menu organizes data at varying levels from Gene-> transcript -> variants.

Mitochondrial disease list page, select "Mitochondrial disease" item from the Top Menu "MSeqDR-LSDB." https://mseqdr.org/MITO/status (production site), or https://mseqdr.org/MITO3/status (development site)

Single disease report pages https://mseqdr.org/mb.php?url=genes.php

a) Explore interesting links in the page (diseases, GeneCard, etc).
b) Disease Associated Genes and Variants
c) Disease Associated ClinVar Variants, full details
d) Disease Associated Phenotypes (Human phenotype ontology, inheritance, frequency, NAMDC key diagnostic phenotypes

Gene list page, select item "Gene View" from the Top Menu "MSeqDR-LSDB" to start work flow
Search/sort by gene, disease associated, sources, mutations

Single Gene pages (https://mseqdr.org/mb.php?url=genes.php)

e) Explore interesting links in the page (diseases, GeneCard, etc).
f) Drilling down to transcript and variants+ blog annotation
g) Data sharing, push variants from current gene to external genome browsers. UCSC is a good example where our variants are shown at top.

***Exercise 3:**

1. **Click on MSeqDR-LSDB and click on transcript variants**
2. **Enter in gene list a mtDNA gene, *ND6***
3. **Click on first entry and scroll down to see all the information and hyperlinks available**
4. **Repeat this process for a nuclear gene *FBXL4***

-> Community blog tool. Try the 3 tab buttons available on variant page for:

h) Quick annotation of existing variant (Add/revise/delete),

i) ClinVar Style full annotation, see how it automatically prefill genomic annotations. And the use of variant real time annotation at the page top.

j) Show/Hide community addon annotations, revise/delete your own annotation.

Example: View genomic variant #0000000006 DB-ID NMNAT1_000002, rs150726175

Chromosome	1
Allele	Unknown
Affects function (reported)	Affects function
Affects function (concluded)	Affects function
Type	-
DNA change (genomic)	g.10042688G>A
Published as	-
GERP	5.010
Segregation	-
DB-ID	NMNAT1_000002
dbSNP ID	rs150726175
Frequency	-
Sources	; clinvar; ensembl;
Reference	22842231;22842229;20301475;22842227;22842230
Variant remarks	-
Genetic origin	-
Automatic mapping	Done (created genes as needed) Map again
Average frequency (large NGS studies)	0.00108 View details
Owner	LOVD
Variant data status	Public
Created by	Lishuang Shen

IV. MSeqDR Genome Browser https://mseqdr.org/gbrowse_bridge.php

a) Users can access and share positional data as community track using gff3 input. Data can be made public, or shared only with specific collaborators.

b) Select tracks to turn on our unique collection of mtDNA data from: Mitomap, HmtDB, Mitobreak, main DB (NCBI, Ensembl, ClinVar).

c) Regional zoom in, export, graphics, track show, hide, download, share to external genome browser.

***Exercise 4**

1. **Click on the gene *POLG* under the search bar**
2. **Click on the first *POLG* hyperlink under "name type description"**
3. **Choose under select tracks "*POLG* mutation database"**
4. **Run curser over different highlighted tracks and icons**

V. General Purpose Tools:

Varian annotation engine for any variant using HGVS input
https://mseqdr.org/variant.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).

***Exercise 5 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'**

***Exercise 6 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 for mtDNA variants.**
3. **Check the population frequencies collected by MSeqDR.**
4. **Is there a dbSNP associated with this position?**
5. ***Click on NCBI and Ensembl* links under 'Disease and phenotype in-house data, ClinVar and more'**

VI. MSeqDR PhenoTips

Search and explore diseases and human phenotype ontology and associations

Click on Upper right hand corner MSeqDRPhenoTips

Login UMDF
Password umdf@mseqdr

Exercise 7

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"**

VII. Phenome and Genomic Data Integrated (OPTIONAL)

Search and explore how the diseases, phenotypes and pathogenic variants annotated in LSDB and 3rd party sources are reported.

Click on Upper menu "Phenome", then click "Disease Portal"
(https://mseqdr.org/diag.php)

Exercise 8

1. **Select a disease from drop-down list, or a common disease under the list.**
2. **Check the annotations and HPO, NAMDC terms associated with that disease**

3. **Check the pathegnic variants at lower right part in MSqDR-LSDB, and ClinVar.**
4. **Click pathogenic variants in bottom and Explore further in LSDB.**

Contact and feedback:

Interested in joining MSeqDR consortium and collaboration?
Please contact Dr. Marni Falk at falkm@email.chop.edu, Dr. Xiaowu Gai at Xiaowu_Gai@meei.harvard.edu

Technical questions about the tutorial, data, tools and website?
Please contact Dr. Lishuang Shen Lishuang_Shen@meei.harvard.edu, or feedback.

ADDITIONAL MSeqDR EXERCISES TO COMPLETE ON YOUR OWN:

1. Data sharing mechanisms:

a) LSDB linkout as real time per-gene variant track, from the LSDB Gene page, try the UCSC Genome browser links:

Graphical displays and utilities	
Graphs	Graphs displaying summary information of all variants in the database »
MSeqDR Genome Browser	Show variants in the MSeqDR Genome Browser (full view, compact view)
UCSC Genome Browser	Show variants in the UCSC Genome Browser (full view, compact view)
Ensembl Genome Browser	Show variants in the Ensembl Genome Browser (full view, compact view)

b) Gbrowse track sharing to collaborators, URL link to gff3 file, or upload a gff3 file to the custom/community section, owner can specify which user can access it.

2. Data mining practice:

a) Search and explore MT-ND1 for LSDB. It links to multiple diseases. And also start with diseases LHON, MELAS will lead to it from non-genomic starting point.

b) Search and explore POLG for LSDB. It links to multiple diseases. It has big community contributed causative variant contribution.

c) Search for disease (OMIM) or phenotype (HPO) , explore the fuzzy-text matched and ranked list that linkto to MSeqDR and OMIM/HPO. https://mseqdr.org/search_phenotype.php?hponame=Leigh%20syndrome&dbsource=OMIM

3. Data Submission: we welcome expert contributions with easy to use submission tools:

The features: quick draft submission by automating genomic data input as much as possible with our backend single variant annotation engine, minimal required items and visit back any time to complete.

a) New single variant submission: minimal typing or no typing annotation (input variant, and click).

b) ClinVar compatible submission online instead of Excel for the 2 ClinVar required sections. **Try Pathogenic variant submission (https://mseqdr.org/ms_variant.php)** , use VCF/HGVS input format, then manage and complete annotations per variant .

4. Data Submission practice:

Single new variant: find a new pathogenic mtDNA mutation, and a new nuclear mutation for mitochondrial disease from recent literature. After the new variant is added, try from its page: Quick Comment blog, full ClinVar annotation. Show hide after comment.

5. Account and Data access practice:
Log in https://mseqdr.org/bblogin.php then "Manage Account and Data Access" for request access, grant access, create your lab/collaborator group and define access rules.