The TDR Targets Database

Prioritizing potential drug targets in complete genomes

Prioritizing targets in whole genomes

- **TDR Targets** facilitates the prioritization of targets in complete genomes by allowing users to search for targets using defined criteria **AND** to assign scores (weight) to these queries.

- Although you can use TDR Targets in the same way you use other genome databases (search, then view), the full potential of the database is exploited by working differently.
  1. Search, Search, Search
  2. Assign scores to each search (weight queries)
  3. Combine all weighted queries to obtain a ranked list of genes

- This is the focus of this tutorial, and in the following slides we will show you how to prioritize a genome in this way.

Prioritizing targets in the genome of *M. tuberculosis*

- In the next slides, we will give you a tour of **TDR Targets**, showing you how you can prioritize potential drug targets in the genome of *Mycobacterium tuberculosis.*

- Remember that because we want to be able to assign different scores to each of our search criteria, we need to separate these criteria in different searches.

Our list of criteria

- Here is a list of criteria we will use to prioritize targets:
  - Target is an **enzyme** (potential for assayability, good druggability precedents)
  - Target has **low molecular weight** and **no transmembrane domains** (higher chances of producing soluble active recombinant proteins
  - Target has either a **known 3D structure** or a **3D model**.
  - Target is **expressed in the dormant stage** (clinically relevant disease phase)
  - Target is **essential**
  - Target is **absent in humans** (or other mammals) and **present in other bacteria**
  - Target has some **precedent for druggability**

Start a new search

TDR Targets Database

Identification and ranking of targets against neglected tropical diseases

Search

Login | Register | Documentation | Contact | FAQ

home | **new search** | history | posted lists of targets | drug targets survey | user manual

1.

1. Select pathogen species of interest

- [ ] *Mycobacterium leprae*
- [ ] *Wolbachia endosymbiont of Brugia malayi*
- [ ] *Schistosoma mansoni*
- [ ] *Plasmodium vivax*
- [ ] *Leishmania major*
- [ ] *Trypanosoma cruzi*

- [x] *Mycobacterium tuberculosis*
- [ ] *Brugia malayi*
- [ ] *Plasmodium falciparum*
- [ ] *Toxoplasma gondii*
- [ ] *Trypanosoma brucei*

All searches begin with the same first step: selecting your species of interest.

Target is an enzyme

Our first search will look for *enzymes* as these are usually good drug targets. For this, we expand the corresponding section, and check the corresponding box for **Functional category.**

2.

Optionally name your queries

Before running your query there is an optional step, in which you can give your query a meaningful name. This is useful so that you can later identify easily each of your queries in the **Query History** page.

3. ➡️

3. Name and run this query

| mtuberculosis enzymes | | [optional]

Run this query   Reset (entire query form)

This section appears at the **bottom of the search page**.

TDR Targets Database
Identification and ranking of targets against neglected tropical diseases   [   ] Search

Login | Register | Documentation | Contact | FAQ

home | new search | history | posted lists of targets | drug targets survey | user manual

☑ **1:** mtuberculosis enzymes, 1790 records. Export | Show parameters | Delete

This is how your query will appear in the **site's History**.

Target has low MW and no TM domains

We have already run our first search. We are not interested in the results for now, so we procceed to our second search. As before we should first specify our species of interest, then our search criteria.

1. →

1. Select pathogen species of interest

☐ Mycobacterium leprae   ☑ Mycobacterium tuberculosis

2. →

Features

☑ ⊞ Search based on target properties and/or features (molecular weight, isoelectric point, length, etc.)

Protein length (AA):	= ▼		in number of residues
Molecular weight:	< ▼	100000	in Daltons
Isoelectric point:	= ▼		
Signal peptide:	— ▼		
GPI Anchor:	— ▼	[presence/absence of glycosylphosphatidylinositol anchor]
# of transmembrane (TM) spans:	= ▼	0	
Number of exons:	= ▼		

[search] [reset]

Target has a 3D structure

In our next query we search for targets that have an **experimental 3D structure**. Note that in the same search section we can also specify that the target should have a 3D model. But because we want to be able to score these two criteria differently, we do separate searches in this case.

1. ➡️

1. Select pathogen species of interest

☐ *Mycobacterium leprae*   ☑ *Mycobacterium tuberculosis*

2. ➡️

Structures

☑ ⊕ Search for targets with associated 3D structures or models.

Retrieve targets with three dimensional data from:

Crystal structures (from PDB):   ☑

OR

Structural models (from Modbase):   ☐

search   reset

Target has a 3D model

All the queries we've been running have been accumulating in the Query History page. Remember to give them meanignful names!

1. →

1. Select pathogen species of interest

☐ Mycobacterium leprae   ☑ Mycobacterium tuberculosis

2. →

Structures

☑ ⊞ Search for targets with associated 3D structures or models.

Retrieve targets with three dimensional data from:

Crystal structures (from PDB):   ☐

OR

Structural models (from Modbase):   ☑

search   reset

Target is expressed in the dormant stage

For queries involving gene expression, we will perform two separate searches, so that we can give different scores to each later. In the first query we will select genes in the top 80-100% expression rank, and in a second query, those in the 60-80% rank.

1. →

1. Select pathogen species of interest

☐ *Mycobacterium leprae*   ☑ *Mycobacterium tuberculosis*

2. →

Expression

☑ ⊕ Search for targets based on evidence on their expression.

| M. tuberculosis | Dormant phase ▼ | murphy ▼ | 80-100% ▼ |
| P. falciparum | Select stage ▼ | Expression level ▼ | |

Expression level
0-20%
20-40%
40-60%
60-80%
80-100%

Notes on available datasets:
murphy: Identification of gene targets against dormant phase Mycobacterium tuberculosis infections. ⊟ →PubMed
hasan: Prioritizing genomic drug targets in pathogens: application to Mycobacterium tuberculosis. ⊟ →PubMed

⊕ More information about this search.

[search] [reset]

Target is essential

1. →

1. Select pathogen species of interest

☐ *Mycobacterium leprae*   ☑ *Mycobacterium tuberculosis*

2. →

Essentiality

☑ ⊞ Search for targets that are essential/inviable.

Retrieve targets for which genome-wide information about their essentiality is available.
If genome-wide information for an organism is not available, you can evaluate the essentiality o
different species from the options below. Also note that essential genes for your organism of int
the **Validation data** search option further down).

☐ **Any** evidence of essentiality in **any** species

Or

Select the species and the type of 'essential' phenotype from the options below: note that if you
genes will be the UNION (boolean OR) of the selection.

C. elegans	— ▼
E. coli	— ▼
M. tuberculosis	essential ▼
S. cerevisiae	— ▼

[search] [reset]

M. tuberculosis is one of the species for which genome-wide knockout data is available.

Target is absent in humans

... and present in other bacteria.

1. → **1. Select pathogen species of interest**

☐ *Mycobacterium leprae*   ☑ *Mycobacterium tuberculosis*

2. → **Phylogenetic distribution**

☑ ⊞ Search for targets based on their phylogenetic distribution.

Restrict to targets with orthologs (present/absent) in:

— ▼	*Drosophila melanogaster*	In ▼	*Escherichia coli*
In ▼	*Mycobacterium leprae*	— ▼	*Mycobacterium tuberculosis*
— ▼	*Wolbachia endosymbiont of Brugia malayi*	— ▼	*Saccharomyces cerevisiae*
— ▼	*Brugia malayi*	— ▼	*Caenorhabditis elegans*
— ▼	*Schistosoma mansoni*	Not in ▼	*Homo sapiens*
Not in ▼	*Mus musculus*	— ▼	*Plasmodium falciparum*
— ▼	*Plasmodium vivax*	— ▼	*Toxoplasma gondii*
— ▼	*Leishmania major*	— ▼	*Trypanosoma brucei*
— ▼	*Trypanosoma cruzi*		

Number of Paralogs: = ▼   [＿＿＿＿]

[search]　[reset]

Target has some precedent for druggability

The **predicted druggability index** (range 0-1) is a combined index with many components, such as: availability of known druggable orthologs, similarity of target vs known druggable targets, structural conservation of binding sites, etc.

1. →

1. Select pathogen species of interest

☐ *Mycobacterium leprae*   ☑ *Mycobacterium tuberculosis*

2. →

Druggability

☑ ⊕ Search for targets based on evidence about their druggability.

Druggability:	> ▾	0.4	Druggability evidence range: 0 to 1]
Compound desirability:	= ▾		[range: 0 to 1]
Associated compounds:	Source of association ▾		

[+/-] More information about this search ...

[search]   [reset]

That should be enough

- In total we have done **10 searches** using different criteria, which we think will allow us to find good targets in the genome of *M. tuberculosis*.

- It's now time to go to the **history page**, add weights to each of these queries and combine them to obtain a ranked list of genes

TDR Targets Database
Identification and ranking of targets against neglected tropical diseases

Search

Login | Register | Documentation | Contact | FAQ

home | new search | history | posted lists of targets | drug targets survey | user manual

Select queries, add weights, calculate UNION

1. → ☑ **1:** mtuberculosis enzymes, 1790 records. Export | Show parameters | Delete

2. → Weight: `50` | Rename: `_____`

☑ **2:** mtuberculosis low MW no TM, 3116 records. Export | Show parameters | Delete

Weight: `50` | Rename: `_____`

☑ **3:** mtuberculosis PDB structures, 229 records. Export | Show parameters | Delete

Weight: `40` | Rename: `_____`

☑ **4:** mtuberculosis modbase models, 2756 records. Export | Show parameters | Delete

Weight: `30` | Rename: `_____`

☑ **5:** mtuberculosis expression murphy 80-100, 799 records. Export | Show parameters | Delete

Weight: `100` | Rename: `_____`

☑ **6:** mtuberculosis expression murphy 60-80, 800 records. Export | Show parameters | Delete

Weight: `80` | Rename: `_____`

☑ **7:** mtuberculosis genes absent from mammals, present in other bacteria, 1330 records. Export |

Weight: `60` | Rename: `_____`

☑ **9:** mtuberculosis essential genes, 802 records. Export | Show parameters | Delete

Weight: `100` | Rename: `_____`

☑ **10:** mtuberculosis druggability > 0.4, 71 records. Export | Show parameters | Delete

Weight: `75` | Rename: `_____`

1. Select query

2. Specify numeric **score/ weight**

3. Combine selected queries

Combine or act on selected queries

◉ **Union:** genes that appear in **any** of the selected queries.

◯ **Intersection:** genes that appear in **all** of the selected lists.
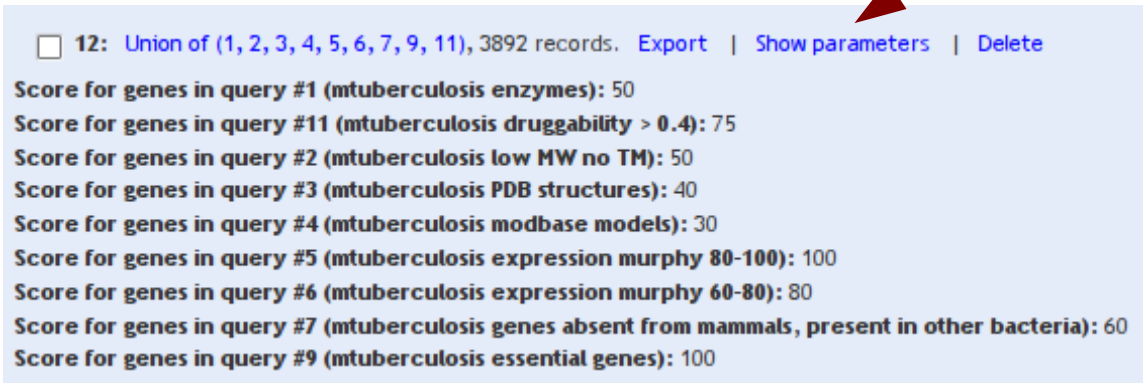
Let's think about this for a second ...

• It is important to understand why you have to calculate the **UNION** of the selected queries to produce a ranked list of genes.

• By calculating the **INTERSECTION** of the selected queries you will get only those genes that are present in **all** queries. And therefore all genes in this final list will have the same score (the sum of all scores). Pretty boring huh?

• When you calculate the **UNION**, tough, any gene that is present in at least one of your queries will end up in the final list. However, genes that are present more than once will have their scores added. **Note also** that the final list will also include genes that were present in all the queries (the same ones you'd get by calculating the INTERSECTION). These genes (if any) will have the maximal score and will be listed at the top.

The ranked list of genes

After calculating the **UNION** of the selected queries, a new query in your **History page** will appear, containing the results of your prioritization.

In this case, the ranked list contains 3892 genes (which is > 95% of the genes in the genome of *M. tuberculosis*).

You can click on **Show parameters** to expand the view and reveal the weighting strategy used (useful if you later decide to delete all previous queries).

☐ **12:** Union of (1, 2, 3, 4, 5, 6, 7, 9, 11), 3892 records. Export | Show parameters | Delete

Score for genes in query #1 (mtuberculosis enzymes): 50
Score for genes in query #11 (mtuberculosis druggability > 0.4): 75
Score for genes in query #2 (mtuberculosis low MW no TM): 50
Score for genes in query #3 (mtuberculosis PDB structures): 40
Score for genes in query #4 (mtuberculosis modbase models): 30
Score for genes in query #5 (mtuberculosis expression murphy 80-100): 100
Score for genes in query #6 (mtuberculosis expression murphy 60-80): 80
Score for genes in query #7 (mtuberculosis genes absent from mammals, present in other bacteria): 60
Score for genes in query #9 (mtuberculosis essential genes): 100

The ranked list of genes

These are the **top targets** in the prioritized list, sorted by descending weight.

Organism	Name	Ortholog group	Product	Source	Weight
M. tuberculosis	Rv1256c	OG1.2_7733	PROBABLE CYTOCHROME P450 130 CYP130	TubercuList	485
M. tuberculosis	Rv0766c	OG1.2_7733	PROBABLE CYTOCHROME P450 123 CYP123	TubercuList	465
M. tuberculosis	Rv2266	OG1.2_7733	Probable cytochrome P450 124 CYP124	TubercuList	465
M. tuberculosis	Rv3545c	OG1.2_7733	PROBABLE CYTOCHROME P450 125 CYP125	TubercuList	465
M. tuberculosis	Rv1094	OG1.2_48054	POSSIBLE ACYL-[ACYL-CARRIER PROTEIN] DESATURASE DESA2 (ACYL-[ACP] DESATURASE) (S ...	TubercuList	430
M. tuberculosis	Rv3581c	OG1.2_4173	PROBABLE 2C-METHYL-D-ERYTHRITOL 2,4-CYCLODIPHOSPHATE SYNTHASE ISPF (MECPS)	TubercuList	430
M. tuberculosis	Rv3710	OG1.2_7356	2-ISOPROPYLMALATE SYNTHASE LEUA (ALPHA-ISOPROPYLMALATE SYNTHASE) (ALPHA-IPM SYNT ...	TubercuList	430
M. tuberculosis	Rv3048c	OG1.2_26217	RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE (BETA CHAIN) NRDF2 (RIBONUCLEOTIDE REDUCTAS ...	TubercuList	430
M. tuberculosis	Rv2225	OG1.2_3086	Probable 3-methyl-2-oxobutanoate hydroxymethyltransferase PanB	TubercuList	430
M. tuberculosis	Rv3014c	OG1.2_4999	PROBABLE DNA LIGASE [NAD DEPENDENT] LIGA (POLYDEOXYRIBONUCLEOTIDE SYNTHASE [NAD+ ...	TubercuList	410
M. tuberculosis	Rv3227	OG1.2_932	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE AROA (5-ENOLPYRUVYLSHIKIMATE-3-PHOS ...	TubercuList	405
M. tuberculosis	Rv2540c	OG1.2_2862	PROBABLE CHORISMATE SYNTHASE AROF (5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE PHOSPHOLYA ...	TubercuList	405
M. tuberculosis	Rv3051c	OG1.2_533	RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE (ALPHA CHAIN) NRDE (RIBONUCLEOTIDE REDUCTAS ...	TubercuList	405
M. tuberculosis	Rv0570	OG1.2_533	PROBABLE RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE (LARGE SUBUNIT) NRDZ (RIBONUCLEOTI ...	TubercuList	405
M. tuberculosis	Rv2552c	OG1.2_4174	PROBABLE SHIKIMATE 5-DEHYDROGENASE AROE (5-DEHYDROSHIKIMATE REDUCTASE)	TubercuList	390
M. tuberculosis	Rv3132c	OG1.2_9875	TWO COMPONENT SENSOR HISTIDINE KINASE DEVS	TubercuList	390
M. tuberculosis	Rv2007c	OG1.2_10530	PROBABLE FERREDOXIN FDXA	TubercuList	390
M. tuberculosis	Rv3398c	OG1.2_2751	PROBABLE MULTIFUNCTIONAL GERANYLGERANYL PYROPHOSPHATE SYNTHETASE IDSA1 (GGPP SYN ...	TubercuList	390

That's all for now

- In this quick tour of the **TDR Targets** database, we showed you how you can use the database as a tool to prioritize targets in a genome, based on a defined set of criteria.

- There are other aspects of **TDR Targets** that we didn't cover in this tutorial. For more quick guides, please head to
  - http://tdrtargets.org/tutorials or
  - http://slideshare.net/tdrtargets

- Some of the slideshows available are:
  - Introduction to the TDR Targets Database
  - Target Surveys in TDR Targets
  - Sharing information with others in TDR Targets