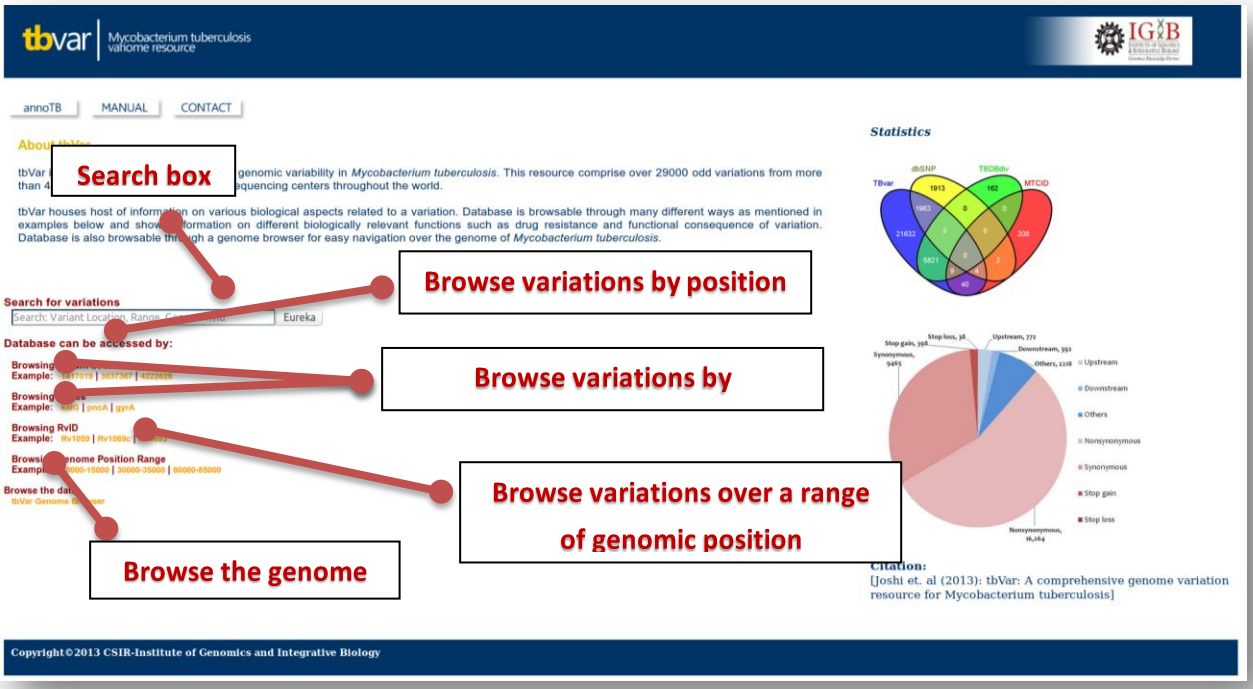
tbVar: A comprehensive genome variation resource for *Mycobacterium tuberculosis*

This is a general manual to help users search, navigate and even annotate their *Mycobacterium tuberculosis* (Mtb) genomic variations.

Homepage

Homepage of the database introduces the database and its purpose to the users. It also gives a brief overview of the data that the database holds.

Users can investigate the database using search box provided on the homepage using multiple parameters as shown in figures below.



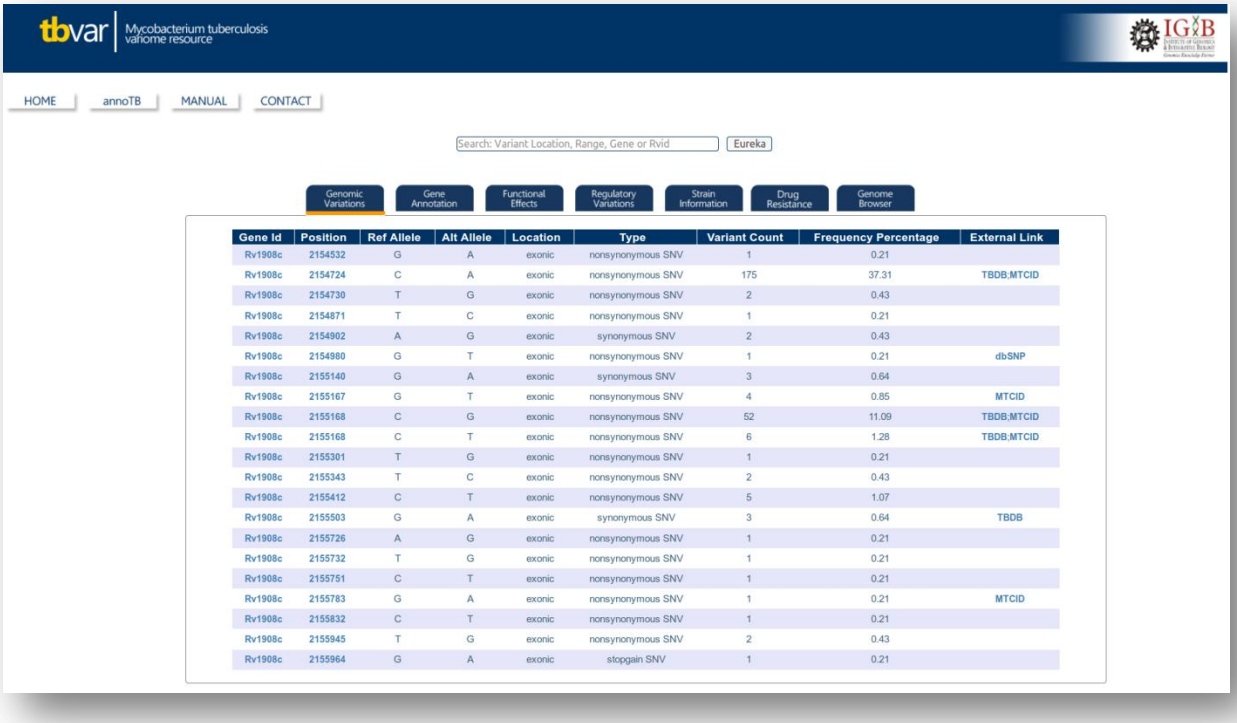
For the convenience of users database is made searchable through multiple features such as variant position, gene name/id or a range of genomic position as mentioned in a number of examples on the homepage.

Search term in any of the above mentioned form takes user to a separate result page which is further divided into different sections and easily.

tbvar

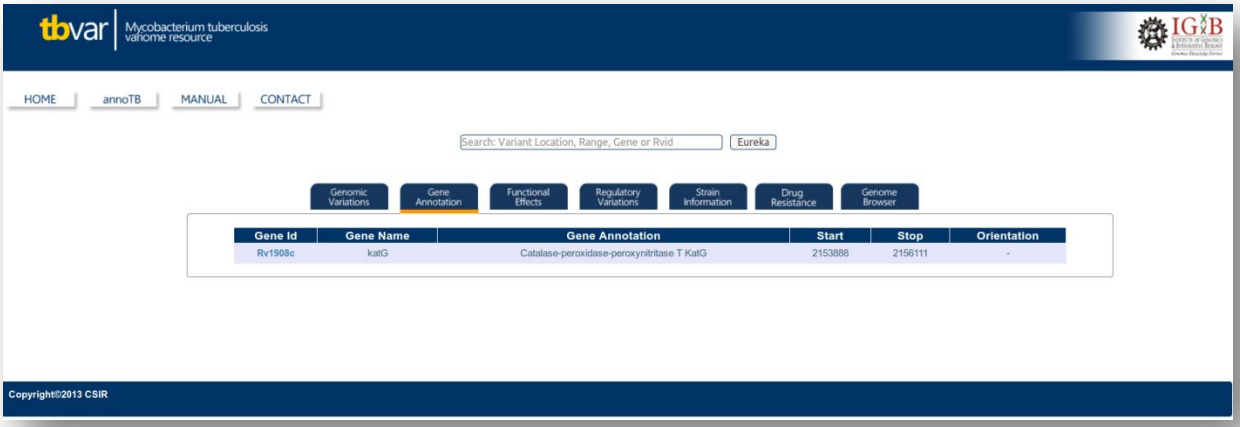
The output page for database query is divided into different sections which can be explored using different tabs at the top of result table. Different sections show different biological and other information related to query.

Genomic variations: This section gives information on the genomic position of the variations along with their genomic loci (e.g. whether they lie in genic part of the inter-genic region of the genome). This section also mentions the count and frequency of occurrence of the particular variation within the population of samples chosen for building database. Finally, this section also links out to other various databases for the variants found in other similar databases.

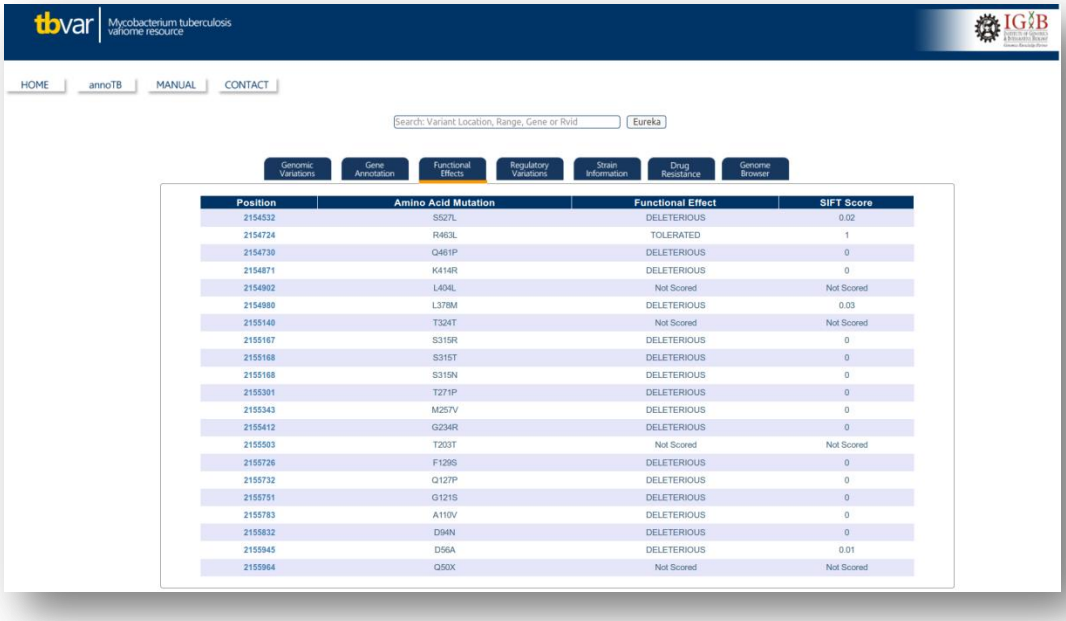


Gene Id	Position	Ref Allele	Alt Allele	Location	Type	Variant Count	Frequency Percentage	External Link
Rv1908c	2154532	G	A	exonic	nonsynonymous SNV	1	0.21	
Rv1908c	2154724	C	A	exonic	nonsynonymous SNV	175	37.31	TBDB;MTCID
Rv1908c	2154730	T	G	exonic	nonsynonymous SNV	2	0.43	
Rv1908c	2154871	T	C	exonic	nonsynonymous SNV	1	0.21	
Rv1908c	2154902	A	G	exonic	synonymous SNV	2	0.43	
Rv1908c	2154980	G	T	exonic	nonsynonymous SNV	1	0.21	dbSNP
Rv1908c	2155140	G	A	exonic	synonymous SNV	3	0.64	
Rv1908c	2155167	G	T	exonic	nonsynonymous SNV	4	0.85	MTCID
Rv1908c	2155168	C	G	exonic	nonsynonymous SNV	52	11.09	TBDB;MTCID
Rv1908c	2155168	C	T	exonic	nonsynonymous SNV	6	1.28	TBDB;MTCID
Rv1908c	2155301	T	G	exonic	nonsynonymous SNV	1	0.21	
Rv1908c	2155343	T	C	exonic	nonsynonymous SNV	2	0.43	
Rv1908c	2155412	C	T	exonic	nonsynonymous SNV	5	1.07	
Rv1908c	2155503	G	A	exonic	synonymous SNV	3	0.64	TBDB
Rv1908c	2155726	A	G	exonic	nonsynonymous SNV	1	0.21	
Rv1908c	2155732	T	G	exonic	nonsynonymous SNV	1	0.21	
Rv1908c	2155751	C	T	exonic	nonsynonymous SNV	1	0.21	
Rv1908c	2155783	G	A	exonic	nonsynonymous SNV	1	0.21	MTCID
Rv1908c	2155832	C	T	exonic	nonsynonymous SNV	1	0.21	
Rv1908c	2155945	T	G	exonic	nonsynonymous SNV	2	0.43	
Rv1908c	2155964	G	A	exonic	stopgain SNV	1	0.21	

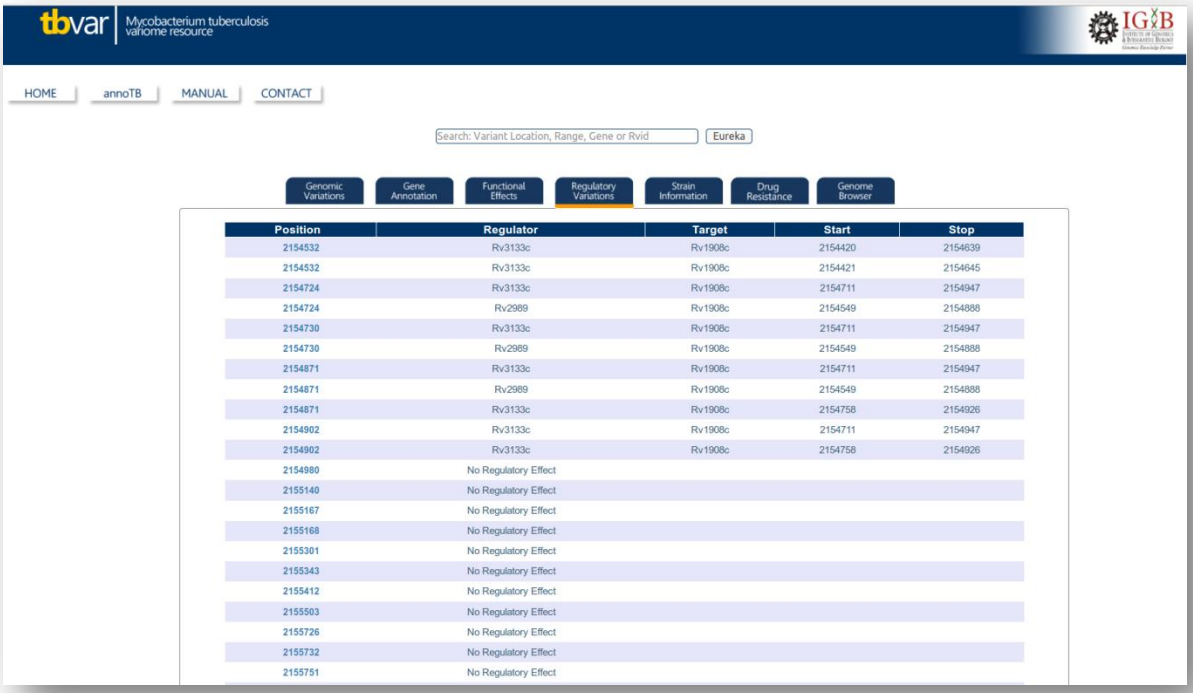
Gene annotation: This section shows all the information pertaining to a gene harboring variations. It shows gene name and it's ID. This tab also gives annotation for the respective gene. Genomic coordinates of genes and its orientation is also presented in this section.



Functional effects: This section shows SIFT prediction and SIFT score for non-synonymous variations. SIFT predicts whether a genomic variation has any functional consequence on the corresponding protein. Here we consider a change in protein structure as predicted by SIFT to be DELETERIOUS, while no change in the protein structure is considered as TOLERATED. SIFT score is a major parameter in prediction and is provided in this tab.

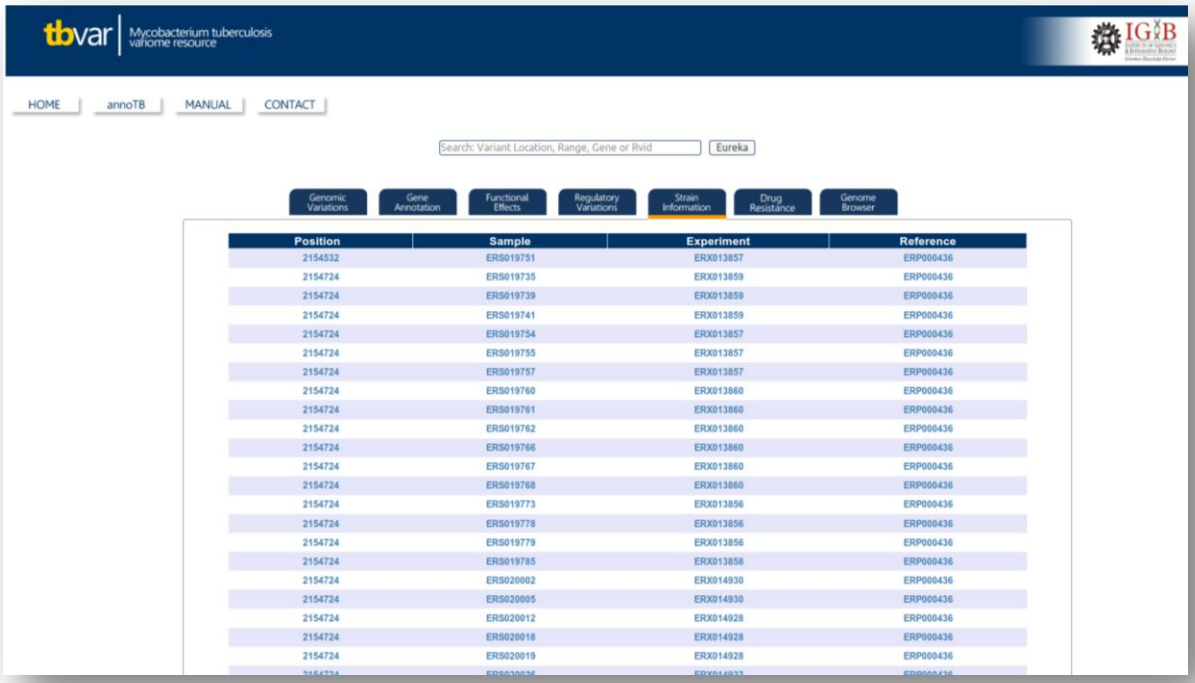


Regulatory variations: Variations lying in the regulatory elements on Mtb genome as found through Transcription factor (TF) ChIP-seq of 50 transcription factors are reported in this section. The targets predicted against these TFs are also reported in this section.

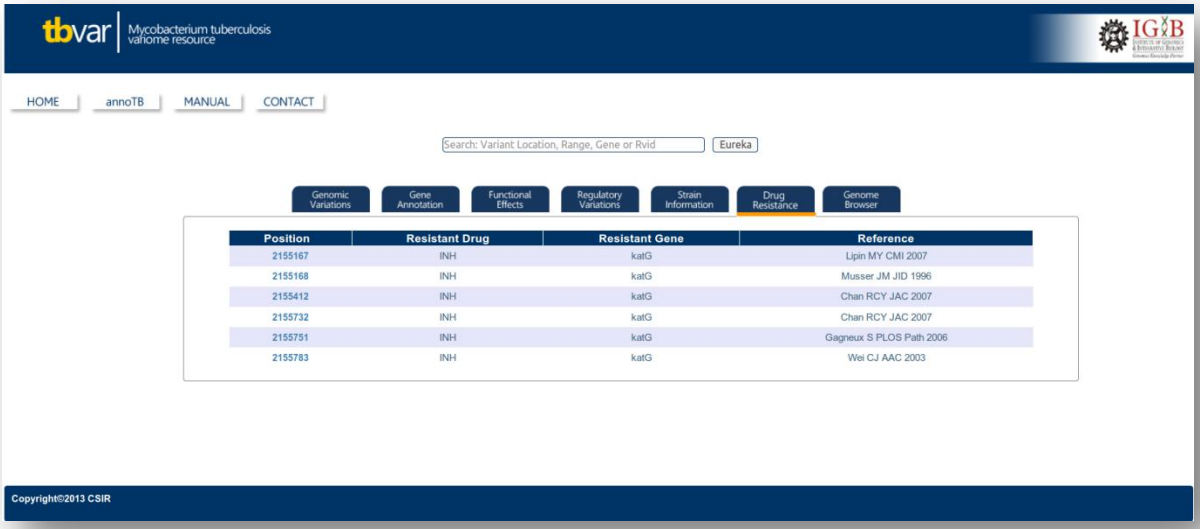


Position	Regulator	Target	Start	Stop
2154532	Rv3133c	Rv1908c	2154420	2154639
2154532	Rv3133c	Rv1908c	2154421	2154645
2154724	Rv3133c	Rv1908c	2154711	2154947
2154724	Rv2989	Rv1908c	2154549	2154888
2154730	Rv3133c	Rv1908c	2154711	2154947
2154730	Rv2989	Rv1908c	2154549	2154888
2154871	Rv3133c	Rv1908c	2154711	2154947
2154871	Rv2989	Rv1908c	2154549	2154888
2154871	Rv3133c	Rv1908c	2154758	2154926
2154902	Rv3133c	Rv1908c	2154711	2154947
2154902	Rv3133c	Rv1908c	2154758	2154926
2154980	No Regulatory Effect			
2155140	No Regulatory Effect			
2155167	No Regulatory Effect			
2155168	No Regulatory Effect			
2155301	No Regulatory Effect			
2155343	No Regulatory Effect			
2155412	No Regulatory Effect			
2155503	No Regulatory Effect			
2155726	No Regulatory Effect			
2155732	No Regulatory Effect			
2155751	No Regulatory Effect			

Strain information: This section reports information on the samples and corresponding experiment and study from which the variations were derived.

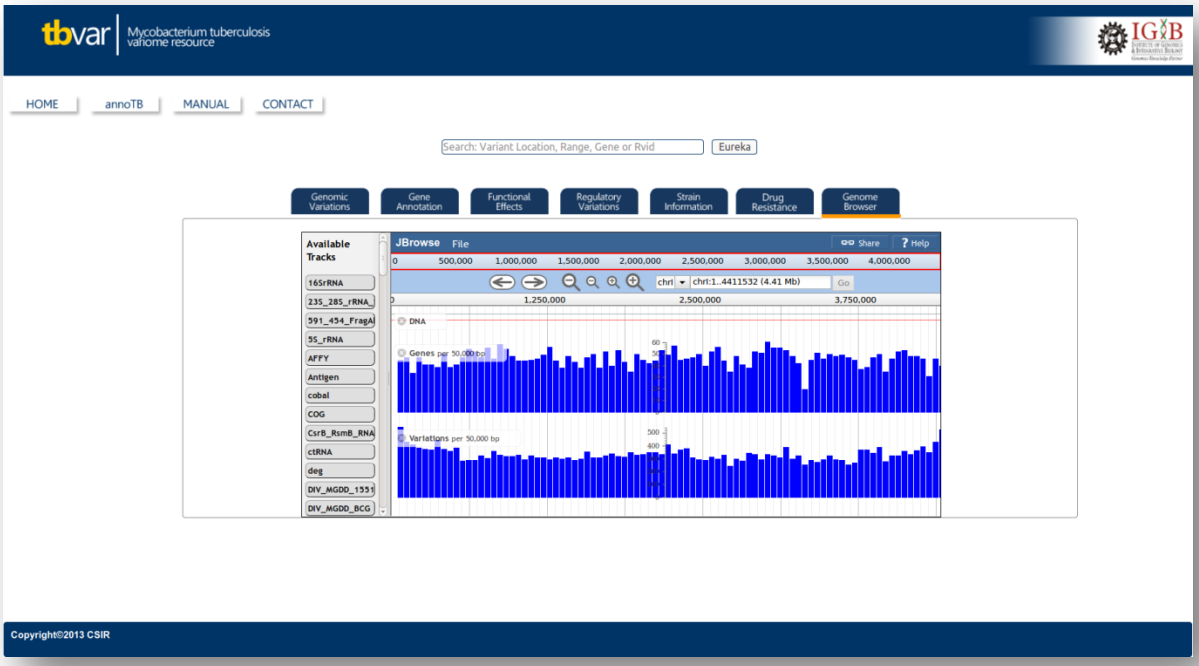


Drug resistance: Variations known to be annotated as drug resistant are reported in this section of the database. Variations having resistance to known anti-tb drug along with the antibiotic and corresponding resistant gene information is shown in this tab. Peer reviewed reference from which the information was derived is also reported. This tab appears only if the query has a drug resistant mutant associated with it.



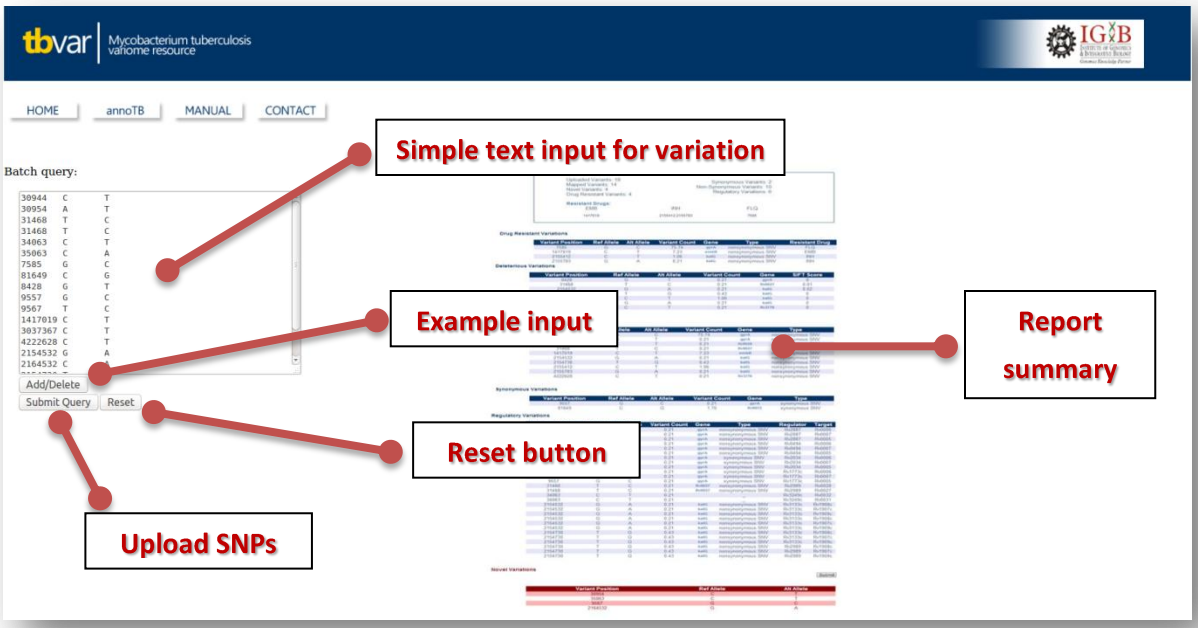
ncRNA: Variations falling in the non-coding RNA regions of the genome are reported in this tab. This is also an optional tab which arrears only if any variation in the query lies in the nc region of the genome.

Genome Browser: A genome browser hosting tracks from TBrowse and other relevant variation information and region surrounding the region of interest is shown in this section. Users can navigate through the genome by clicking and dragging the browser area. New tracks of interest can be added to the browser by a simple drag and drop from the vertical panel on the left. Users can also upload their own tracks of interest to show up in the browser by choosing File -> Open -> Select Files… / Paste remote URLs. Users can upload files from one of these file formats (GFF, BigWig, BAM & BAI). Configuration of the information to show in browser is customizable. Users have an option to either open the file or directly add it as a track.



annoTB

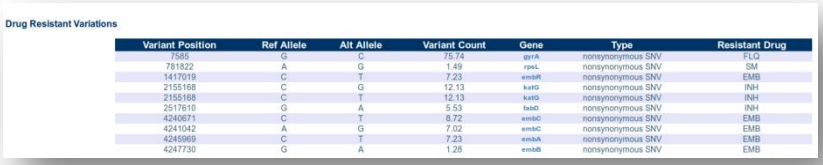
This is novel feature of tbvar, where doctors can simply upload Mtb genomes in simple SNP file format which is used by a parser to map those variations onto variations housed in 'tbvar'. Since tbvar is a comprehensive compendium of known genomic variations in Mtb, most of the variations in clinical sample genome are annotated.



A report showing variation annotation and corresponding information is available for doctors to analyze. The report gives a summary of annotated variations and also shows the variations known to confer drug resistance.

Different panels below report summary give information on annotation of individual variation section-wise.

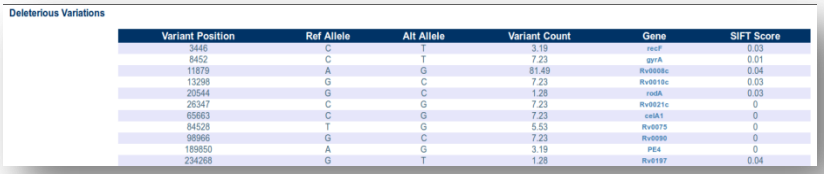
Drug resistance panel lists the variations annotated to be drug resistant.



Drug Resistant Variations

Variant Position	Ref Allele	Alt Allele	Variant Count	Gene	Type	Resistant Drug
7585	G	C	75.74	gyrA	nonsynonymous SNV	FLQ
781822	A	G	1.49	rpsL	nonsynonymous SNV	SM
1417019	C	T	7.23	embR	nonsynonymous SNV	EMB
2155168	C	G	12.13	katG	nonsynonymous SNV	INH
2155168	C	T	12.13	katG	nonsynonymous SNV	INH
2517610	G	A	5.53	fabD	nonsynonymous SNV	INH
4240671	C	T	8.72	embC	nonsynonymous SNV	EMB
4241042	A	G	7.02	embC	nonsynonymous SNV	EMB
4245969	C	T	7.23	embA	nonsynonymous SNV	EMB
4247730	G	A	1.28	embB	nonsynonymous SNV	EMB

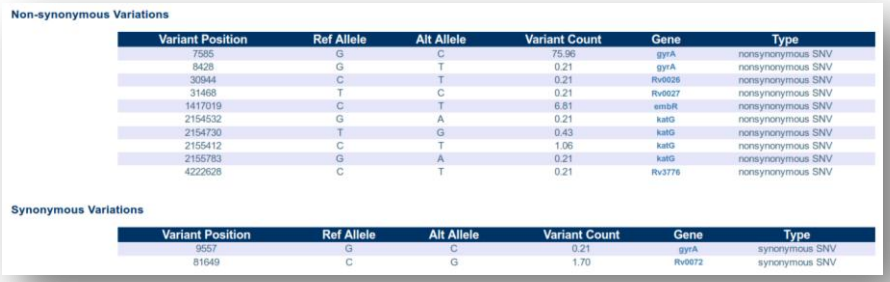
Deleterious variations panel lists deleterious variations predicted by SIFT which exist in database.



Deleterious Variations

Variant Position	Ref Allele	Alt Allele	Variant Count	Gene	SIFT Score
3446	C	T	3.19	recF	0.03
8452	C	T	7.23	gyrA	0.01
11879	A	G	81.49	Rv0008c	0.04
13298	G	C	7.23	Rv0010c	0.03
20544	G	C	1.28	rodA	0.03
26347	C	G	7.23	Rv0021c	0
65663	C	G	7.23	ceiA1	0
84528	T	G	5.53	Rv0075	0
98966	G	C	7.23	Rv0090	0
189850	A	G	3.19	PE4	0
234268	G	T	1.28	Rv0197	0.04

Non-synonymous and synonymous variations are listed in next two panels.



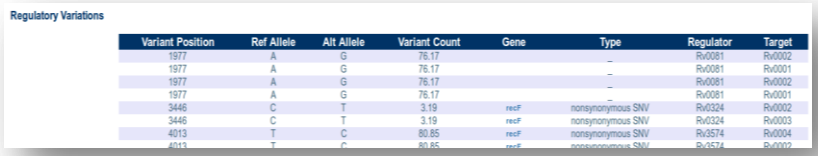
Non-synonymous Variations

Variant Position	Ref Allele	Alt Allele	Variant Count	Gene	Type
7585	G	C	75.96	gyrA	nonsynonymous SNV
8428	G	T	0.21	gyrA	nonsynonymous SNV
30944	C	T	0.21	Rv0026	nonsynonymous SNV
31468	T	C	0.21	Rv0027	nonsynonymous SNV
1417019	C	T	6.81	embR	nonsynonymous SNV
2154532	G	A	0.21	katG	nonsynonymous SNV
2154730	T	G	0.43	katG	nonsynonymous SNV
2155412	C	T	1.06	katG	nonsynonymous SNV
2155783	G	A	0.21	katG	nonsynonymous SNV
4222628	C	T	0.21	Rv3776	nonsynonymous SNV

Synonymous Variations

Variant Position	Ref Allele	Alt Allele	Variant Count	Gene	Type
9557	G	C	0.21	gyrA	synonymous SNV
81649	C	G	1.70	Rv0072	synonymous SNV

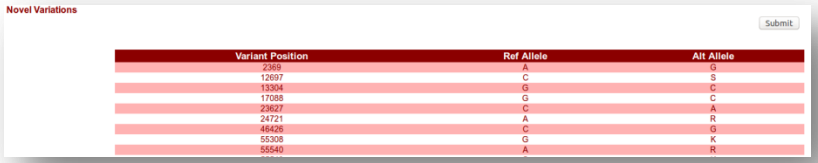
Regulatory variations in uploaded SNP file matching to those present in database are shown in regulatory variation panel.



Regulatory Variations

Variant Position	Ref Allele	Alt Allele	Variant Count	Gene	Type	Regulator	Target
1977	A	G	76.17		_	Rv0081	Rv0002
1977	A	G	76.17		_	Rv0081	Rv0001
1977	A	G	76.17		_	Rv0081	Rv0002
1977	A	G	76.17		_	Rv0081	Rv0001
3446	C	T	3.19	recF	nonsynonymous SNV	Rv0324	Rv0002
3446	C	T	3.19	recF	nonsynonymous SNV	Rv0324	Rv0003
4013	T	C	80.85	recF	nonsynonymous SNV	Rv3574	Rv0004
4013	T	C	80.85	recF	nonsynonymous SNV	Rv3574	Rv0002

The last panel lists those variations not present in ‘tbvar’. Users have an option to submit these novel variations to the database.



Novel Variations

Submit

Variant Position	Ref Allele	Alt Allele
2369	A	G
12697	C	S
13304	G	C
17088	G	C
23627	C	A
24721	A	R
46426	C	G
55308	G	K
55540	A	R

When users press the submit button, a form asking for information on the submitter and submission. By submitting the form users submit the SNP file they uploaded to the server where manual curation is done and the data is included into the database.