

VCF-Server User Manual

(1st Edition)

2018.6

Content

Background.....	3
Study Management	4
New Study (Now opens up to login users)	4
Open Study	5
Close Study	6
Delete Study	6
Upload VCF File	7
VCF File Procession.....	9
Analyze VCF File	9
Customize Columns	10
Customize Filters	11
Add Filter.....	11
Filter Example	13
Remove Filter	14
Export Filtered Variants to CSV	14
Build Columns Indexes (Speed up filtering)	15
Annotate VCF File.....	16
Select Annotation Databases	16
Annotate Selected Items in Databases	18
Information of annotation databases	18
Download VCF File	19
Delete VCF File	19

Background

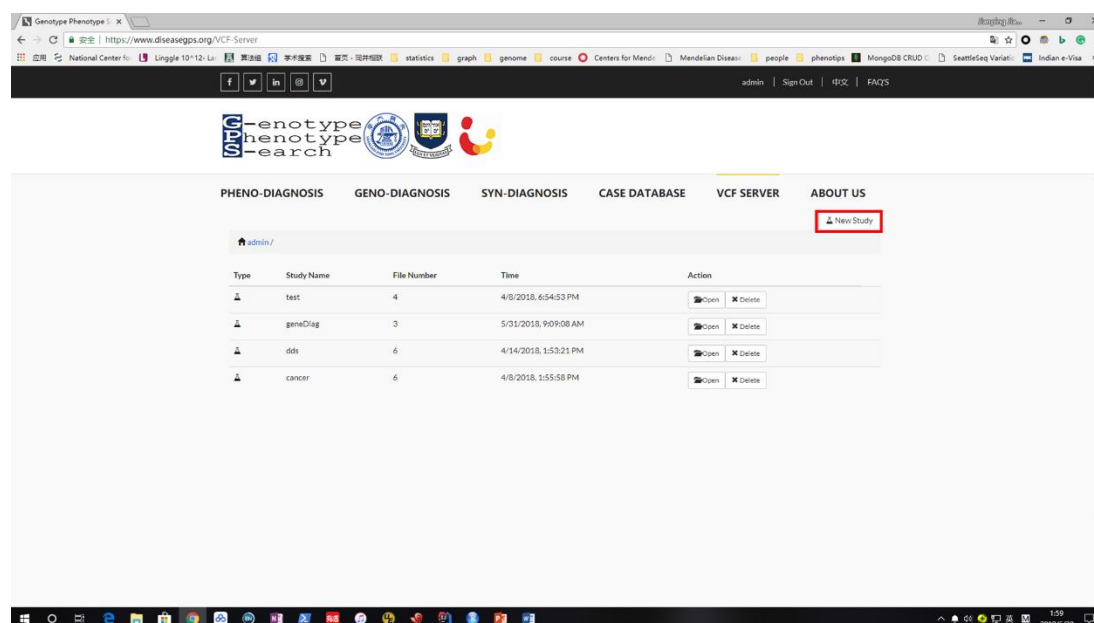
Next generation sequencing (NGS) has facilitated the discovery of disease-linked genetic variants and has been widely used for disease aetiology research and clinical genetic diagnostics. Mutations identified by NGS are usually recorded in variant call format (VCF) file, which has become the community standard for storing mutations data. The VCF file describes each mutation as a deviation from the indicated reference genome assembly, such as hg19 and hg38. Although it is a common and standard format for most of the bioinformatics tools, VCF lacks readability and is difficult to process for scientists without programming skills, especially when the volume of VCF file becomes large. Furthermore, as the number of VCF files generated by NGS increases from several to thousands, there is growing concern about the management of these huge mutation data. Therefore, there is a big gap between VCF files and the ‘end users’ urges to develop a simple tool to mine mutations data in VCF files.

We report development of VCF-Server, which overcomes many limitations found in currently available tools, and providing fast and one-stop service for VCF files processing expert. It helps the user manage, query, annotate and prioritize their variants data in a web platform based on a browser interface and is suitable for Windows, Mac and UNIX systems. VCF-Server manages different sequencing studies created from users’ uploaded VCF files and visualizes them into dynamic tables on a web browser. User can annotate VCF files with common used databases and filter variants by flexible rule chains. The filtered variants can be export to local after prioritization. VCF-Server has optimized VCF file loading and indexing to fulfill rapid and frequent querying and filtering of variants. It allows researchers with little bioinformatics backgrounds to explore and interpret mutation data, thereby fostering translational research in the field of genetics.

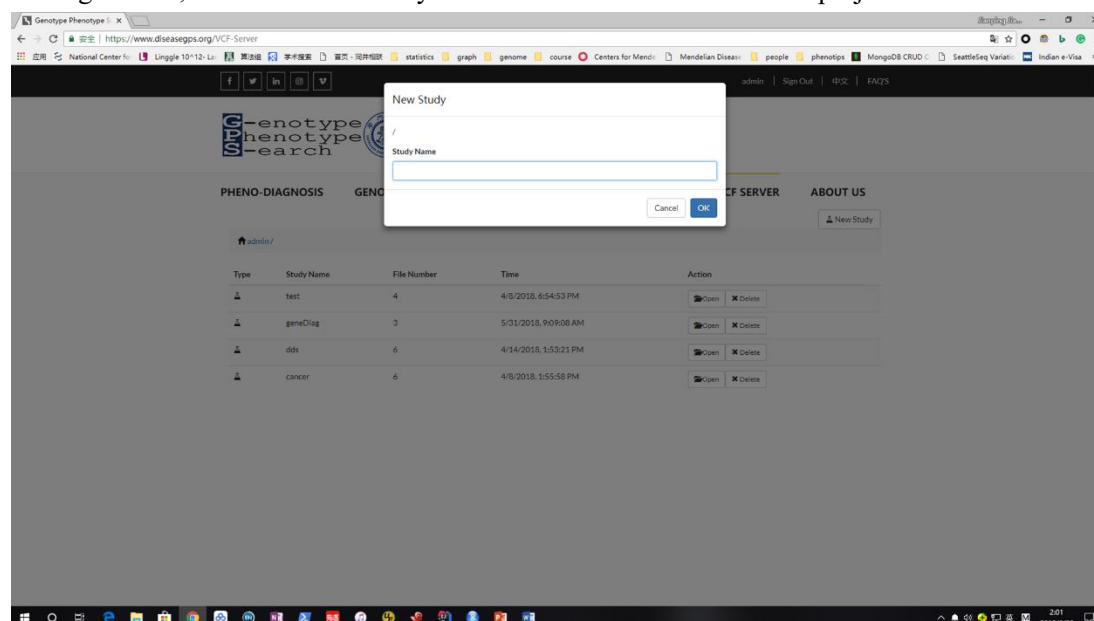
Study Management

The huge amount of mutations data generated by NGS poses a challenge on variants management. VCF-Server designs to manage sequencing studies created from uploaded VCF files, making them traceable and easy to locate. The user can create studies with meaningful names on VCF-Server according to research purposes.

New Study (Now opens up to login users)

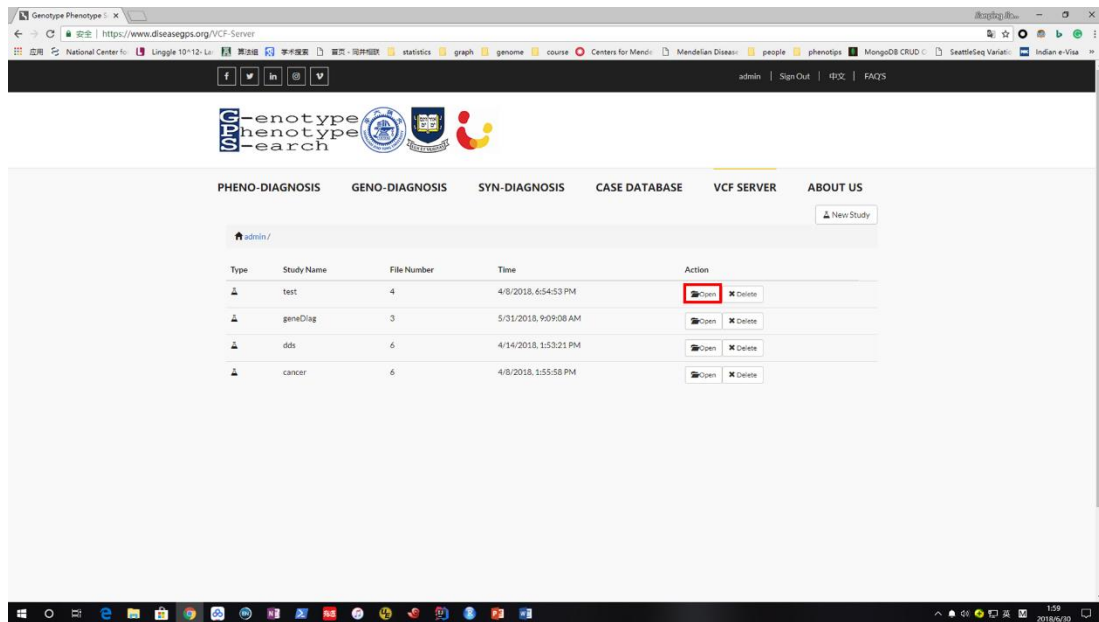


For login users, click the 'New Study' button and create a new research project.



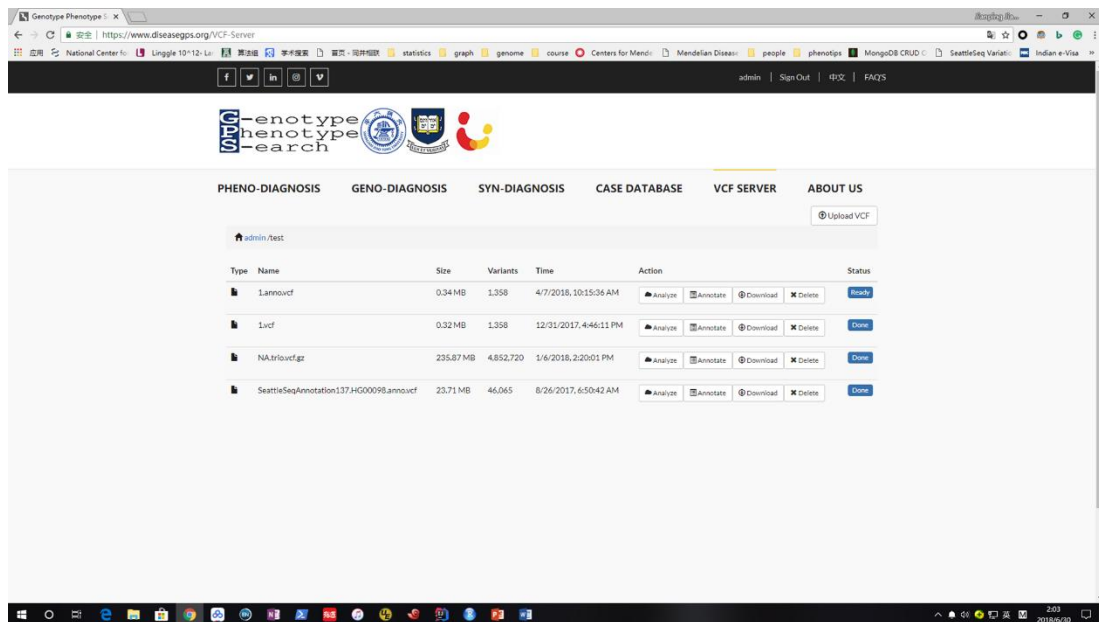
After entering project name and click 'OK' button, the new project will be showed in the project list.

Open Study



The screenshot shows the VCF-Server web interface. The top navigation bar includes links for PHENO-DIAGNOSIS, GENO-DIAGNOSIS, SYN-DIAGNOSIS, CASE DATABASE, VCF SERVER, and ABOUT US. The 'PHENO-DIAGNOSIS' tab is active. Below the navigation bar, there is a 'New Study' button. The main content area displays a table of studies. The 'test' study is highlighted with a red box around the 'Open' button.

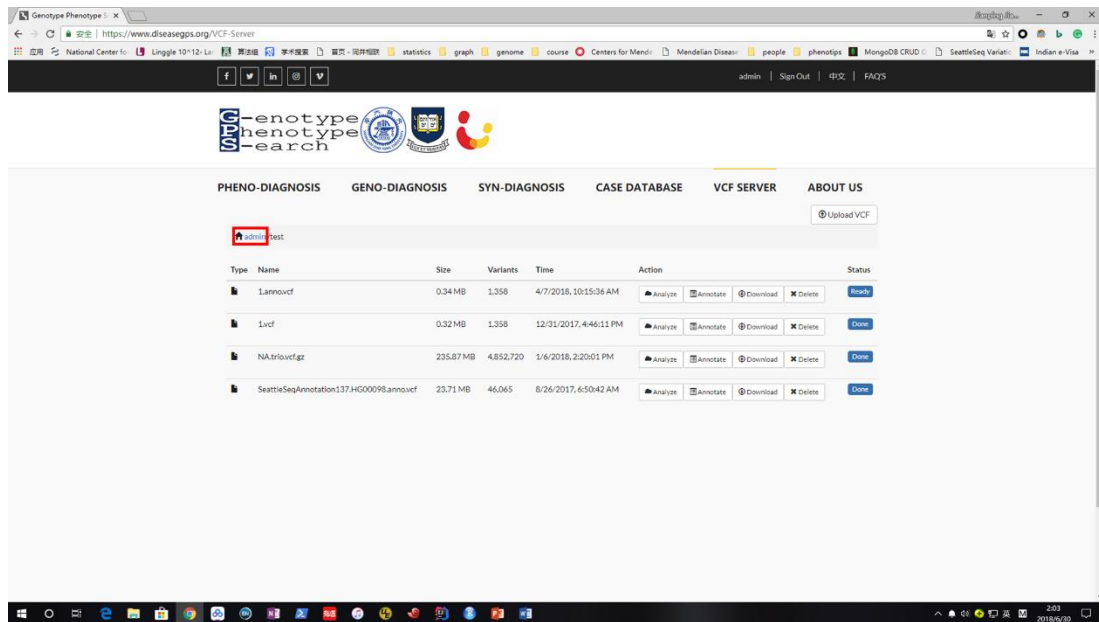
Type	Study Name	File Number	Time	Action
test	test	4	4/8/2018, 6:54:53 PM	Open Delete
geneDiag	geneDiag	3	5/31/2018, 9:09:08 AM	Open Delete
dds	dds	6	4/14/2018, 1:53:21 PM	Open Delete
cancer	cancer	6	4/8/2018, 1:55:58 PM	Open Delete



The screenshot shows the VCF-Server web interface. The top navigation bar includes links for PHENO-DIAGNOSIS, GENO-DIAGNOSIS, SYN-DIAGNOSIS, CASE DATABASE, VCF SERVER, and ABOUT US. The 'VCF SERVER' tab is active. Below the navigation bar, there is an 'Upload VCF' button. The main content area displays a table of VCF files. The 'test' study is highlighted with a red box around the 'Open' button.

Type	Name	Size	Variants	Time	Action	Status
1annoVcf	1annoVcf	0.34 MB	1,358	4/7/2018, 10:15:36 AM	Analyze Annotate Download Delete	Ready
1vcf	1vcf	0.32 MB	1,358	12/31/2017, 4:46:11 PM	Analyze Annotate Download Delete	Done
NA.trioVcf.gz	NA.trioVcf.gz	235.87 MB	4,852,720	1/4/2018, 2:20:01 PM	Analyze Annotate Download Delete	Done
SeattleSeqAnnotation137.HC000098.annoVcf	SeattleSeqAnnotation137.HC000098.annoVcf	23.71 MB	46,065	8/26/2017, 6:50:42 AM	Analyze Annotate Download Delete	Done

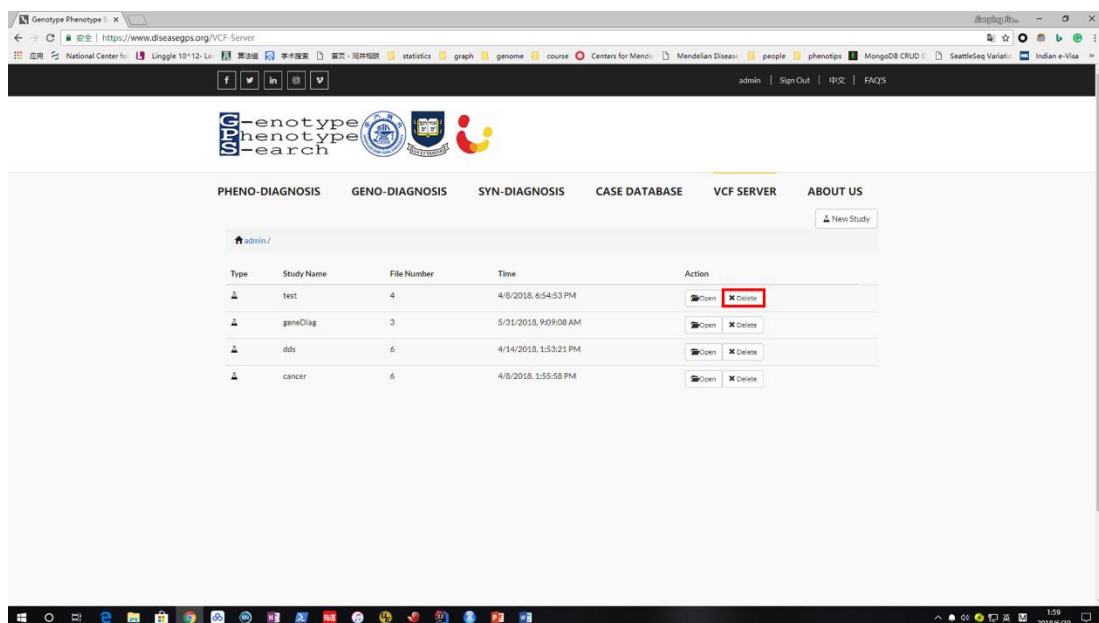
Close Study



The screenshot shows the VCF-Server web application. The 'admin' button is highlighted in red. The table below lists the studies available for deletion.

Type	Name	Size	Variants	Time	Action	Status
1	Lannouv	0.34 MB	1,358	4/7/2018, 10:15:36 AM	Analyze Annotate Download Delete	Ready
1	Luvf	0.32 MB	1,358	12/31/2017, 4:46:11 PM	Analyze Annotate Download Delete	Done
1	NA.triouv.gz	235.87 MB	4,852,720	1/6/2018, 2:20:01 PM	Analyze Annotate Download Delete	Done
1	SeattleSeqAnnotation137+HG00098.lannouv	23.71 MB	46,065	8/26/2017, 6:50:42 AM	Analyze Annotate Download Delete	Done

Delete Study



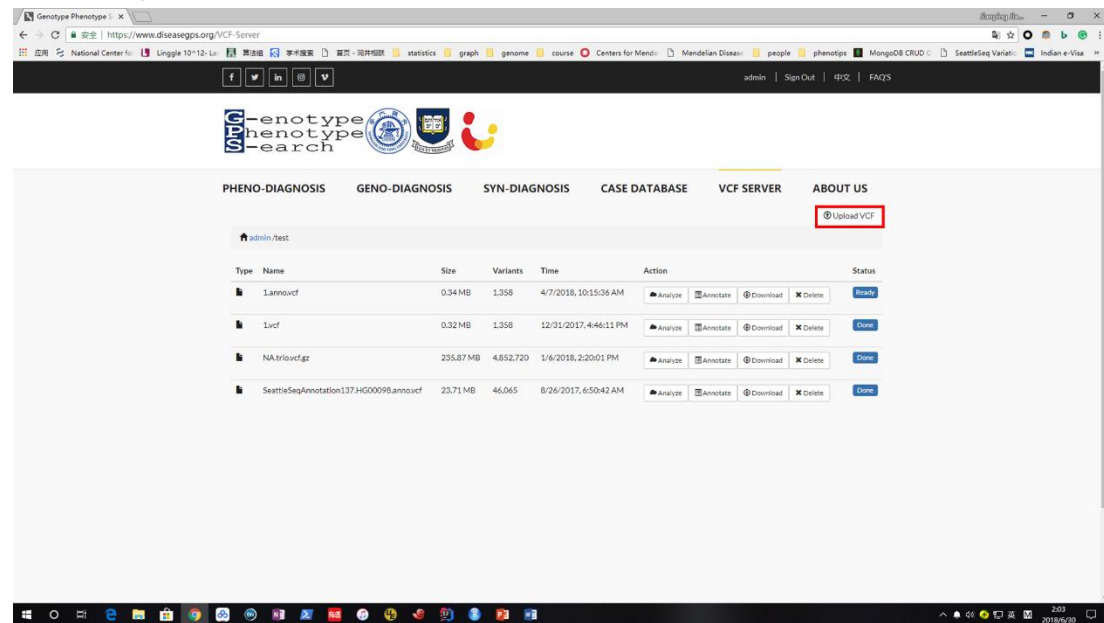
The screenshot shows the VCF-Server web application. The 'Delete' button is highlighted in red. The table below lists the studies available for deletion.

Type	Study Name	File Number	Time	Action
1	test	4	4/8/2018, 6:54:53 PM	Open Delete
1	geneDiag	3	5/31/2018, 9:09:08 AM	Open Delete
1	dds	6	4/14/2018, 1:53:21 PM	Open Delete
1	cancer	6	4/8/2018, 1:55:58 PM	Open Delete

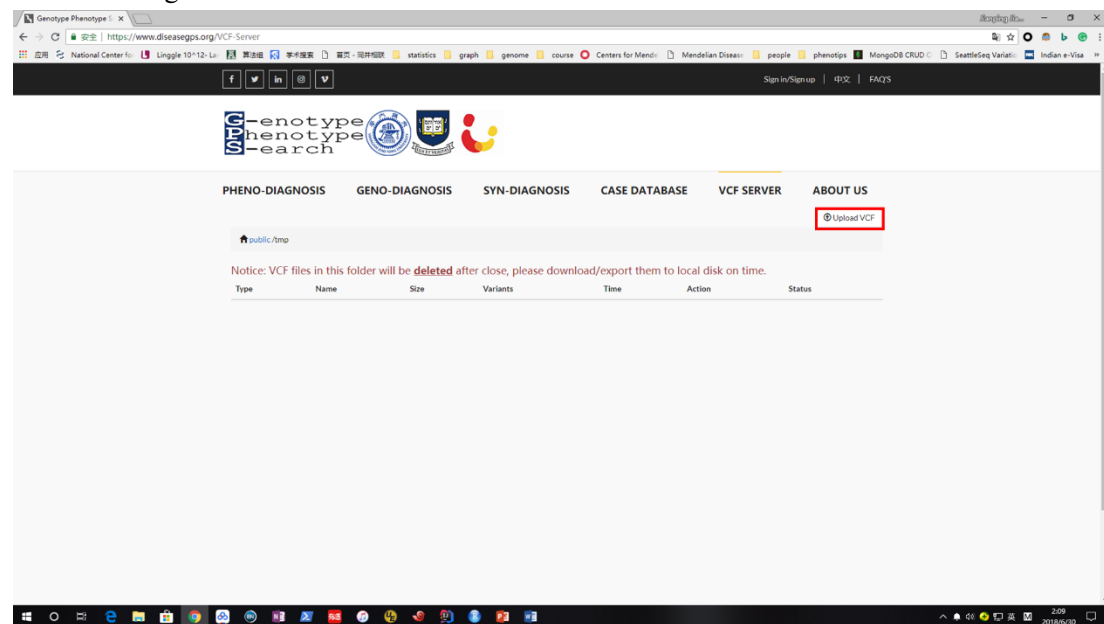
Mention that only projects without any VCF file could be delete.

Upload VCF File

● For login users

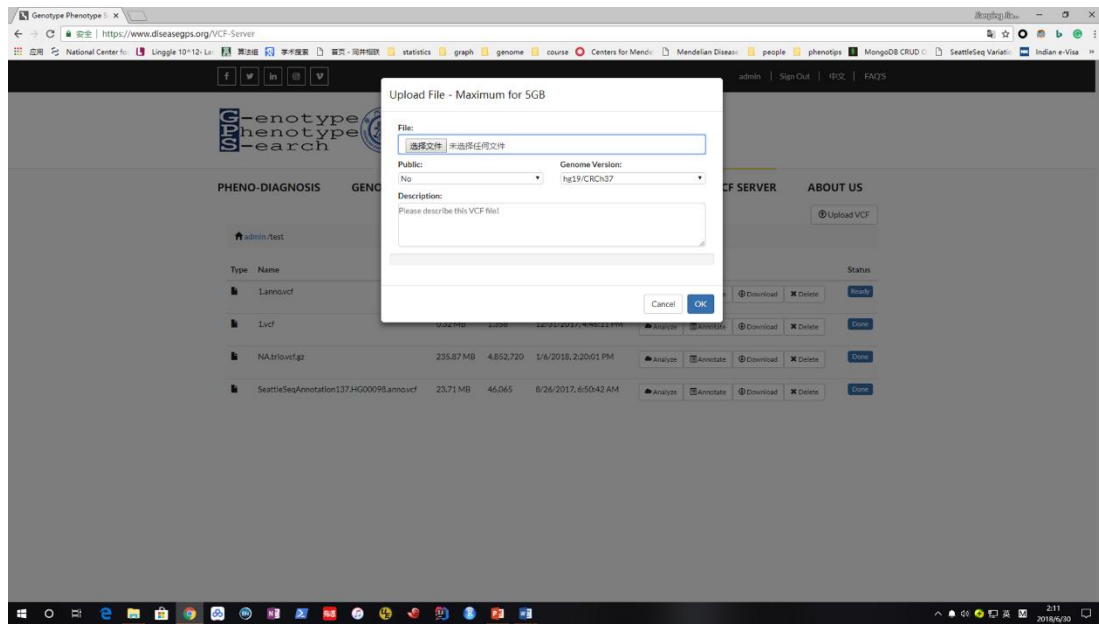


● For unlogin users



Mention that unlogin users could only upload VCF files to 'tmp' project, which will be cleared after closing the web browser.

● VCF File Description



VCF File Processing

VCF-Server is a user-friendly and one-stop platform that provides rich user interfaces to assist scientists with little bioinformatics background to mine disease-linked variants in VCF files. The platform provides services on VCF file management, variants annotation, visualization and prioritization, and filtered variants export that facilitate user to manage and interrogate their VCF files.

Analyze VCF File

The screenshot shows the VCF-Server web interface. The 'VCF SERVER' tab is selected. The table lists the following VCF files:

Type	Name	Size	Variants	Time	Action	Status
L1annovcf		0.34 MB	1,358	4/7/2018, 10:15:36 AM	Analyze Annotate Download Delete	Ready
L1vcf		0.32 MB	1,358	12/31/2017, 4:46:11 PM	Analyze Annotate Download Delete	Done
NK1vcf.gz		235.87 MB	4,852,720	1/6/2018, 2:20:01 PM	Analyze Annotate Download Delete	Done
SeattleSeqAnnotation137HG00098.annovcf		23.71 MB	46,065	8/26/2017, 6:50:42 AM	Analyze Annotate Download Delete	Done

The screenshot shows the VCF-Server web interface with a detailed view of a VCF file. The table displays the following variants:

CHROM	POS	ID	REF	ALT	QUAL	FILTER	Samples	#
chr1	948,846	rs3841266	T	TA	1537.8	PASS	GP1	1
chr1	948,870	rs4615788	C	G	1412.8	PASS	GP1	1
chr1	948,921	rs15842	T	C	23230.8	PASS	GP1	1
chr1	949,654	rs5997	A	G	2092.8	PASS	GP1	1
chr1	949,925	rs2799070	C	T	25934.8	PASS	GP1	1
chr1	1,147,422	rs17568	C	T	3532.7	PASS	GP1	1
chr1	9,770,690	-	C	CAG	3216.8	PASS	GP1	1
chr1	9,776,575	rs2230735	A	G	889.8	PASS	GP1	1
chr1	9,776,715	rs182137610	A	C	3028.8	PASS	GP1	1
chr1	9,777,669	rs28730671	C	T				

Customize Columns

The screenshot shows the VCF-Server web interface. The top navigation bar includes links for PHENO-DIAGNOSIS, GENO-DIAGNOSIS, SYN-DIAGNOSIS, CASE DATABASE, VCF SERVER, and ABOUT US. The VCF SERVER section is active, displaying a table of VCF entries. The table has columns: CHROM, POS, ID, REF, ALT, QUAL, FILTER, Samples, and #_San. The table shows 10 entries, with the first entry being chr1:948,846:rs3841285:T:TA:1283.7:PASS:GP1:1. The table is paginated, showing 1 to 10 of 1,000 entries.

The screenshot shows the VCF-Server web interface with the 'Customize Columns' dialog box open. The dialog box lists the following columns and their descriptions:

- ☐ Name: Description
- ☒ CHROM: The chromosome
- ☒ POS: The reference position, with the 1st base having position 1
- ☒ ID: Semi-colon separated list of unique identifiers
- ☒ REF: The reference base(s). Each base must be one of A,C,G,T,N (case insensitive)
- ☒ ALT: Comma separated list of alternate non-reference alleles called on at least one of the samples
- ☒ QUAL: Phred-scaled quality score for the assertion made in ALT, i.e. $-10\log_{10}(\text{prob}(\text{call in ALT is wrong}))$
- ☒ FILTER: PASS if this position has passed all filterSteps, i.e. a call is made at this position. Otherwise, if the site has not passed all filterSteps, a semicolon-separated list of codes for filterSteps that fail, e.g. q10s50 might indicate that at this site the quality is below 10 and the number of samples with data is below 50% of the total number of samples
- ☒ Samples: The names of samples
- ☒ #_Samples: The number of samples
- ☐ FORMAT:GP.LAD: Allelic depths for the ref and alt alleles in the order listed
- ☐ FORMAT:GP.LDP: Approximate read depth (reads with MQ
- ☐ FORMAT:GP.LGQ: Genotype Quality
- ☐ FORMAT:GP.LGT: Genotype
- ☐ FORMAT:GP.LPL: Normalized, Phred-scaled likelihoods for genotypes as defined in the VCF specification
- ☐ INFO:AC: Allele count in genotypes, for each ALT allele, in the same order as listed
- ☐ INFO:AF: Allele Frequency, for each ALT allele, in the same order as listed
- ☐ INFO:AN: Total number of alleles in called genotypes

The dialog box has a 'Close' button at the bottom right.

Customize Filters

Genotype Phenotype Search

PHENO-DIAGNOSIS GENO-DIAGNOSIS SYN-DIAGNOSIS CASE DATABASE VCF SERVER ABOUT US

Show: 10 entries Search:

CHROM	POS	ID	REF	ALT	QUAL	FILTER	Samples	#_Sam
chr1	948,846	rs3841266	T	TA	1283.7	PASS	GP1	1
chr1	948,870	rs4615788	C	G	1537.8	PASS	GP1	1
chr1	948,921	rs155842	T	C	1412.8	PASS	GP1	1
chr1	949,654	rs5997	A	G	23230.8	PASS	GP1	1
chr1	949,925	rs2799070	C	T	2092.8	PASS	GP1	1
chr1	1,147,422	rs17568	C	T	25934.8	PASS	GP1	1
chr1	9,770,690	-	C	CAG	3532.7	PASS	GP1	1
chr1	9,776,575	rs2230735	A	G	3216.8	PASS	GP1	1
chr1	9,776,715	rs182137610	A	C	889.8	PASS	GP1	1
chr1	9,777,669	rs28730671	C	T	3028.8	PASS	GP1	1

Showing 1 to 10 of 1,000 entries

Previous 1 2 3 4 5 100 Next

Add Filter

Genotype Phenotype Search

PHENO-DIAGNOSIS GENO-DIAGNOSIS SYN-DIAGNOSIS CASE DATABASE VCF SERVER ABOUT US

Show: 10 entries Search:

CHROM	POS	ID	REF	ALT	QUAL	FILTER	Samples	#_Sam
chr1	948,846	rs3841266	T	TA	1283.7	PASS	GP1	1
chr1	948,870	rs4615788	C	G	1537.8	PASS	GP1	1
chr1	948,921	rs155842	T	C	1412.8	PASS	GP1	1
chr1	949,654	rs5997	A	G	23230.8	PASS	GP1	1
chr1	949,925	rs2799070	C	T	2092.8	PASS	GP1	1
chr1	1,147,422	rs17568	C	T	25934.8	PASS	GP1	1
chr1	9,770,690	-	C	CAG	3532.7	PASS	GP1	1
chr1	9,776,575	rs2230735	A	G	3216.8	PASS	GP1	1
chr1	9,776,715	rs182137610	A	C	889.8	PASS	GP1	1
chr1	9,777,669	rs28730671	C	T	3028.8	PASS	GP1	1

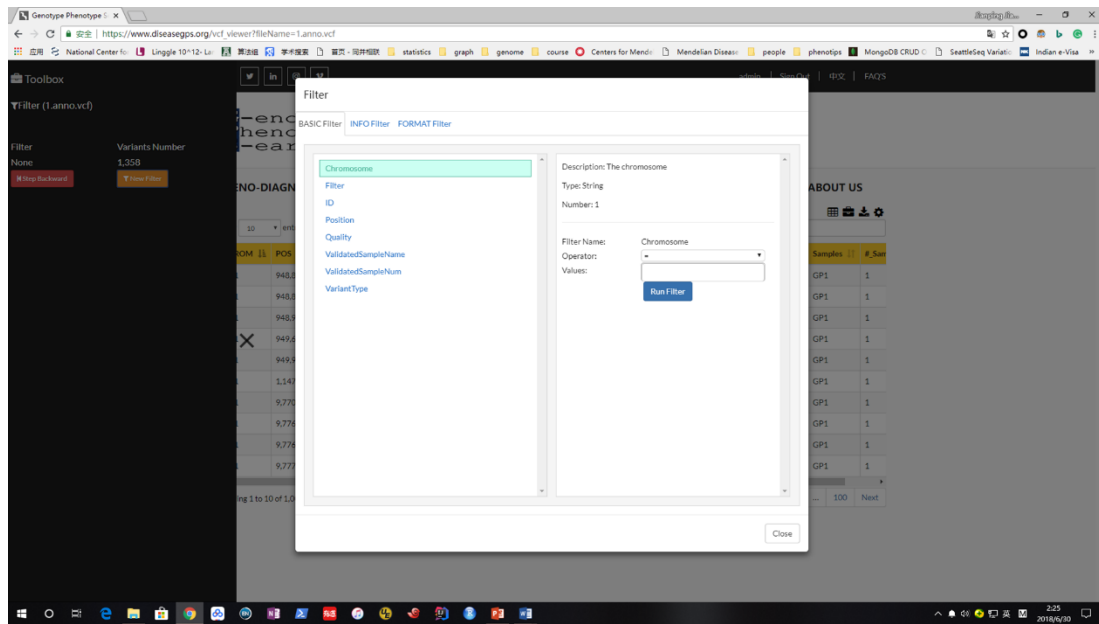
Showing 1 to 10 of 1,000 entries

Previous 1 2 3 4 5 100 Next

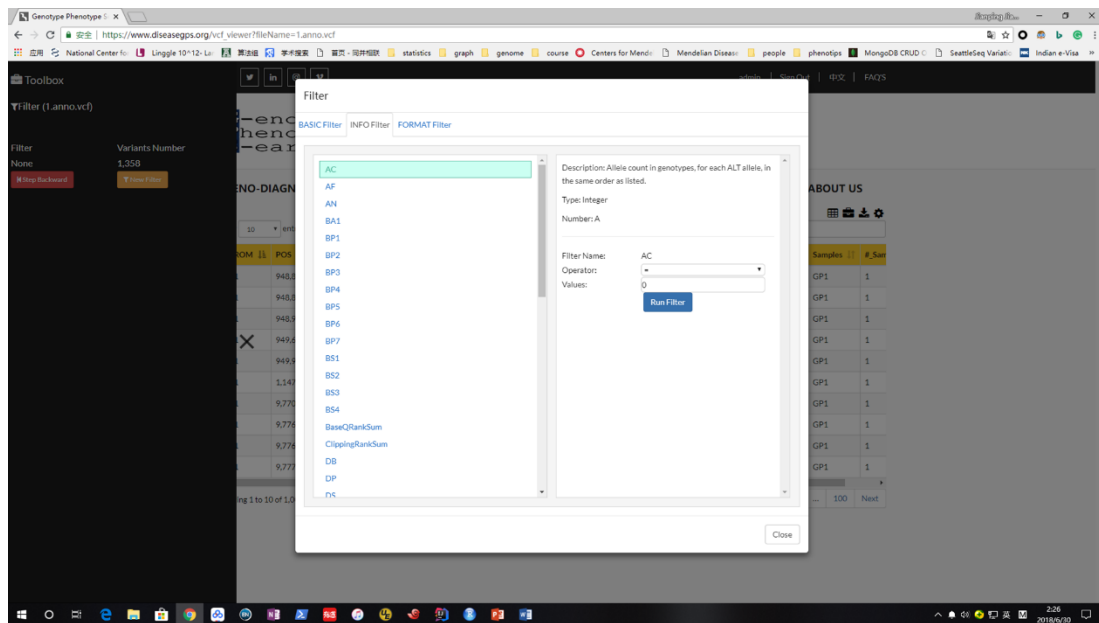
Filter: None Variants Number: 1,398

Stop Backward New Filter

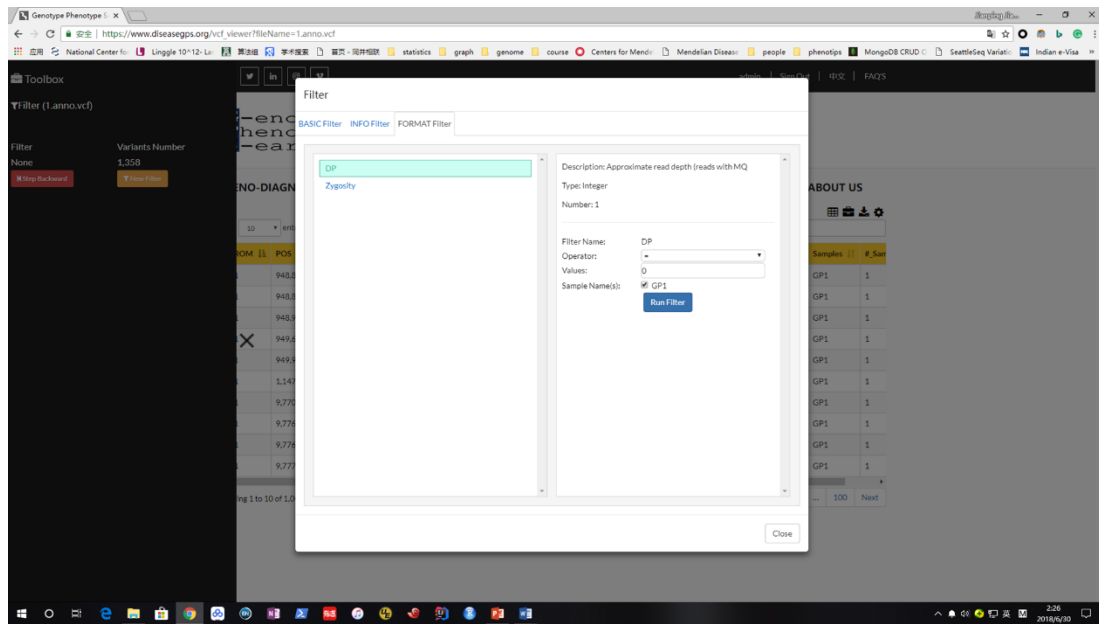
● Basic Filter



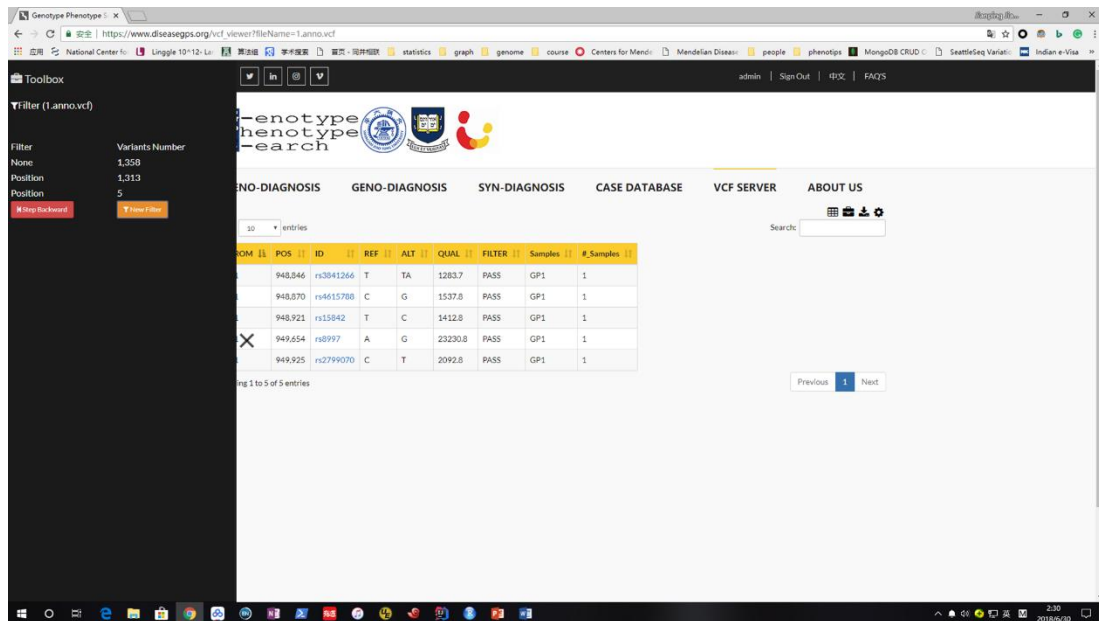
● Information Filter



● Format Filter



Filter Example



Get variants between 900K and 1M region on chromosome 1.

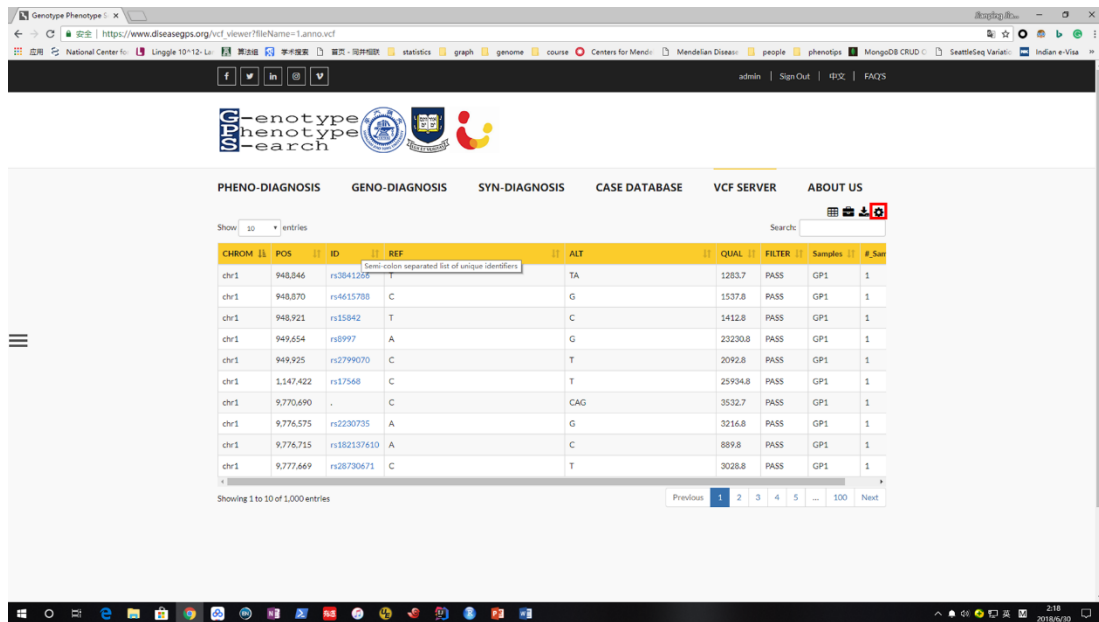
Remove Filter

The screenshot shows the VCF-Server web application. On the left sidebar, under the 'Filter (1.anno.vcf)' section, there is a 'Remove Filter' button. The main content area displays a table of variants with columns: CHROM, POS, ID, REF, ALT, QUAL, FILTER, Samples, and #Samples. The table shows 5 entries, with the first entry having a 'PASS' filter. The 'Remove Filter' button is highlighted in red.

Export Filtered Variants to CSV

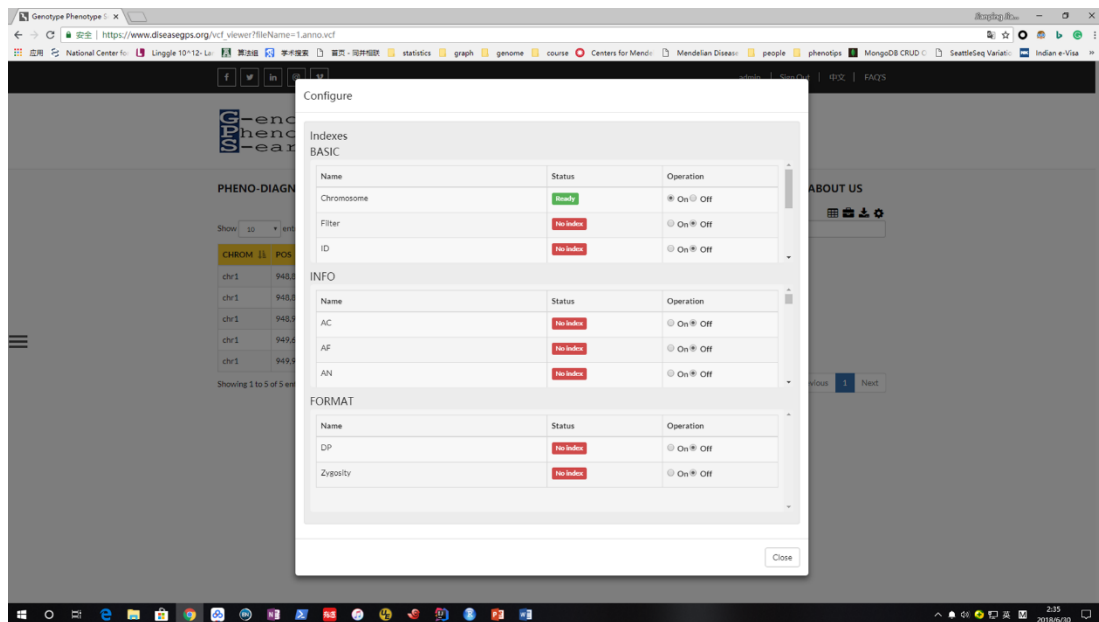
The screenshot shows the VCF-Server web application. In the top right corner, there is an 'Export to CSV' button. The main content area displays a table of variants with columns: CHROM, POS, ID, REF, ALT, QUAL, FILTER, Samples, and #Samples. The table shows 10 entries, with the first entry having a 'PASS' filter. The 'Export to CSV' button is highlighted in red.

Build Columns Indexes (Speed up filtering)



The screenshot shows the VCF-Server interface with a table of VCF entries. The table has columns: CHROM, POS, ID, REF, ALT, QUAL, FILTER, and Samples. The data is displayed in a grid format. A tooltip is visible over the ID column, showing a semi-colon separated list of unique identifiers. The interface includes a search bar, a 'Show' dropdown, and a 'Previous' button.

CHROM	POS	ID	REF	ALT	QUAL	FILTER	Samples
chr1	948,846	rs3841285	T	TA	1283.7	PASS	GP1
chr1	948,870	rs4615788	C	G	1537.8	PASS	GP1
chr1	948,921	rs15842	T	C	1412.8	PASS	GP1
chr1	949,654	rs5997	A	G	23230.8	PASS	GP1
chr1	949,925	rs2799070	C	T	2092.8	PASS	GP1
chr1	1,147,422	rs17568	C	T	25934.8	PASS	GP1
chr1	9,770,690	.	C	CAG	3532.7	PASS	GP1
chr1	9,776,575	rs2230735	A	G	3216.8	PASS	GP1
chr1	9,776,715	rs182137610	A	C	889.8	PASS	GP1
chr1	9,777,669	rs28730671	C	T	3028.8	PASS	GP1



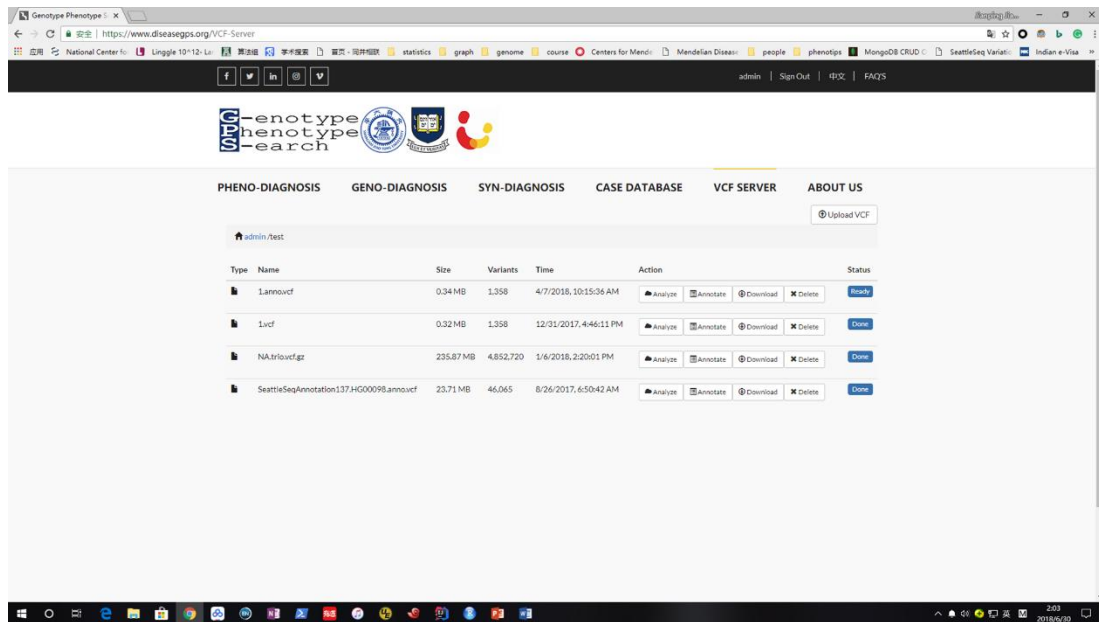
The screenshot shows the VCF-Server interface with a 'Configure' dialog box open. The dialog box has three sections: BASIC, INFO, and FORMAT. Each section contains a table with columns: Name, Status, and Operation. The 'BASIC' section shows 'Chromosome' as 'Ready', 'Filter' as 'No Index', and 'ID' as 'No Index'. The 'INFO' section shows 'AC', 'AF', and 'AN' as 'No Index'. The 'FORMAT' section shows 'DP' and 'Zygosity' as 'No Index'. The 'Operation' column for all items shows 'On * Off'.

Name	Status	Operation
Chromosome	Ready	On * Off
Filter	No Index	On * Off
ID	No Index	On * Off

Name	Status	Operation
AC	No Index	On * Off
AF	No Index	On * Off
AN	No Index	On * Off

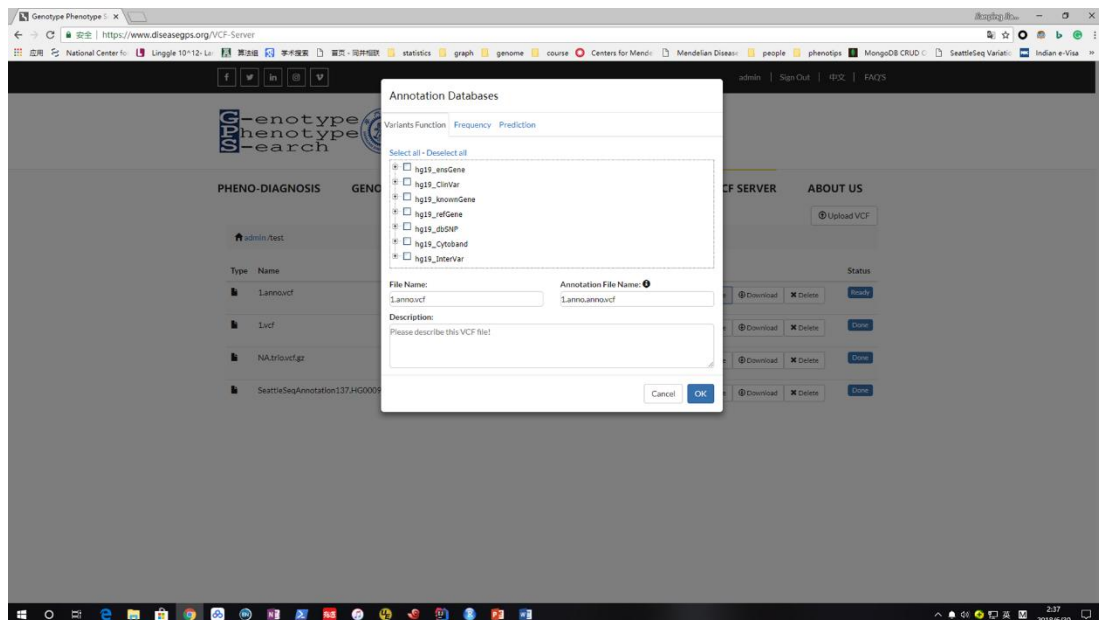
Name	Status	Operation
DP	No Index	On * Off
Zygosity	No Index	On * Off

Annotate VCF File



Select Annotation Databases

● Variants Function Databases



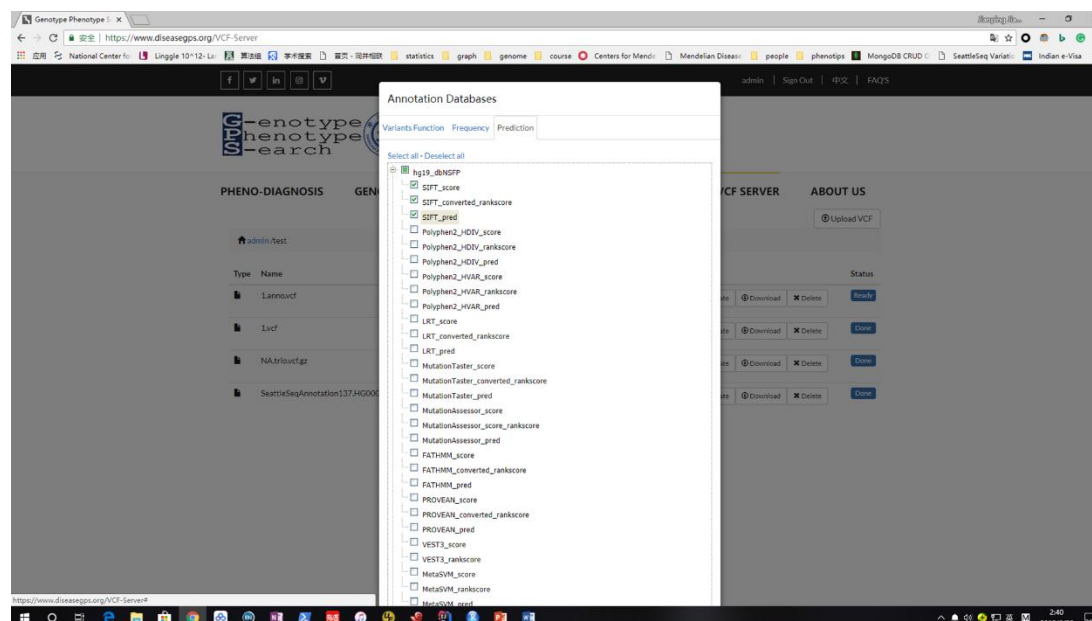
Genotype Phenotype X 1/2/2016 10:10:10 AM



Genotype Phenotype X Rungtuplin - 0 x



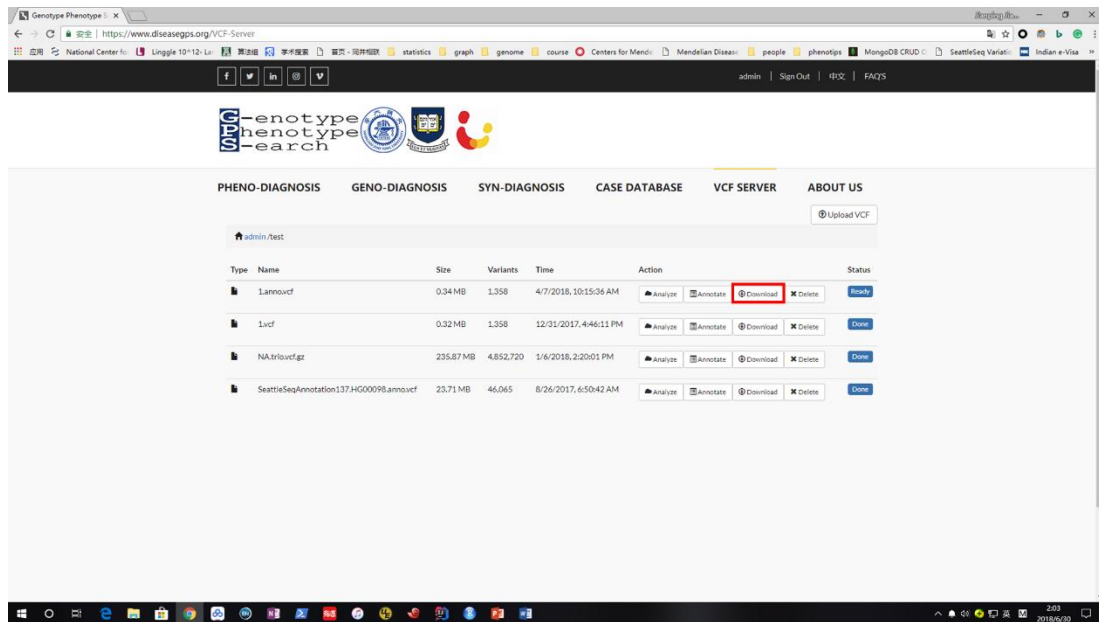
Annotate Selected Items in Databases



Information of annotation databases

Category	Database	Version/Release Date	Description
Variants Function	dbSNP	150	The Single Nucleotide Polymorphism database (dbSNP) is a public-domain archive for a broad collection of simple genetic polymorphisms.
	knownGene	201706	The University of California Santa Cruz (UCSC) Known Genes dataset is constructed by a fully automated process, based on protein data from Swiss-Prot/TrEMBL (UniProt) and the associated mRNA data from Genbank.
	ensGene	201706	The Ensembl gene set is based on evidence, and includes manual annotation for our most used species.
	refGene	201706	The RefGene database was created from the UCSC database. RefGene specifies known human protein-coding and non-protein-coding genes taken from the NCBI RNA reference sequences collection (RefSeq).
	Cytoband	201707	The chromosome band track represents the approximate location of bands seen on Giemsa-stained chromosomes.
	ClinVar	201709	ClinVar is a freely accessible, public archive of reports of the relationships among human variations and phenotypes, with supporting evidence.
	InterVar	201702	InterVar: clinical interpretation of missense variants based on ACMG standards and guidelines.
Allele Frequency	ExAC	0.3	ExAC 65000 exome allele frequency data for ALL, AFR (African), AMR (Admixed American), EAS (East Asian), FIN (Finnish), NFE (Non-finnish European), OTH (other), SAS (South Asian).
	1000Genomes	201708	1000 Genome allele frequency data for ALL, AFR (African), AMR (Admixed American), EAS (East Asian), EUR (European), SAS (South Asian).
	GME	201610	Great Middle East allele frequency including NWA (northwest Africa), NEA (northeast Africa), AP (Arabian peninsula), Israel, SD (Syrian desert), TP (Turkish peninsula) and CA (Central Asia).
Function Prediction	M-CAP	201610	M-CAP scores for non-synonymous variants.
	REVEL	201610	Rare Exome Variant Ensemble Learner scores for all possible human missense variants.
	dbNSFP	0.33a	dbNSFP is an annotation database for non-synonymous SNPs assembled by Xiaoming Liu from the University of Texas School of Public Health.

Download VCF File



Delete VCF File

