VCF-Server User Manual

（1st Edition）

2018.6

Content

Background

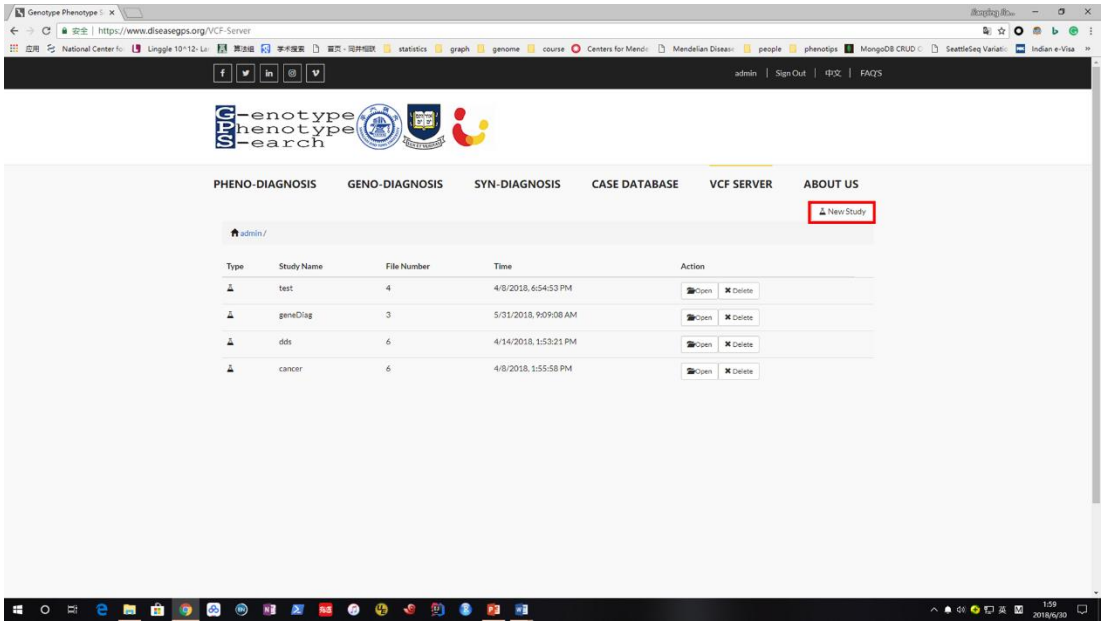
Next generation sequencing (NGS) has facilitated the discovery of disease-linked genetic variants and has been widely used for disease aetiology research and clinical genetic diagnostics. Mutations identified by NGS are usually recorded in variant call format (VCF) file, which has become the community 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.

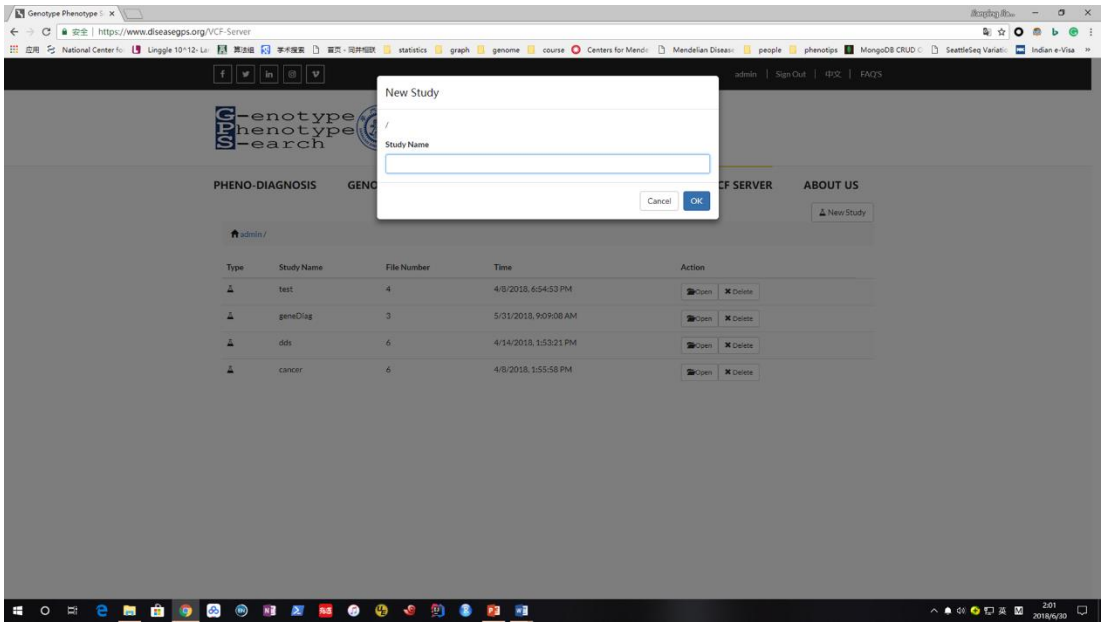
Study Management

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)

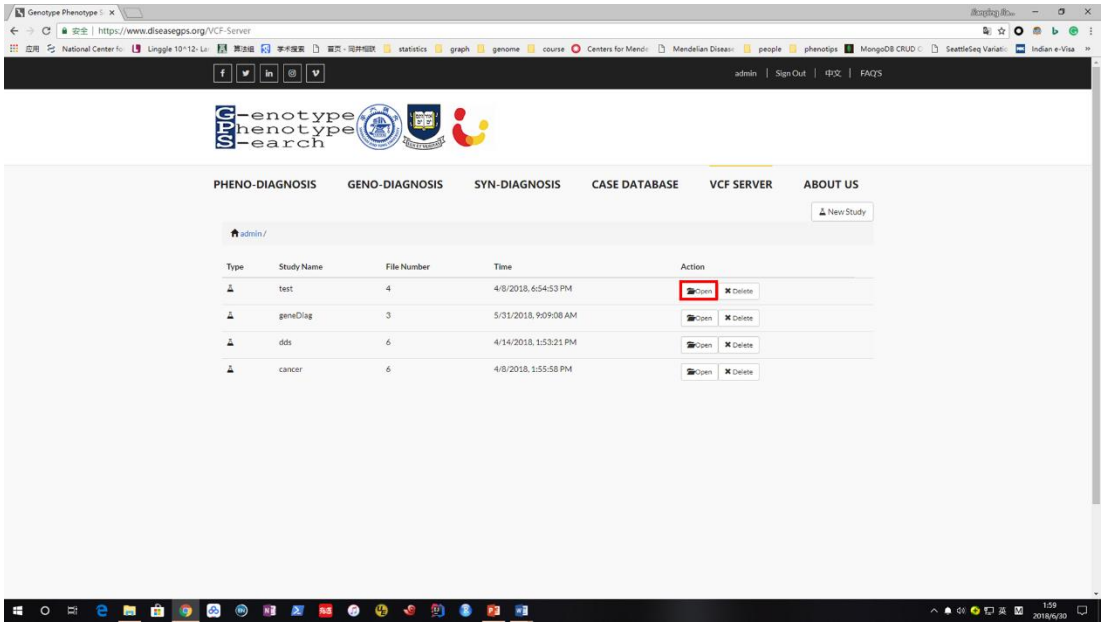


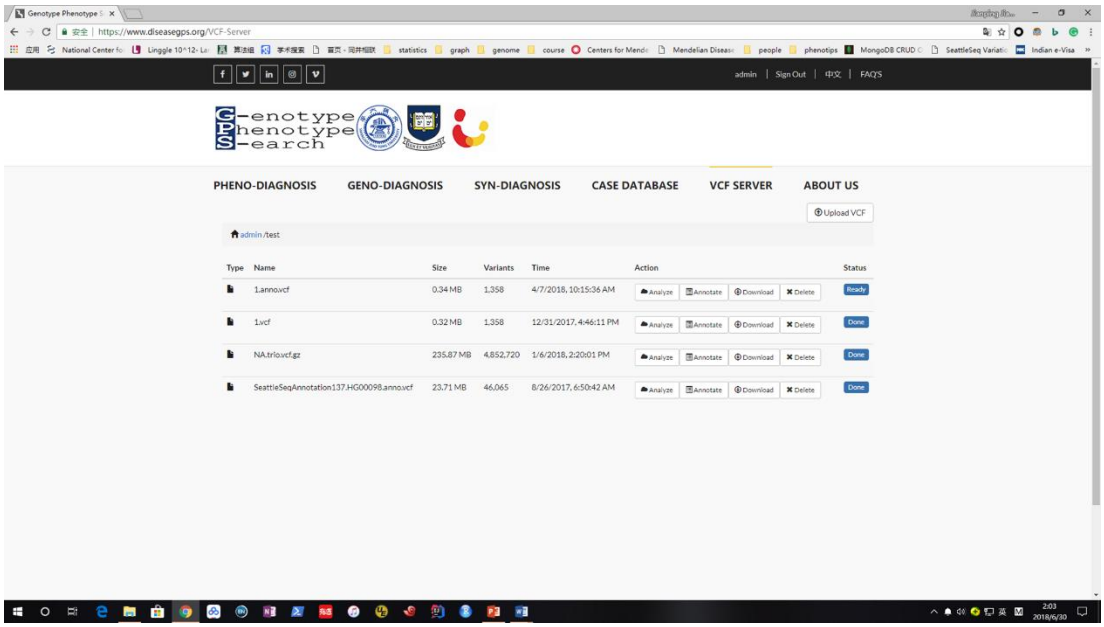
For login users, click the 'New Study' button and create a new research project.



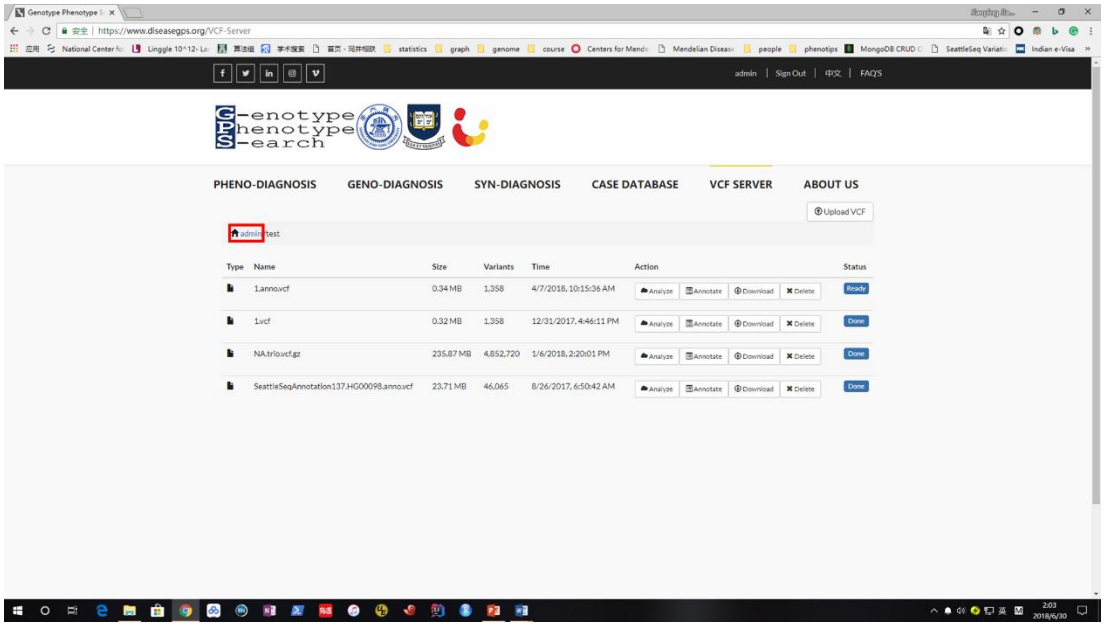
After entering project name and click 'OK' button, the new project will be showed in the project list.

Open Study

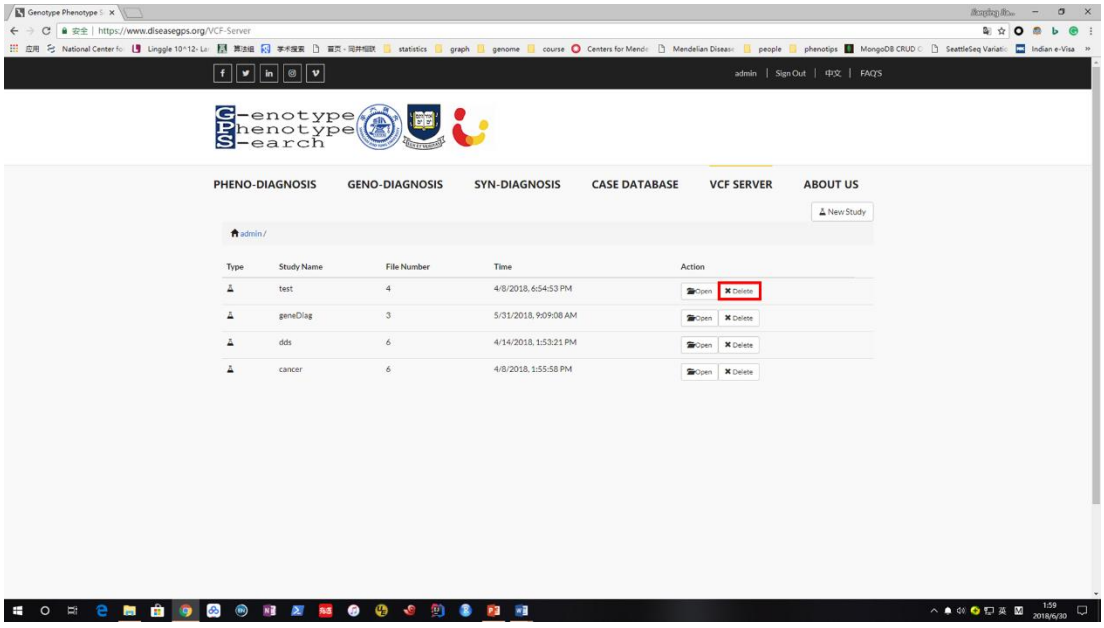

admin | Sign Out | 中文 | FAQ'S
PHENO-DIAGNOSIS
GENO-DIAGNOSIS
SYN-DIAGNOSIS
CASE DATABASE
VCF SERVER
ABOUT US
New Study
admin /
Type
Study Name
File Number
Time
Action
test
4
4/8/2018, 6:54:53 PM
geneDiag
3
5/31/2018, 9:09:08 AM
dds
6
4/14/2018, 1:53:21 PM
cancer
6
4/8/2018, 1:55:58 PM
Open
Delete


admin | Sign Out | 中文 | FAQ'S
PHENO-DIAGNOSIS
GENO-DIAGNOSIS
SYN-DIAGNOSIS
CASE DATABASE
VCF SERVER
ABOUT US
Upload VCF
admin /test
Type
Name
Size
Variants
Time
Action
Status
1.anno.vcf
0.34 MB
1,358
4/7/2018, 10:15:36 AM
Ready
1.vcf
0.32 MB
1,358
12/31/2017, 4:46:11 PM
Done
NA.trio.vcf.gz
235.87 MB
4,852,720
1/6/2018, 2:20:01 PM
Done
SeattleSeqAnnotation137.HG00098.anno.vcf
23.71 MB
46,065
8/26/2017, 6:50:42 AM
Done
Analyze
Annotate
Download
Delete

Close Study



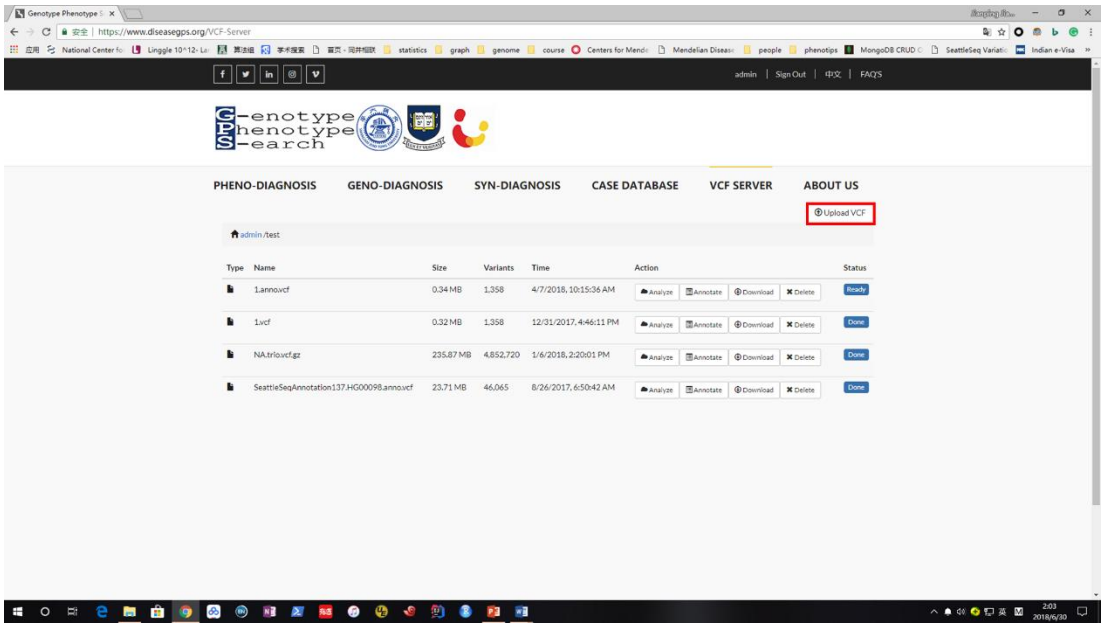
Delete Study



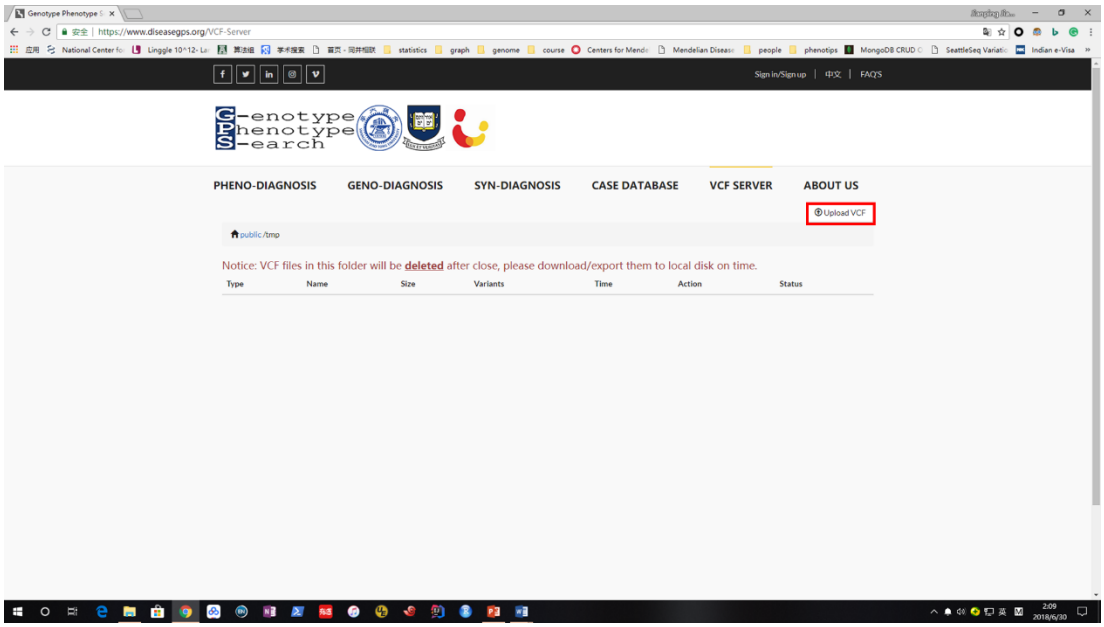
Mention that only projects without any VCF file could be delete.

Upload VCF File

- For login users

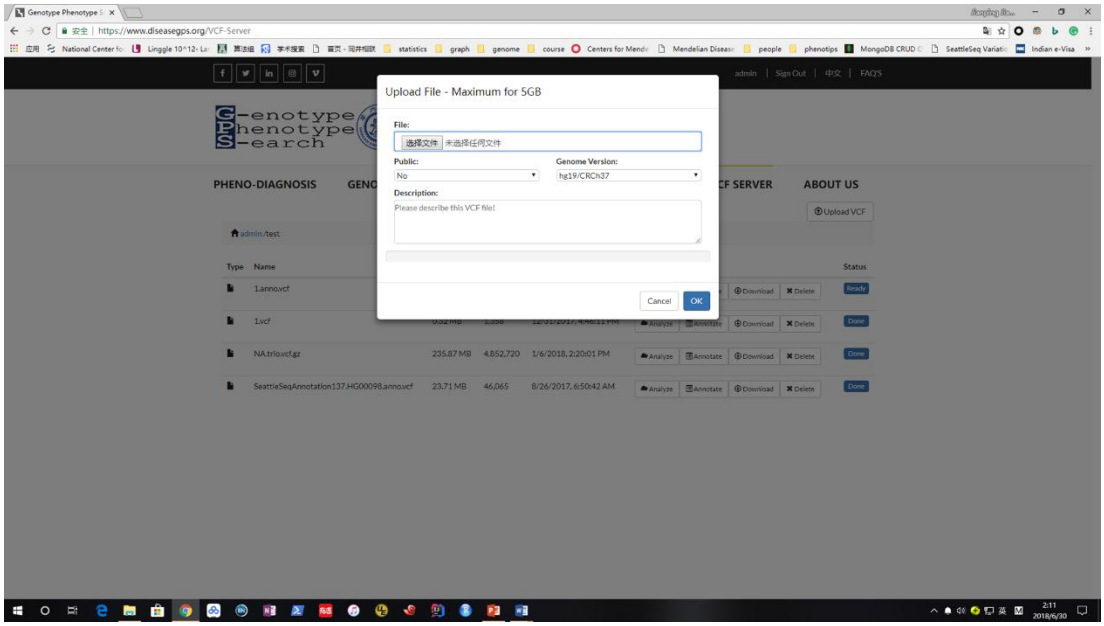


- For unlogin users



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

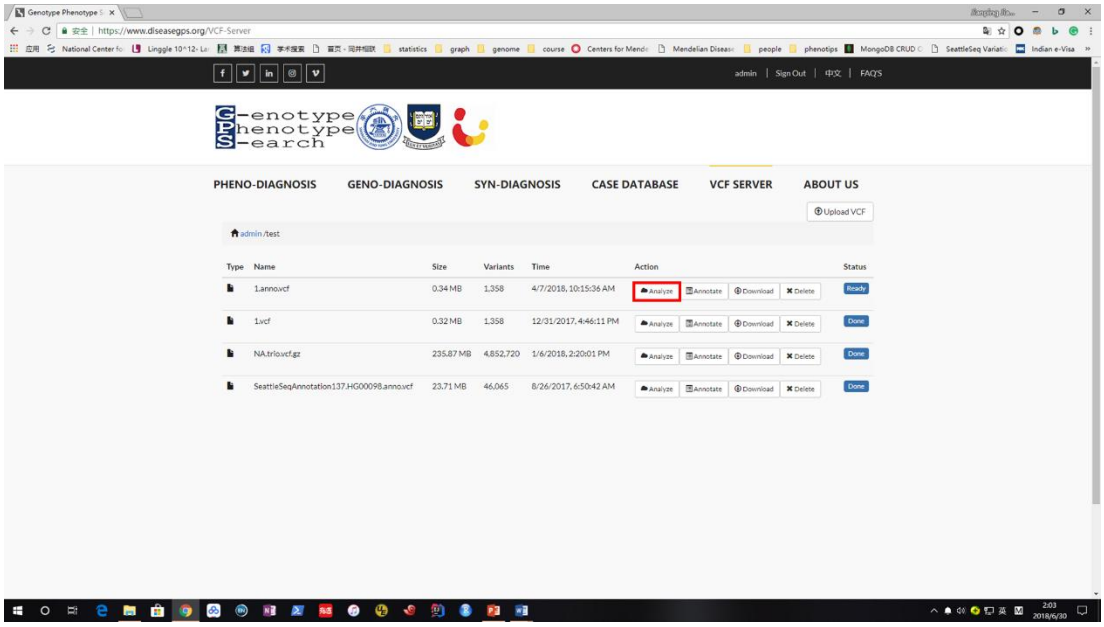
- VCF File Description

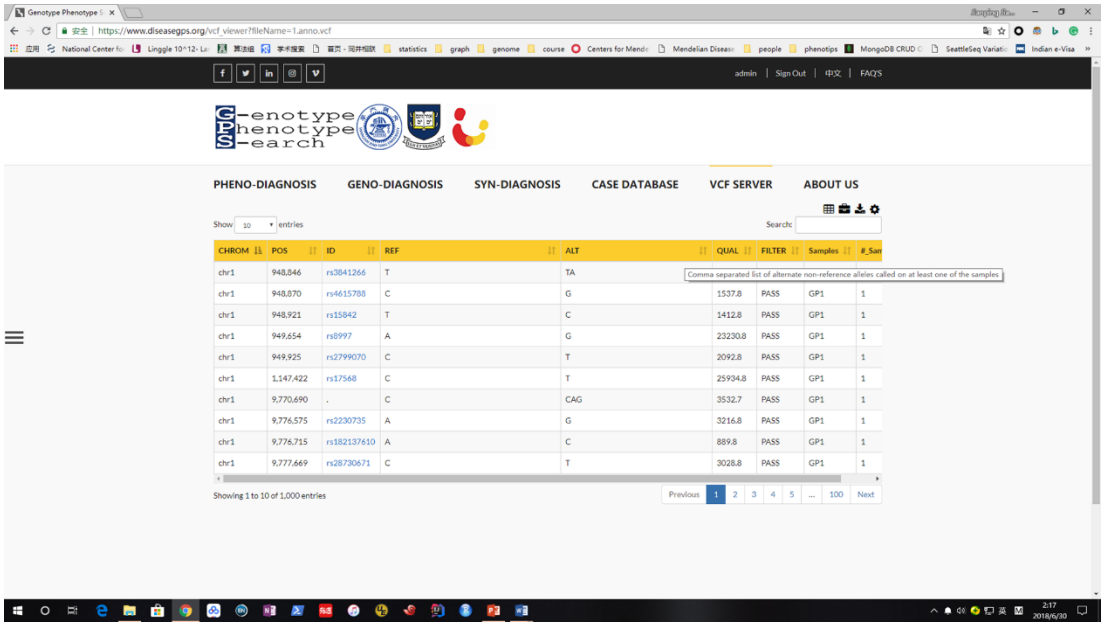


VCF File Procession

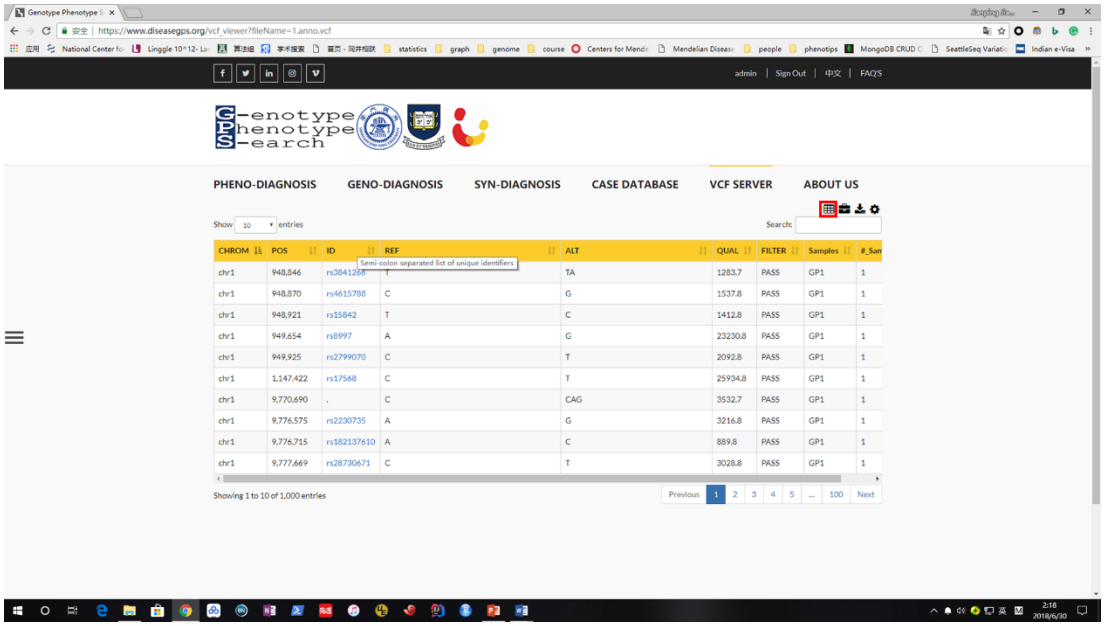
VCF-Server is a user-friendly and one-stop platform that provides rich user interfaces to assist scientists with little bioinformatics background to mine disease-linked variants in VCF files. The platform provides services on VCF file management, variants annotation, visualization and prioritization,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

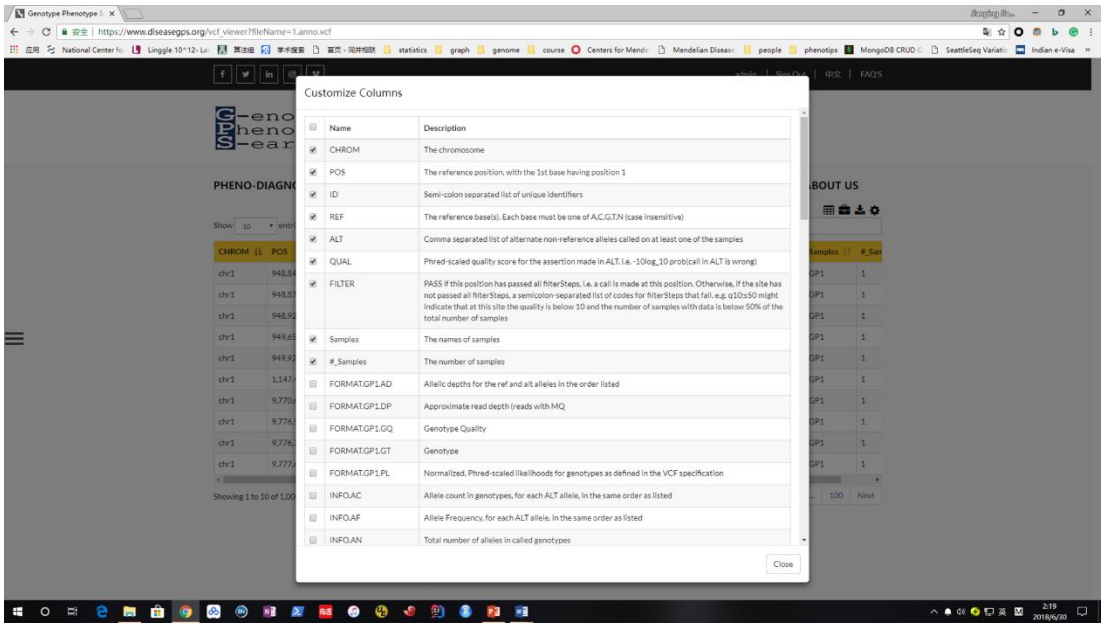
Analyze VCF File



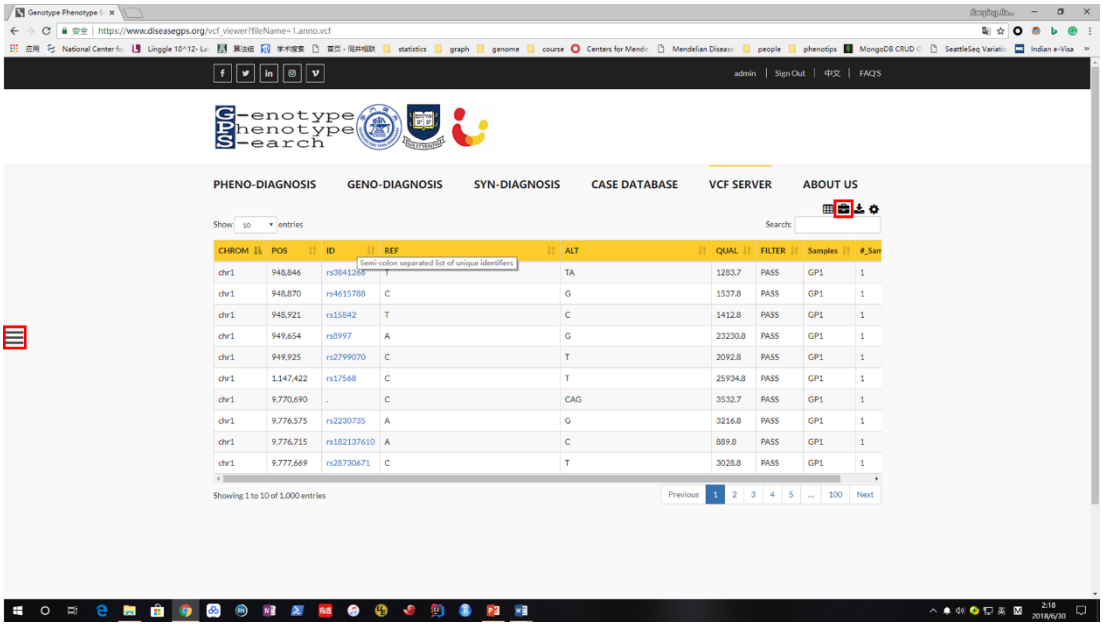


Customize Columns

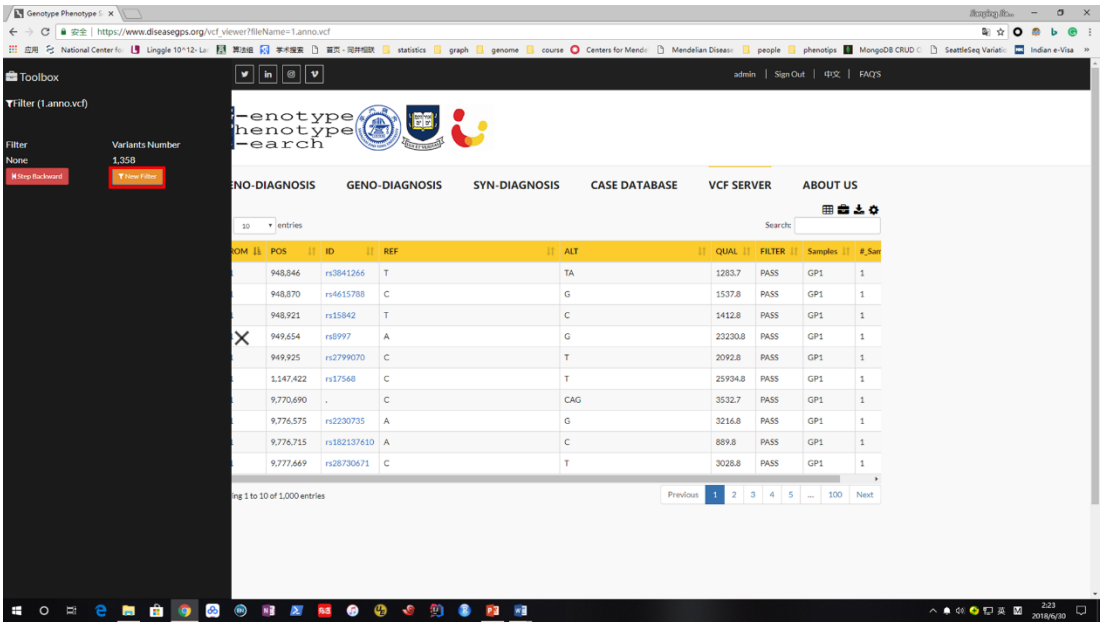




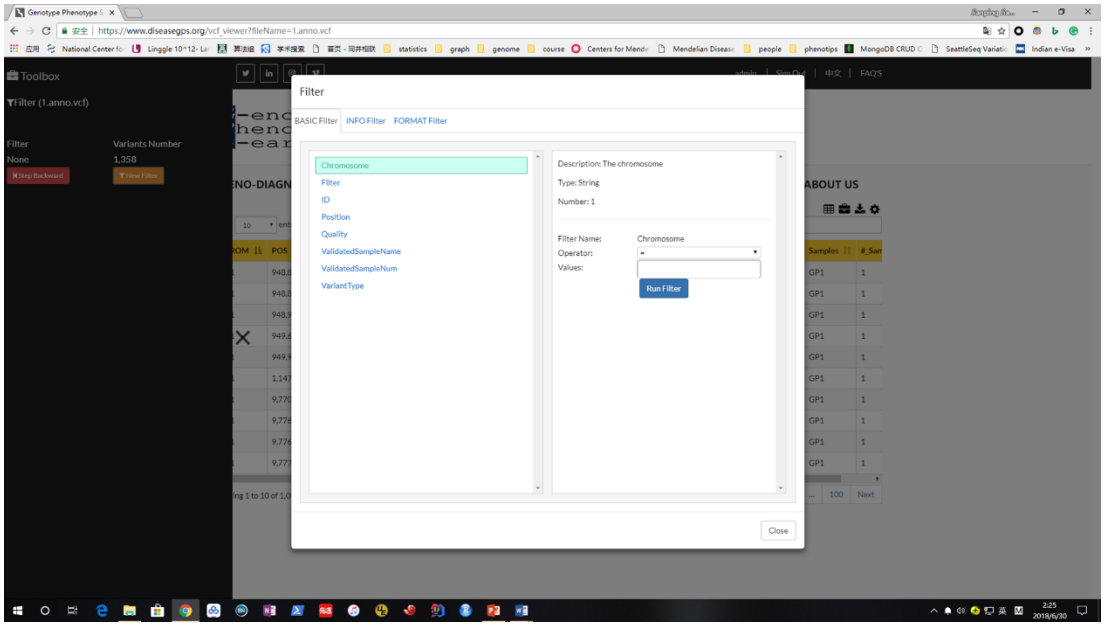
Customize Filters



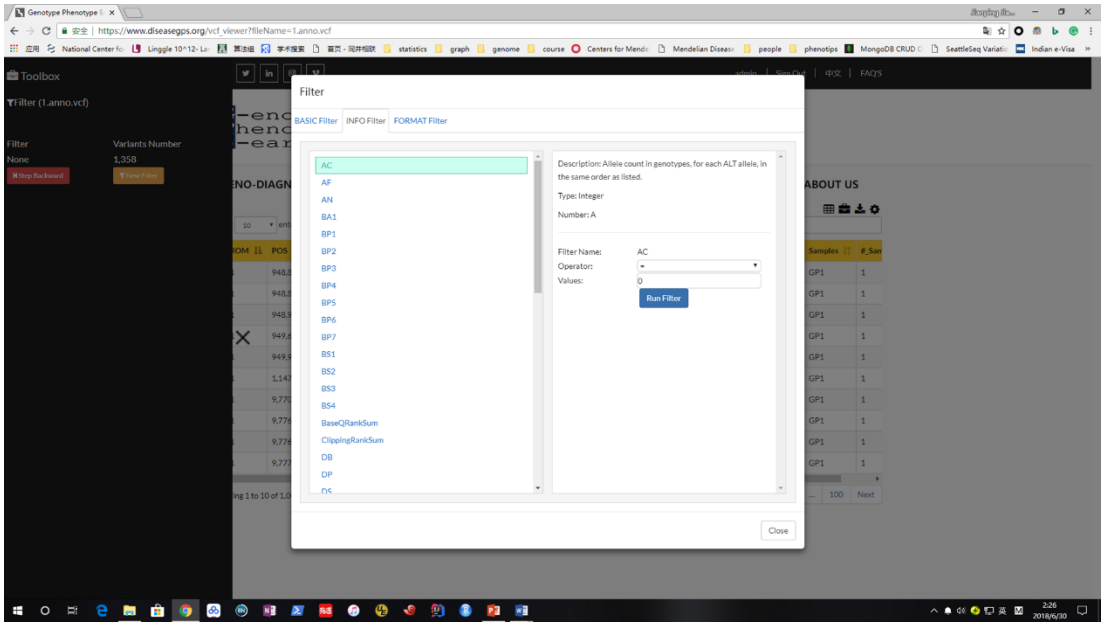
Add Filter



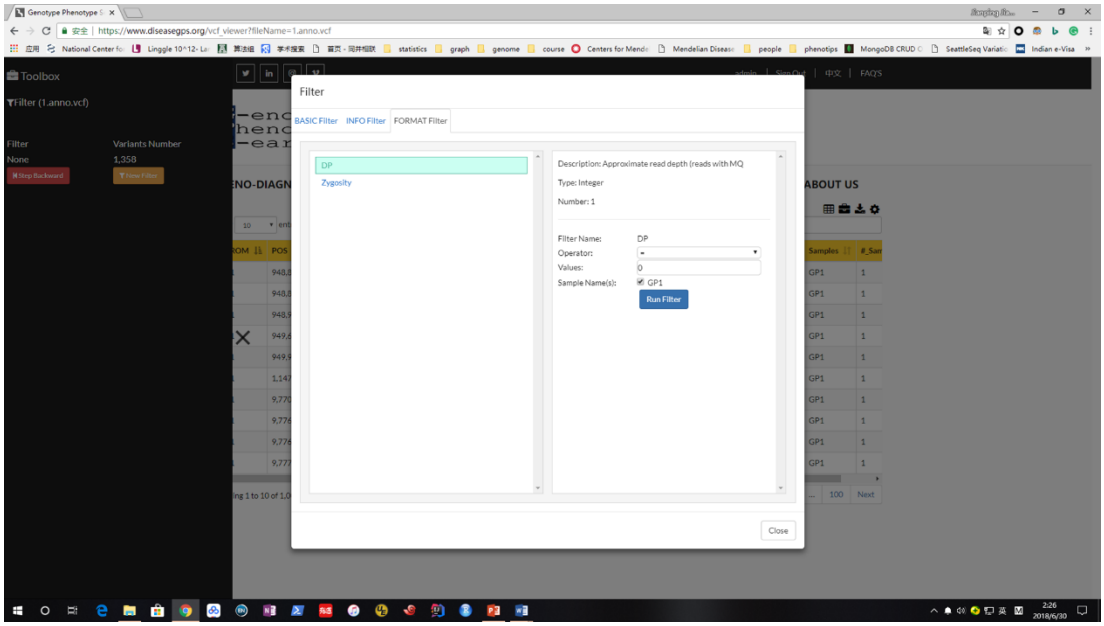
- **Basic Filter**



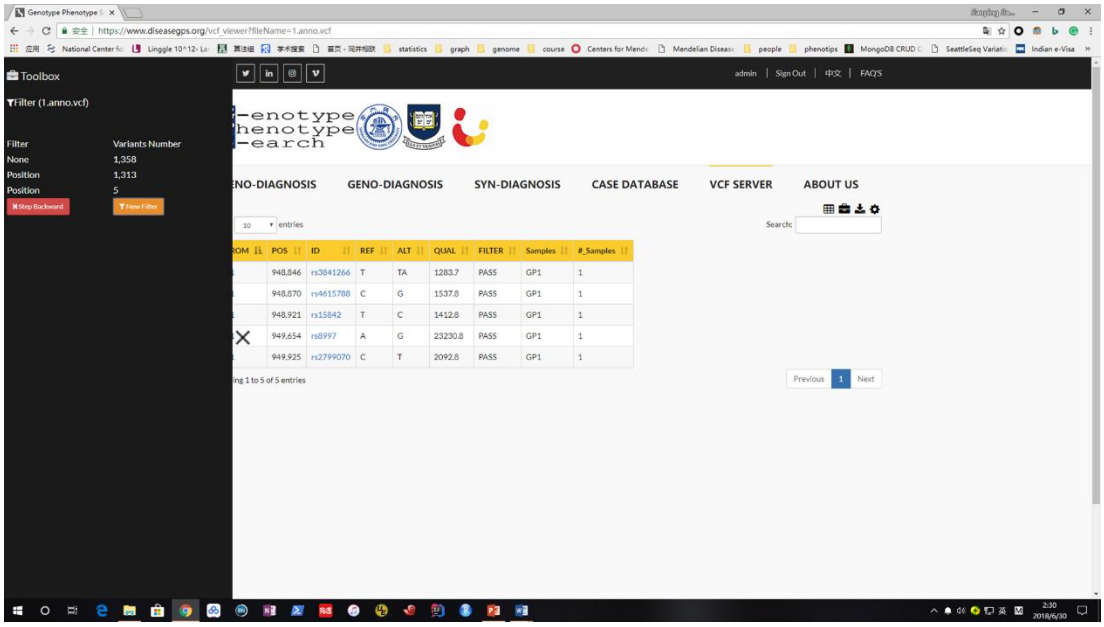
- **Information Filter**



- **Format Filter**

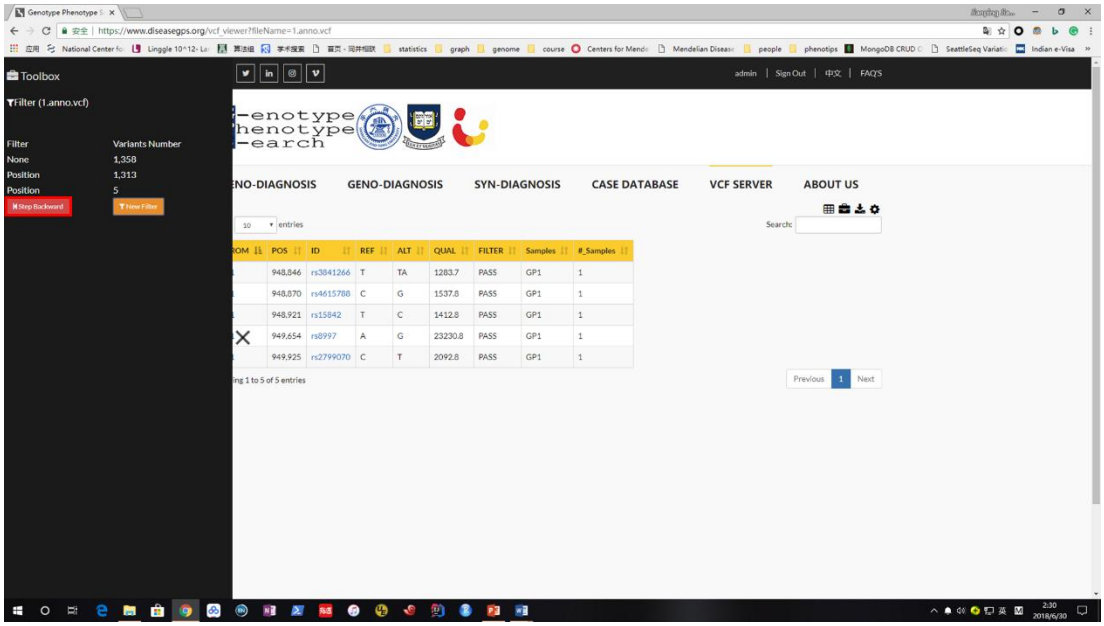


Filter Example

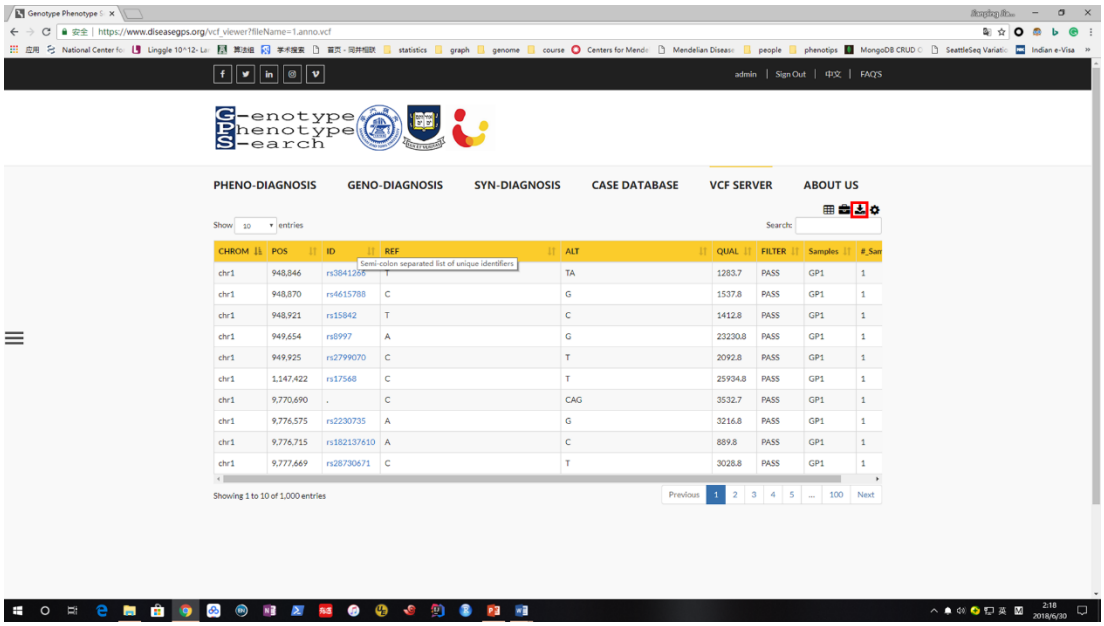


Get variants between 900K and 1M region on chromosome 1.

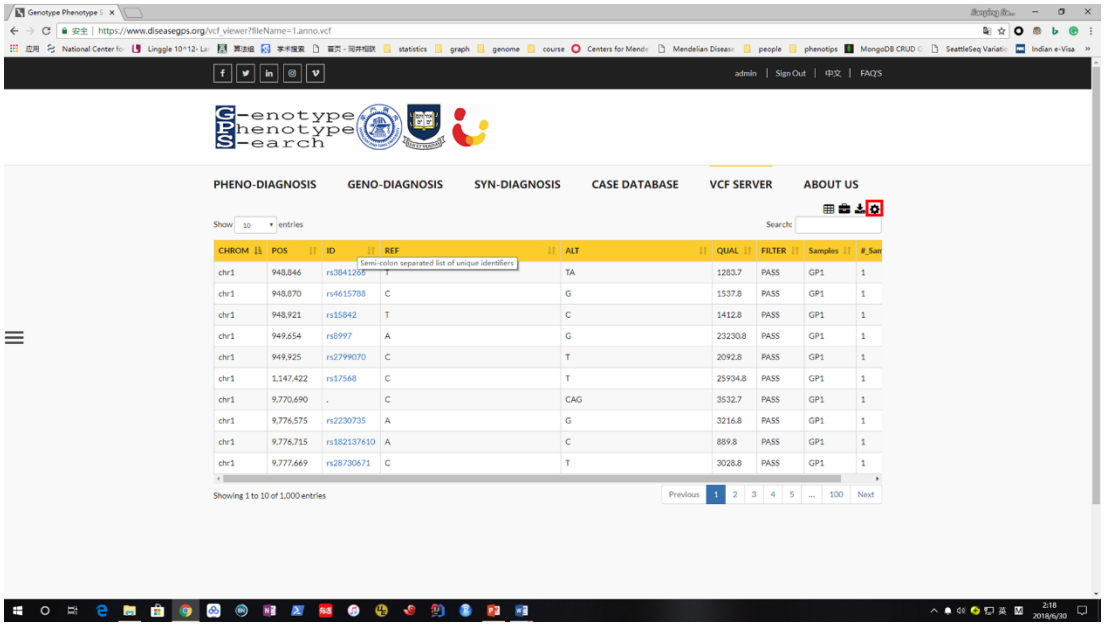
Remove Filter

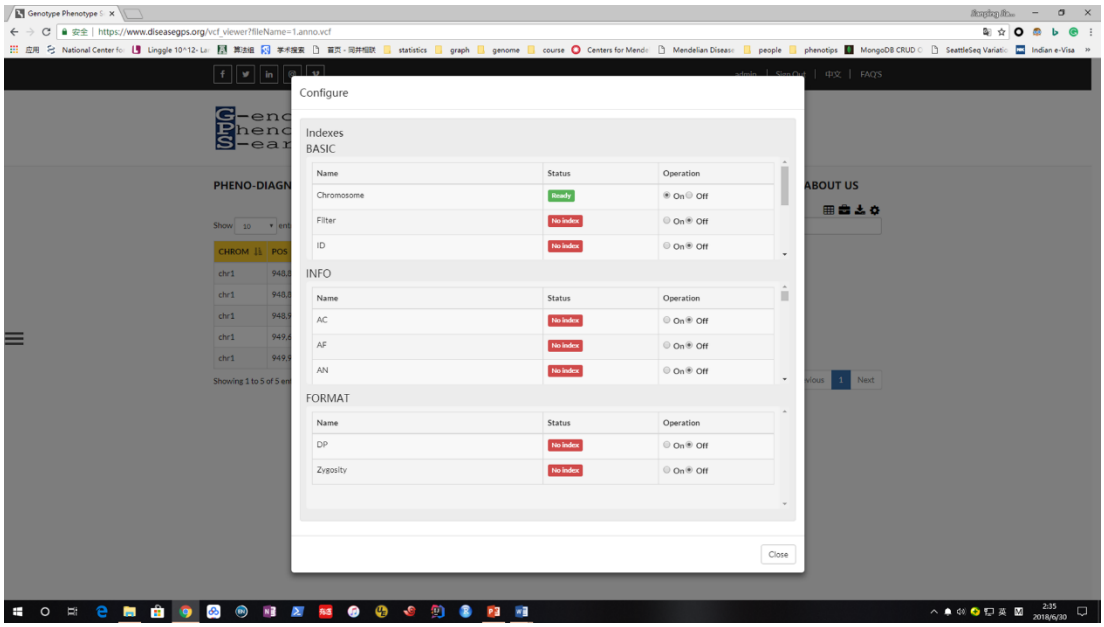


Export Filtered Variants to CSV

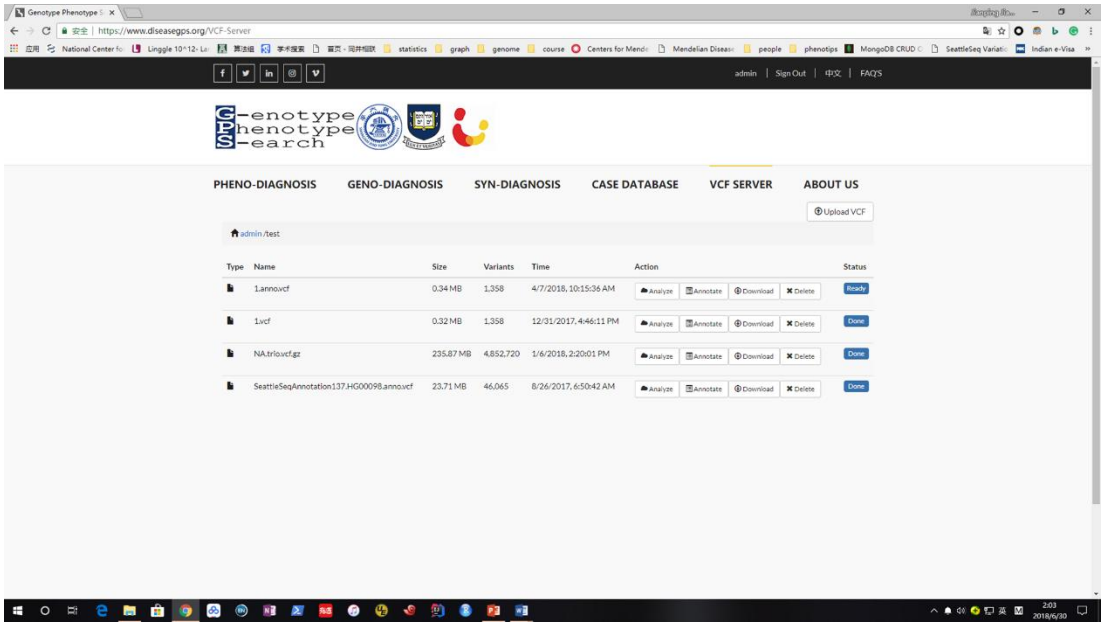


Build Columns Indexes (Speed up filtering)



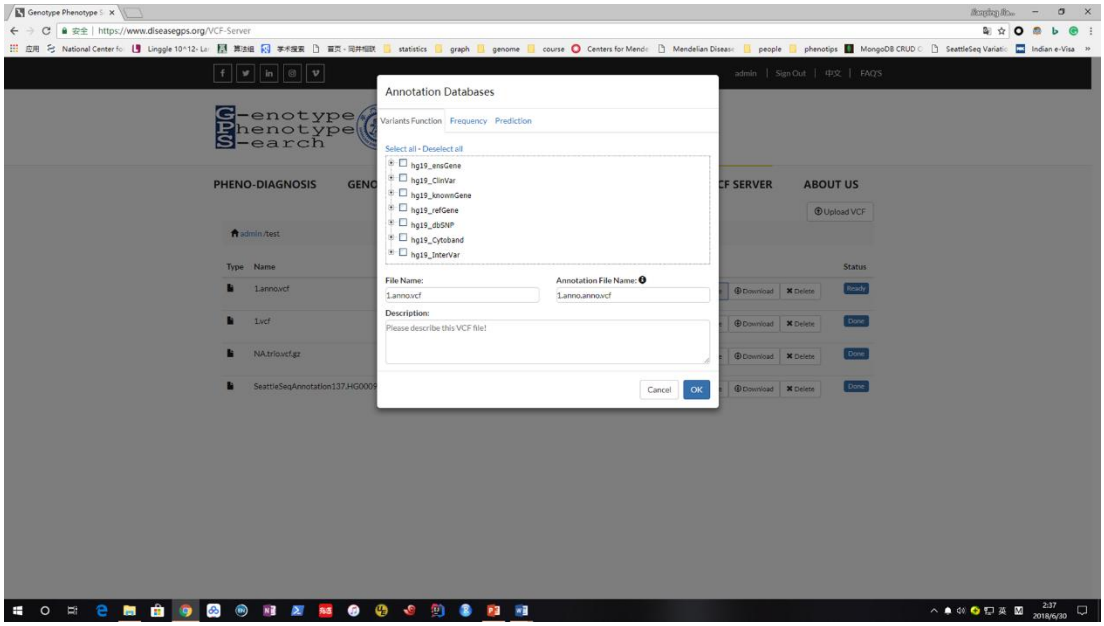


Annotate VCF File

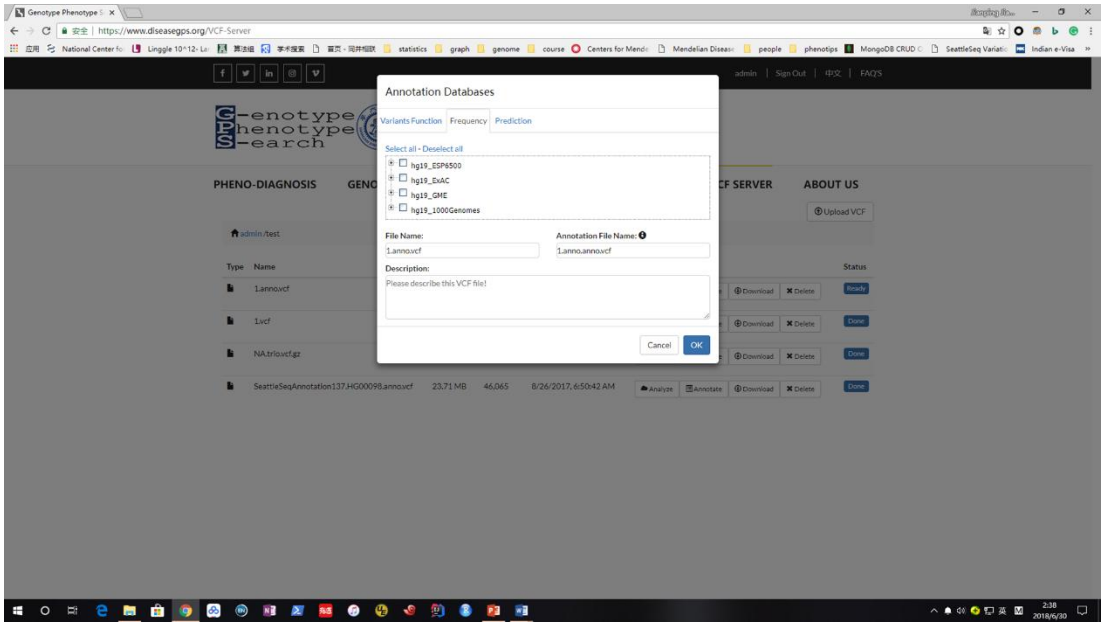


Select Annotation Databases

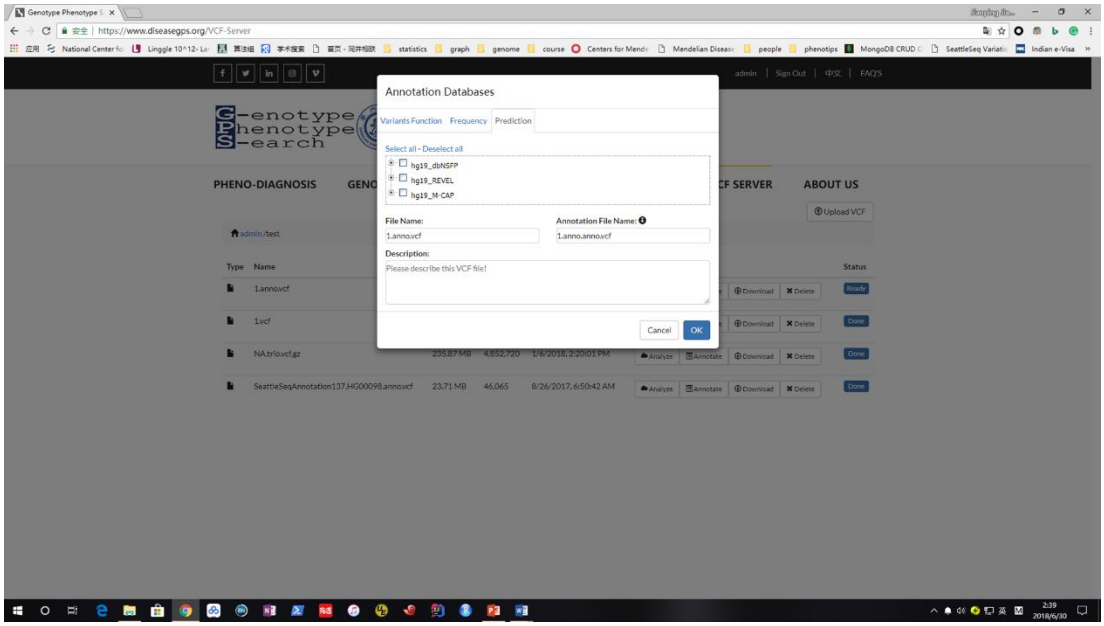
- **Variants Function Databases**



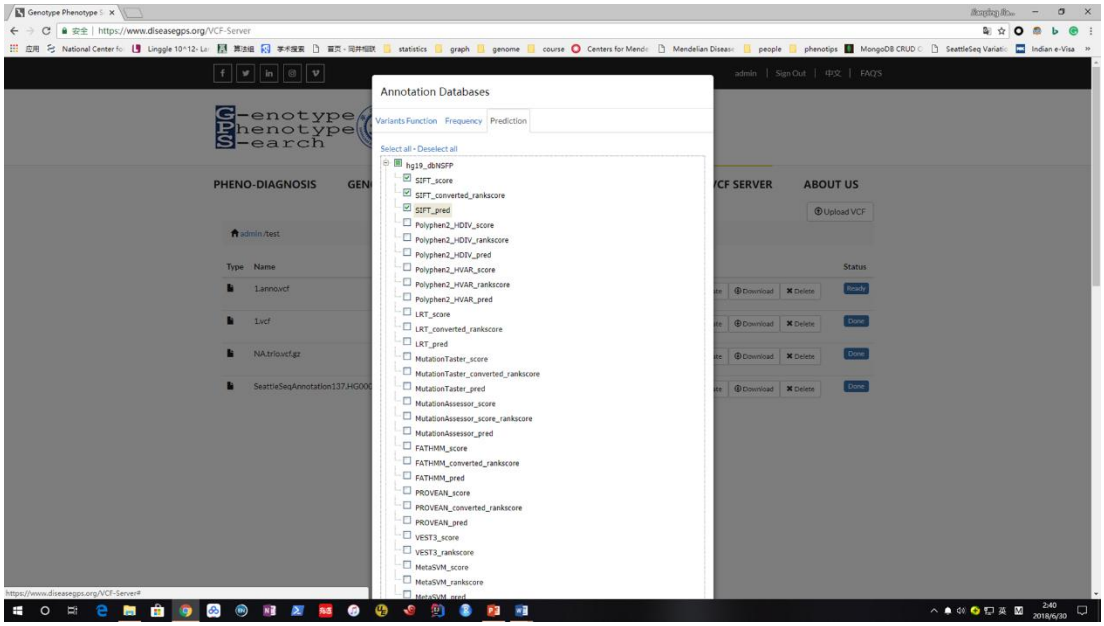
- **Population Allele Frequency Databases**



- **Variants Function Prediction Databases**



Annotate Selected Items in Databases

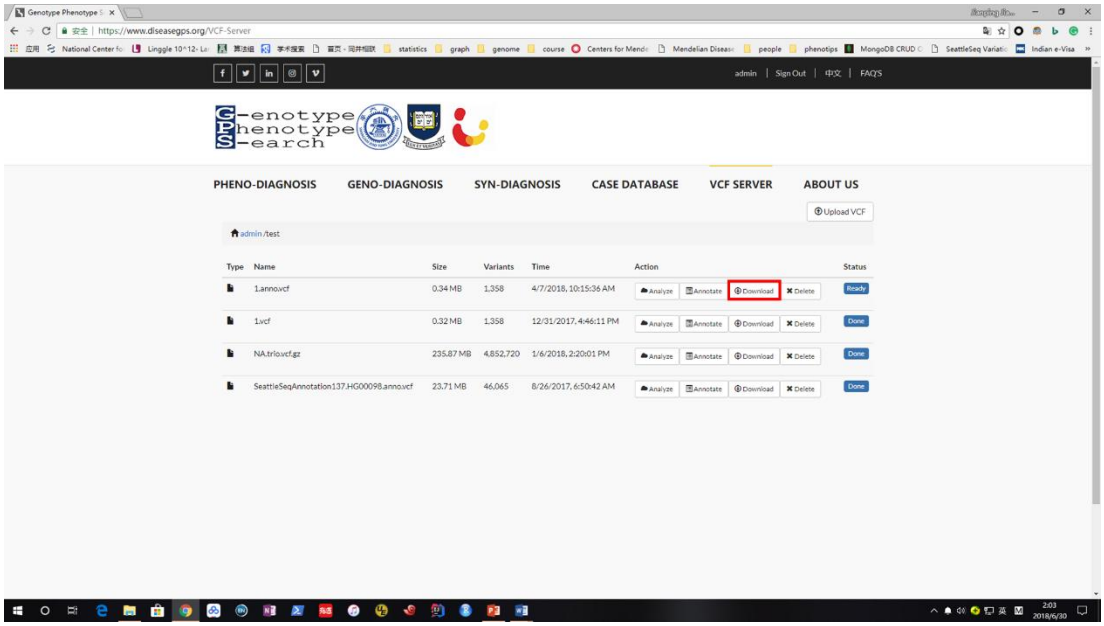


Information of annotation databases

Table S1. Summary of databases used in VCF Server annotation

Category	Database	Version/RelaseDate	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