

# **VCF-Server 用户手册**

**(第一版)**

**2018.6**

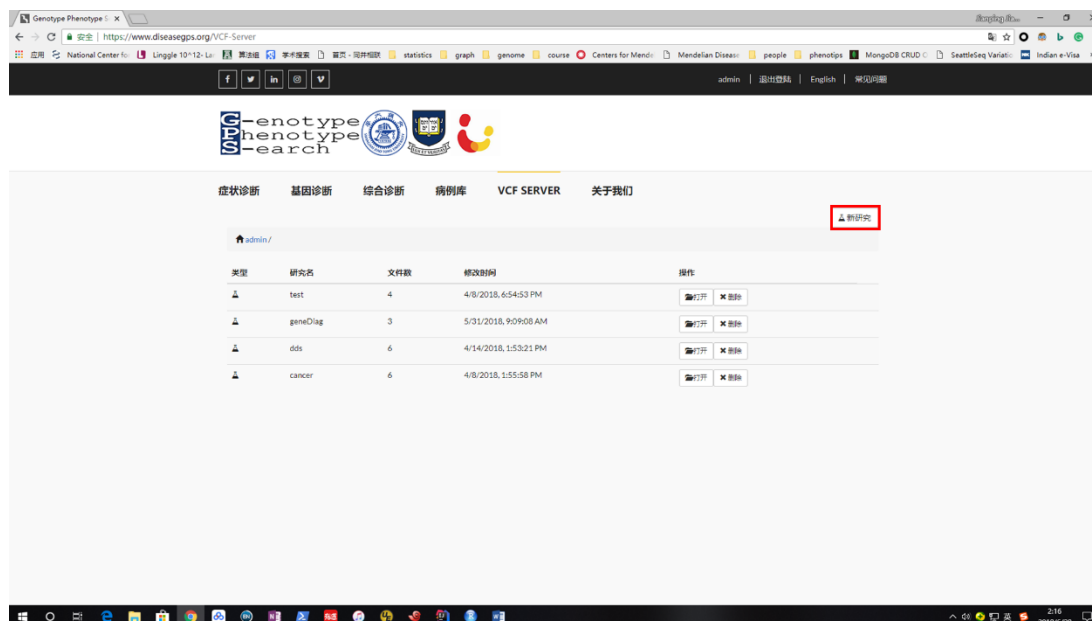
## 目录

研究项目管理 .....	3
新建研究项目（目前只对登录用户开放） .....	3
打开研究项目 .....	4
返回研究项目列表 .....	5
删除研究项目 .....	5
VCF 文件上传 .....	6
VCF 文件处理 .....	8
VCF 文件解析 .....	8
自定义显示列 .....	9
自定义筛选 .....	10
添加筛选条件 .....	10
筛选示例 .....	12
撤销筛选条件 .....	13
将筛选结果导出为 CSV .....	13
建立列索引(加快筛选速度) .....	14
VCF 文件注释 .....	15
注释数据库选择 .....	15
注释数据库中的某些条目 .....	17
注释数据库信息 .....	17
VCF 文件下载 .....	18
VCF 文件删除 .....	18

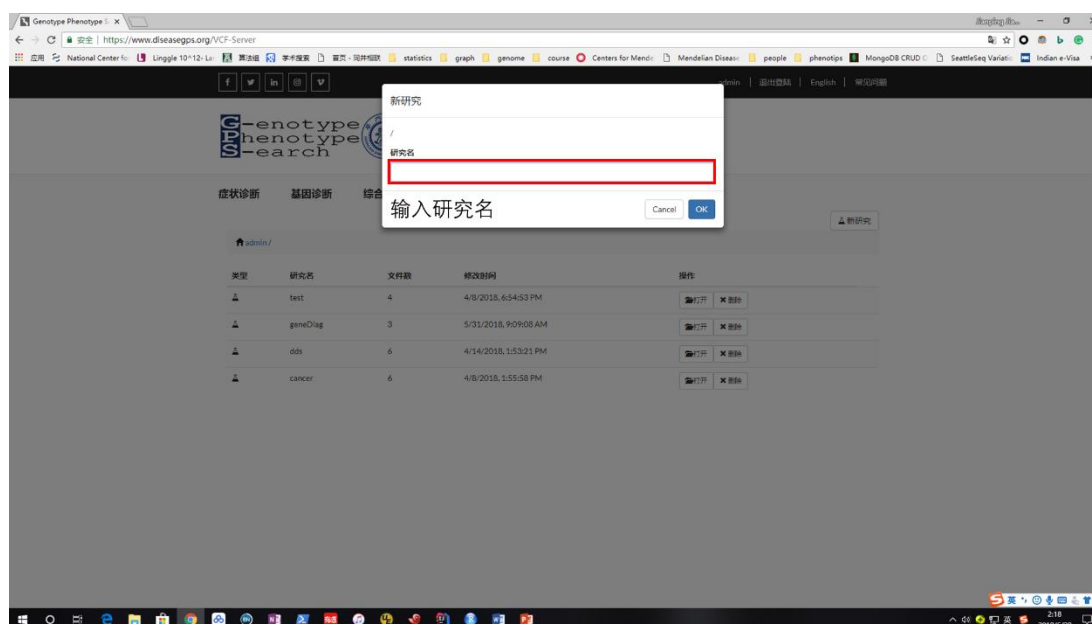
# 研究项目管理

研究项目管理是为了将不同研究项目中产生的 VCF 文件分别放置到不同的研究项目中，以便于统一有效的对 VCF 文件进行管理，这样可以大大提高 VCF 文件查找和处理的效率。

## 新建研究项目（目前只对登录用户开放）

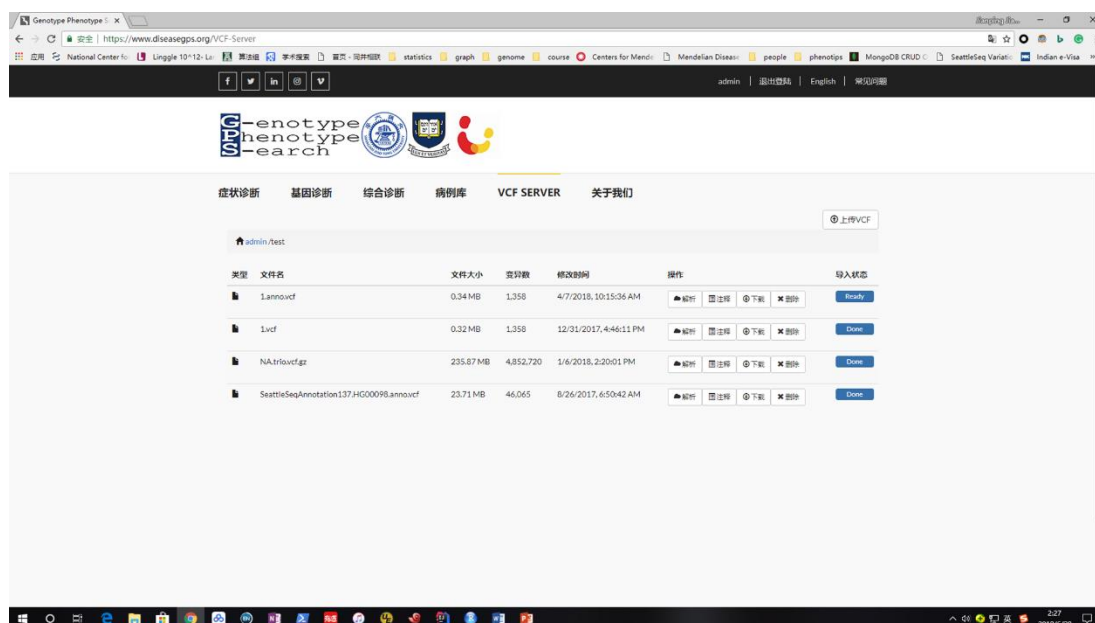
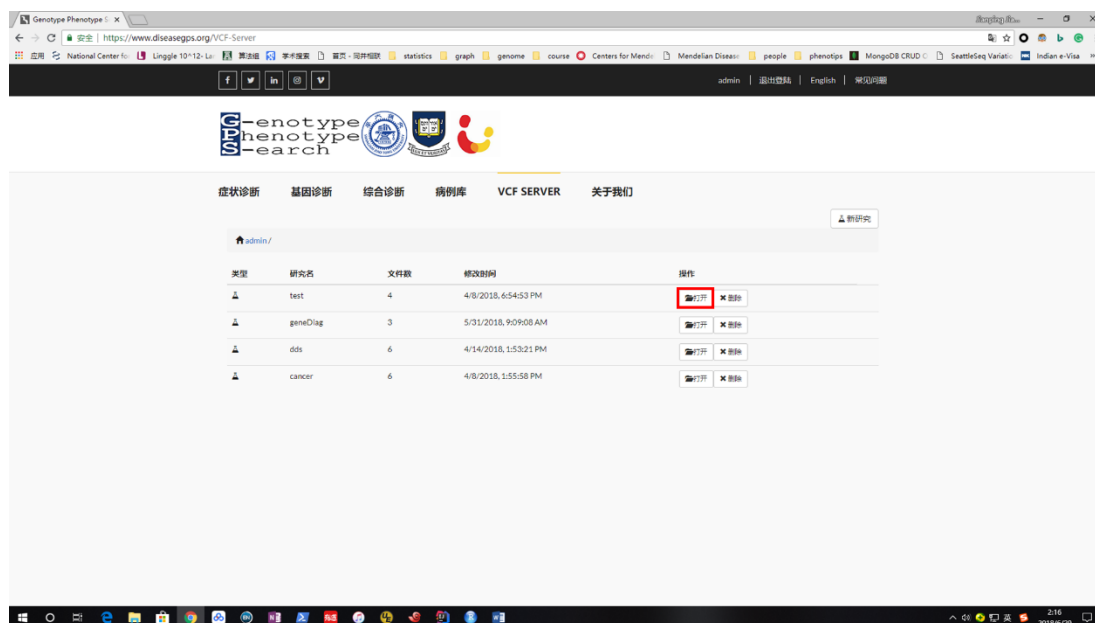


用户登陆后，点击“新研究”图标，可以新建一个研究项目。



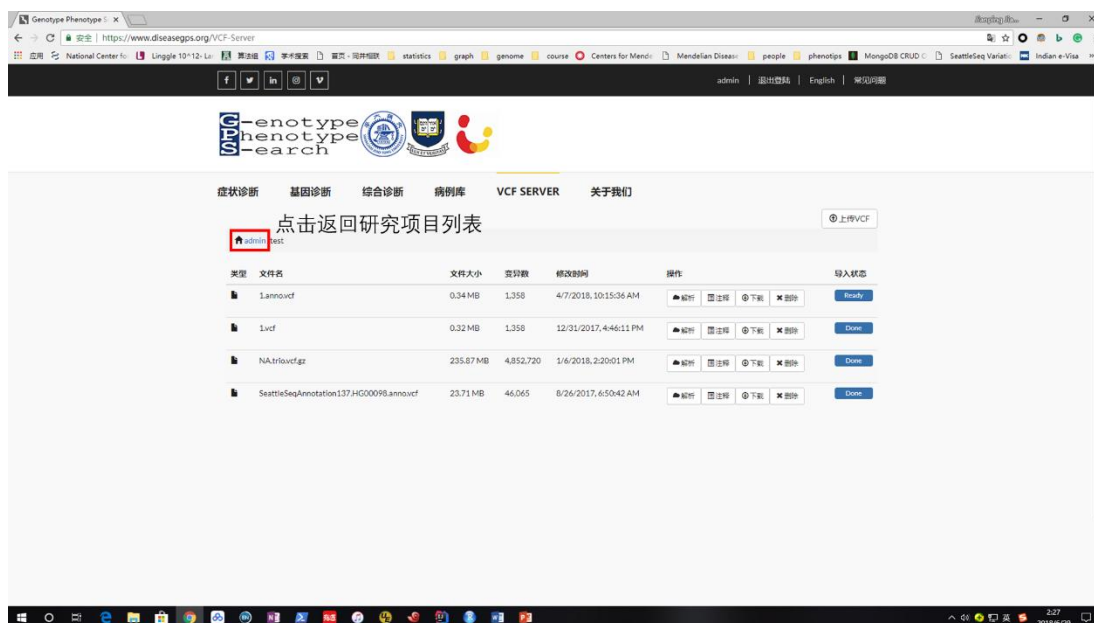
输入研究项目名并点击 OK 后，在页面中就可以看到新建的研究项目了。

## 打开研究项目

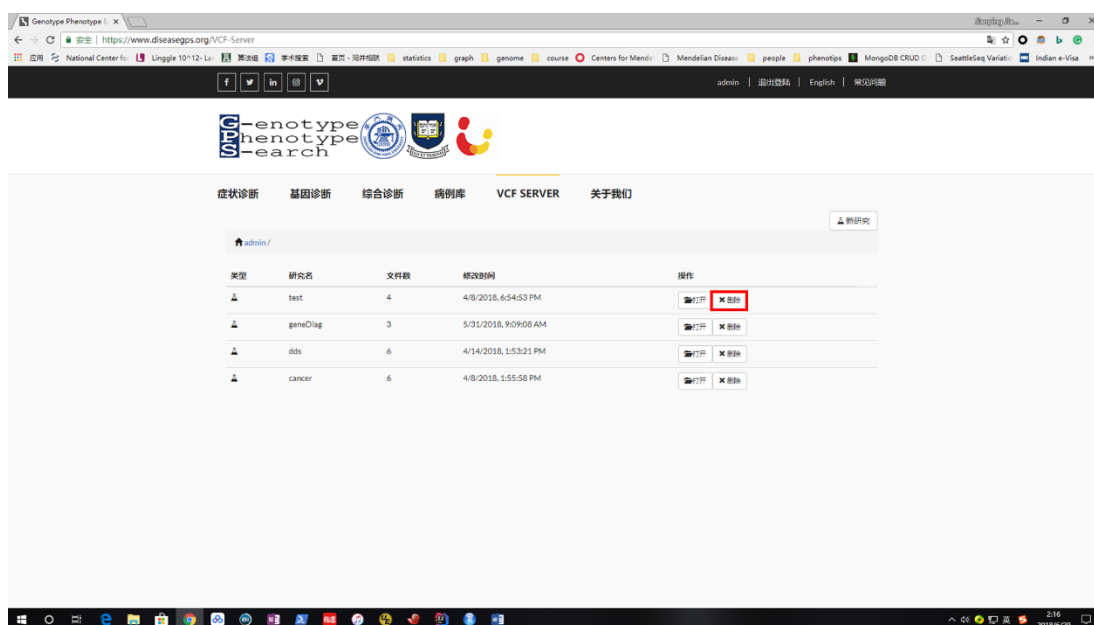


打开研究项目后，显示该项目中所有的 VCF 文件。

## 返回研究项目列表



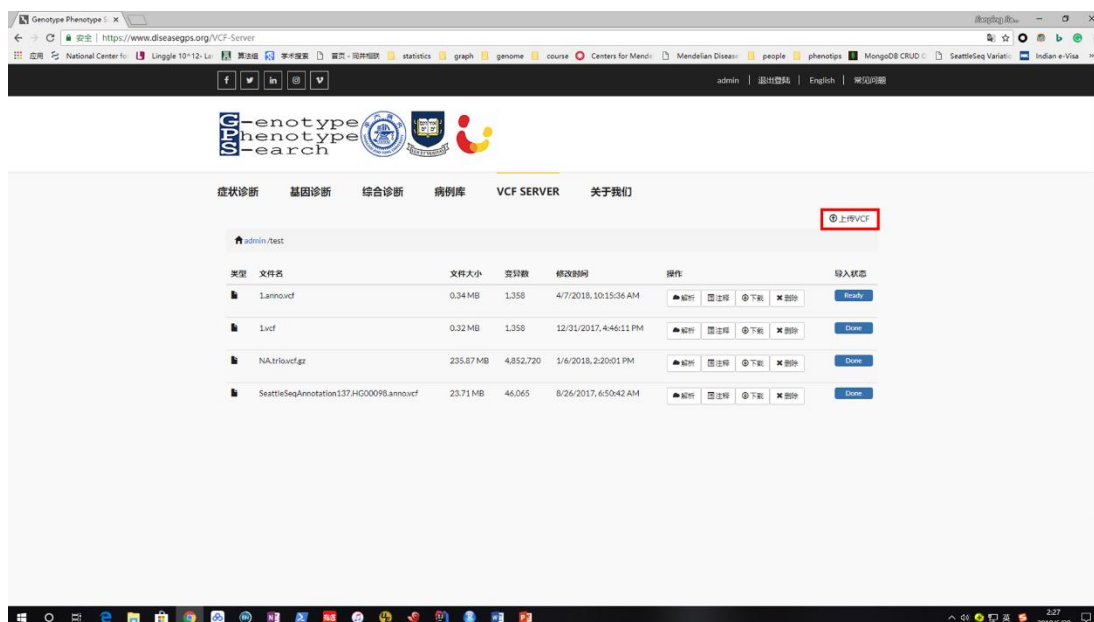
## 删除研究项目



只有当研究项目中没有文件时，才可以删除该研究项目。

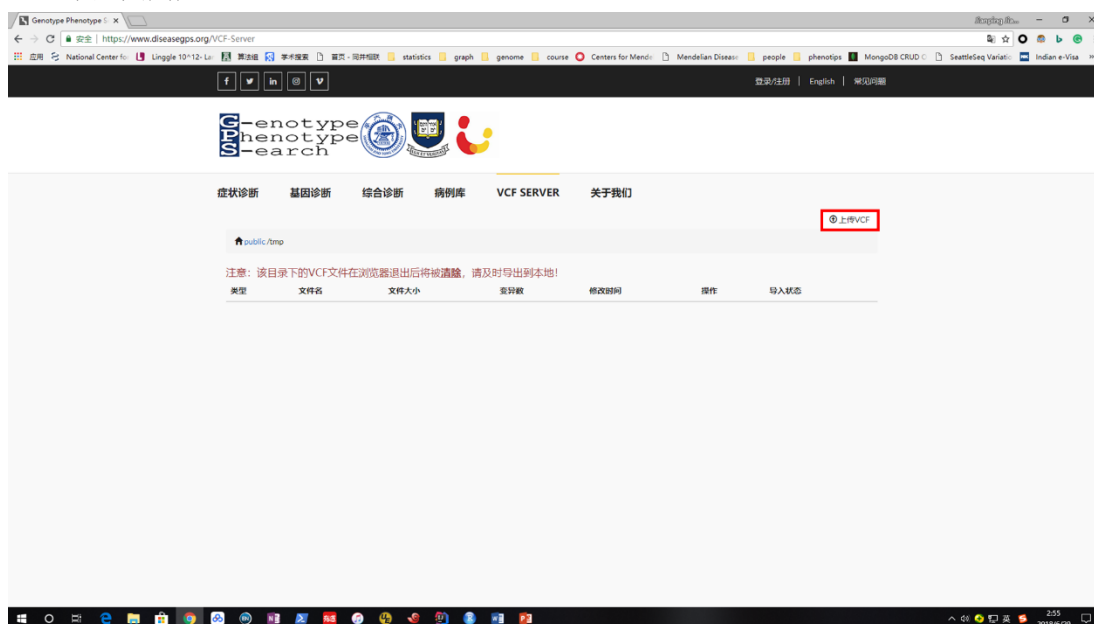
## VCF 文件上传

### ● 登陆用户



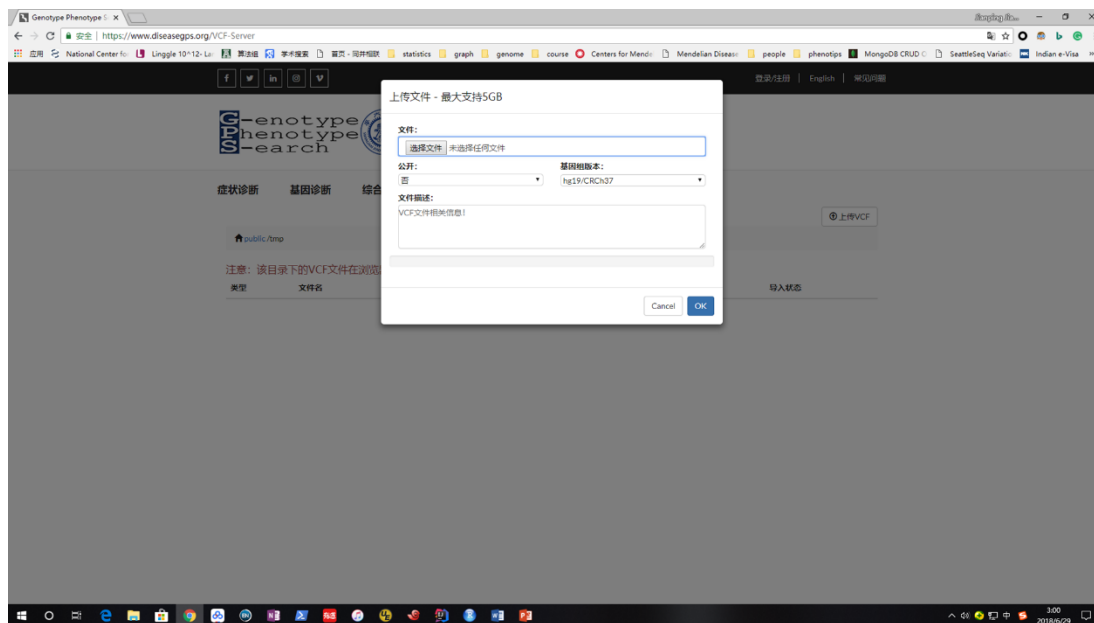
对于登陆用户，可以将 VCF 文件上传至新建的研究项目中。

### ● 未登陆用户



对于未登陆用户，只能将 VCF 文件上传至 tmp 研究项目中，并且浏览器关闭后上传的 VCF 文件及处理后的 VCF 文件会被清除。

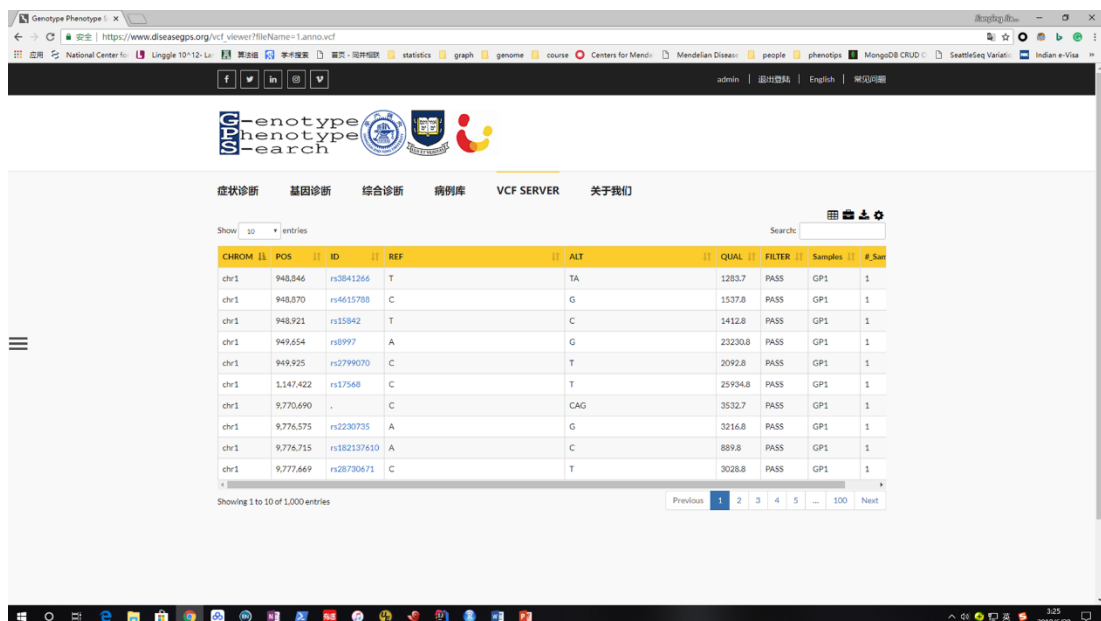
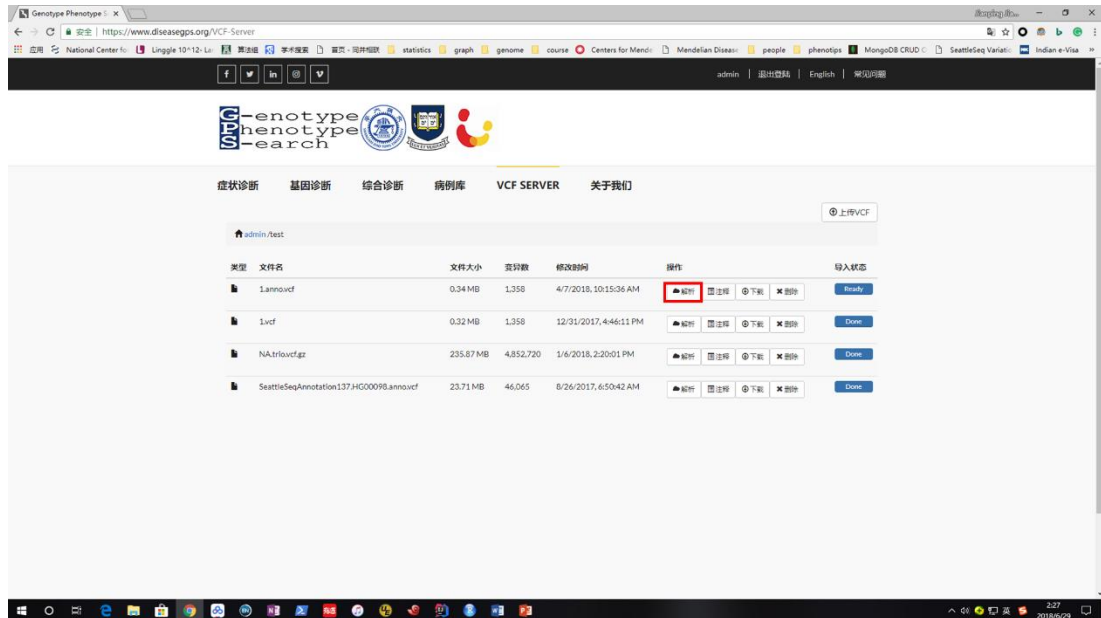
## ● 上传 VCF 文件信息



上传 VCF 文件时需要注明 VCF 使用的基因组版本号。

# VCF 文件处理

## VCF 文件解析



将 VCF 文件进行在线的解析，并显示 VCF 中的基本信息，如染色体，位置，dbSNP 的 ID 等等。



## 自定义显示列

The screenshot shows the VCF-Server web interface. At the top, there's a navigation bar with tabs for 症状诊断, 基因诊断, 综合诊断, 病例库, VCF SERVER, and 关于我们. Below the navigation bar, there's a search bar and a table of variants. The table has columns: CHROM, POS, ID, REF, ALT, QUAL, FILTER, Samples, and #\_Samples. The table displays 10 entries, showing variants across different chromosomes (chr1) with their positions, IDs, reference alleles, alternative alleles, quality scores, filter status, sample names, and the number of samples.

The screenshot shows the VCF-Server web interface with the '自定义列' (Custom Columns) dialog box open. The dialog lists various columns and their descriptions, allowing users to select which columns to display in the table. The columns listed are:

- ☐ 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. -10log<sub>10</sub> prob(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 also has a 'Close' button at the bottom right.

## 自定义筛选

Genotype Phenotype Search

症状诊断 基因诊断 综合诊断 病例库 VCF SERVER 关于我们

Show: 10 entries

Search:

CHROM	POS	ID	REF	ALT	QUAL	FILTER	Samples	#_Var
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	rs15842	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

## 添加筛选条件

Genotype Phenotype Search

症状诊断 基因诊断 综合诊断 病例库 VCF SERVER 关于我们

筛选 (1.anno.vcf)

筛选器: None 变异位点数: 1,358

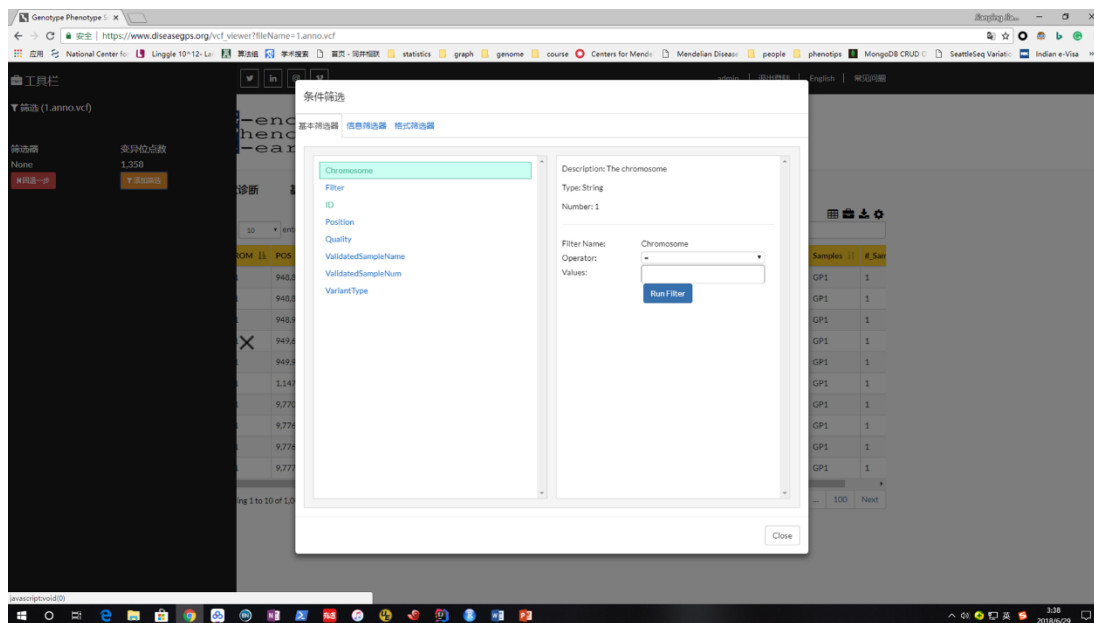
筛选

CHROM	POS	ID	REF	ALT	QUAL	FILTER	Samples	#_Var
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	rs15842	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

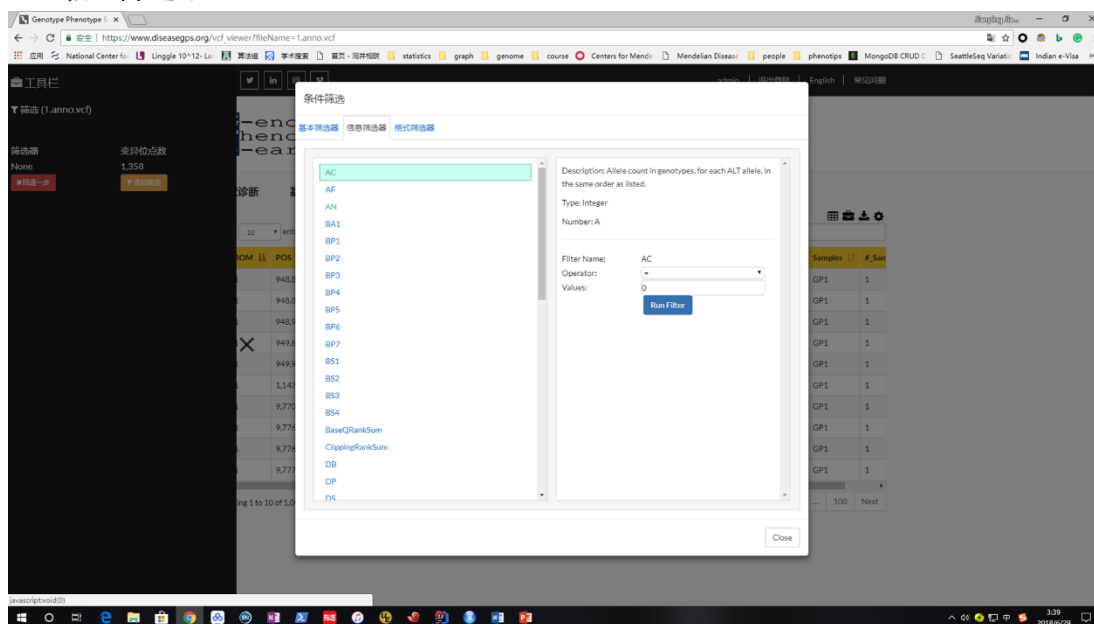
Showing 1 to 10 of 1,000 entries

Previous 1 2 3 4 5 100 Next

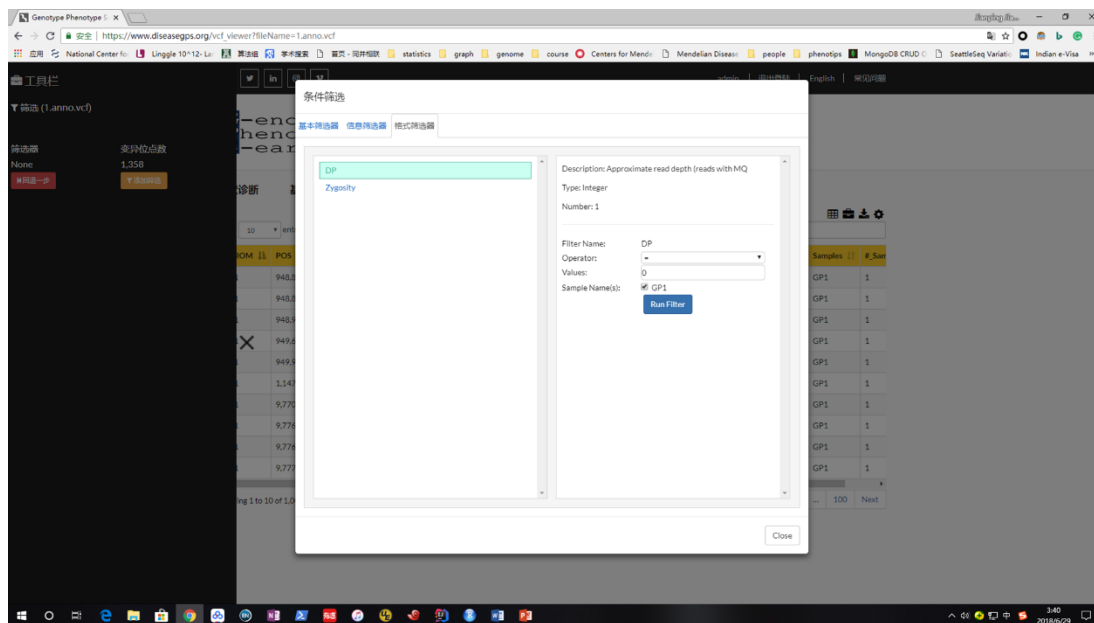
## ● 基本筛选器



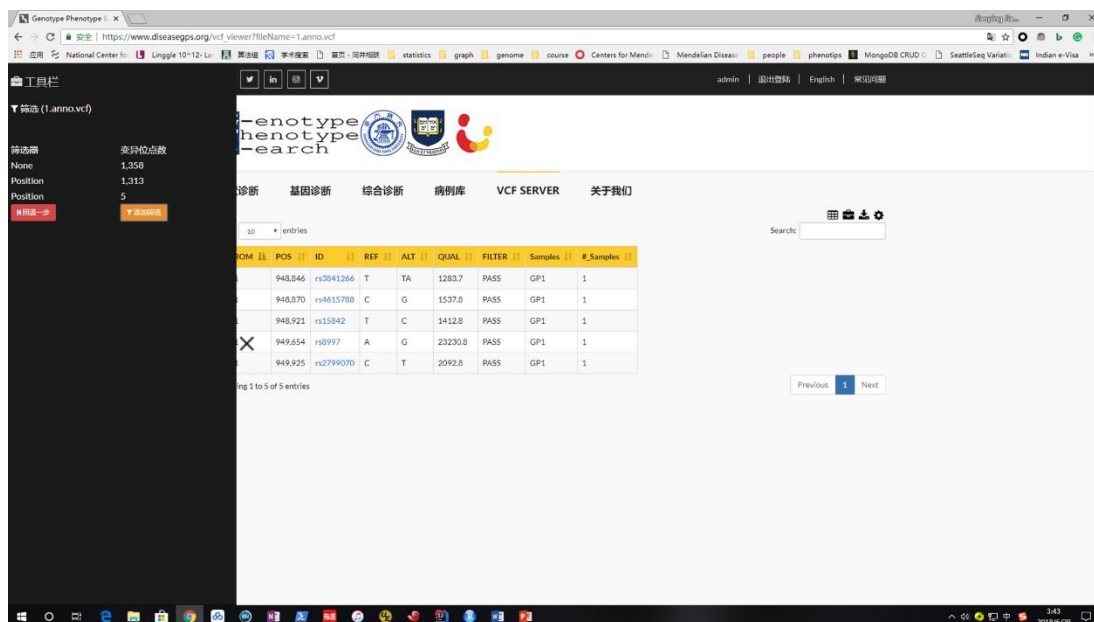
## ● 信息筛选器



## ● 格式筛选器

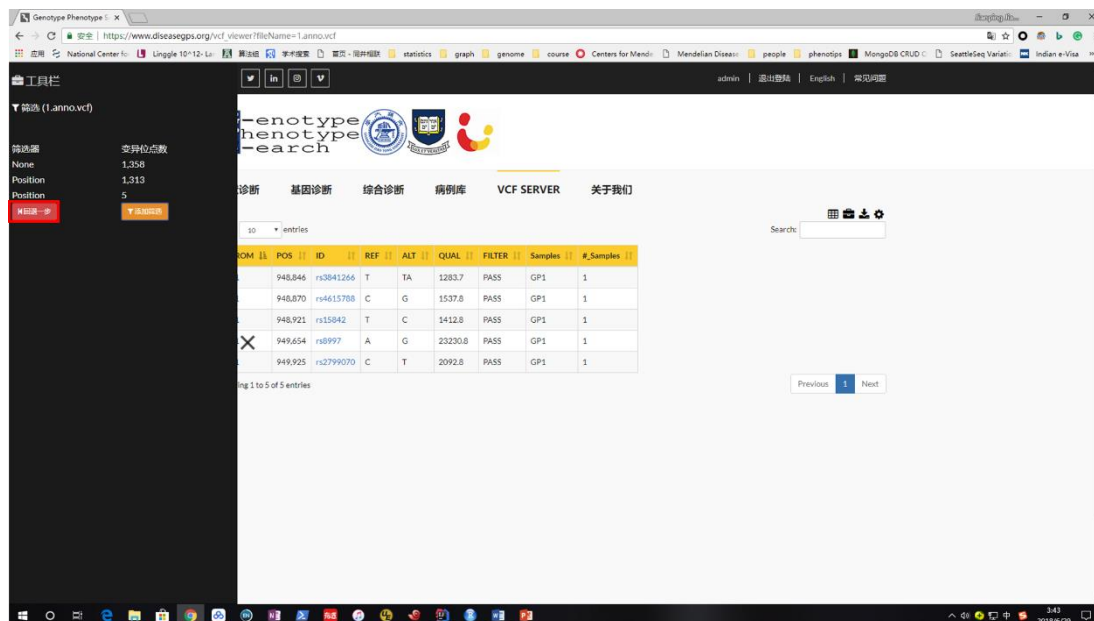


## 筛选示例

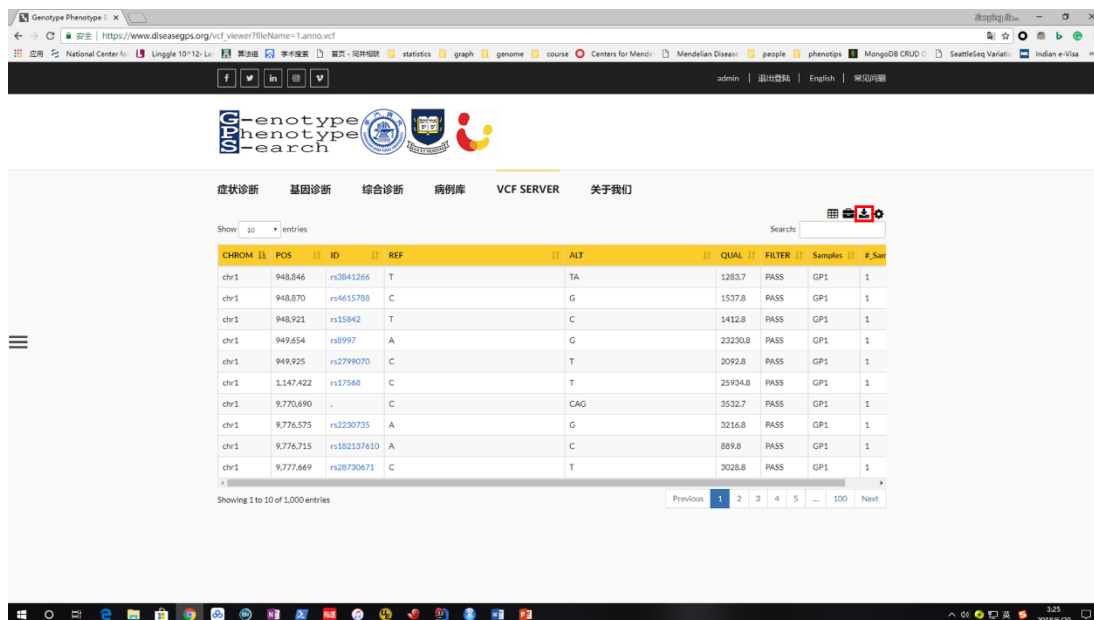


筛选基因组位置 900K-1M 区域内的位点。

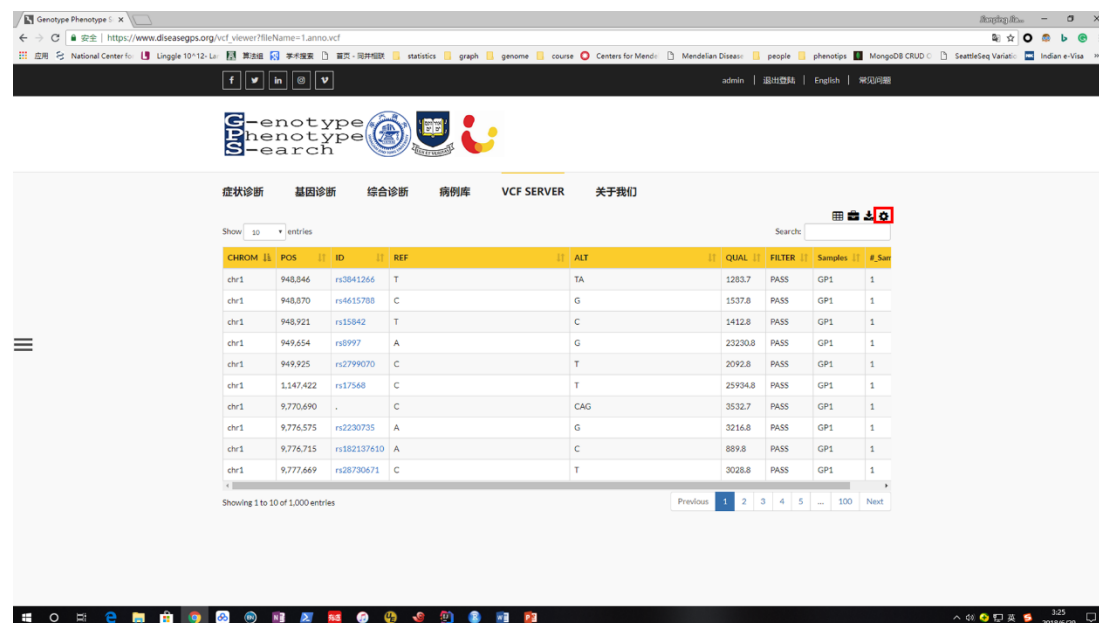
## 撤销筛选条件



## 将筛选结果导出为 CSV



## 建立列索引(加快筛选速度)



Genotype Phenotype Search

虚状诊断 基因诊断 综合诊断 病例库 VCF SERVER 关于我们

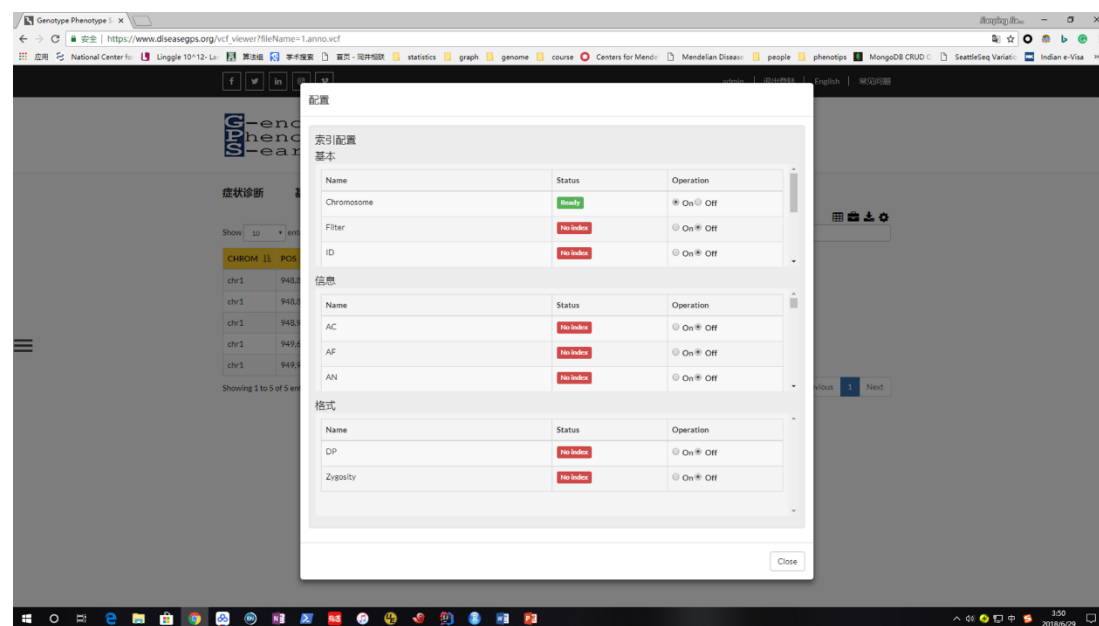
Show: 10 entries

Search:

CHROM	POS	ID	REF	ALT	QUAL	FILTER	Samples	#_Sam
chr1	948,046	rs3841266	T	TA	1283.7	PASS	GP1	1
chr1	948,070	rs4615788	C	G	1537.8	PASS	GP1	1
chr1	948,921	rs15842	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



Genotype Phenotype Search

虚状诊断 基因诊断 综合诊断 病例库 VCF SERVER 关于我们

Show: 10 entries

索引配置

基本

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

信息

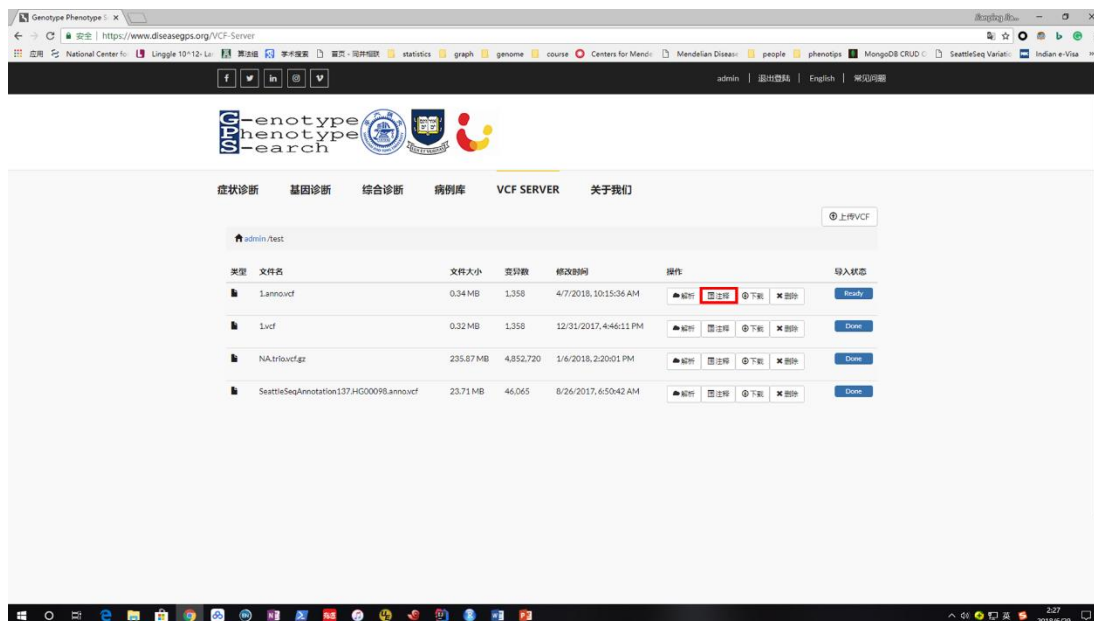
Name	Status	Operation
AC	No Index	On/Off
AF	No Index	On/Off
AN	No Index	On/Off

格式

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

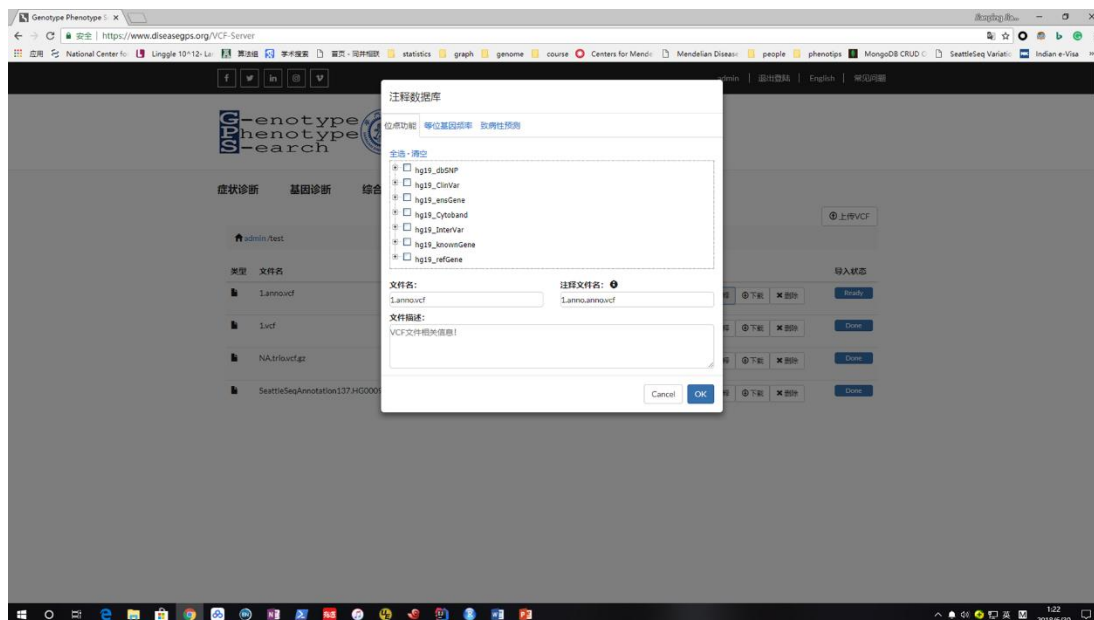
Close

## VCF 文件注释

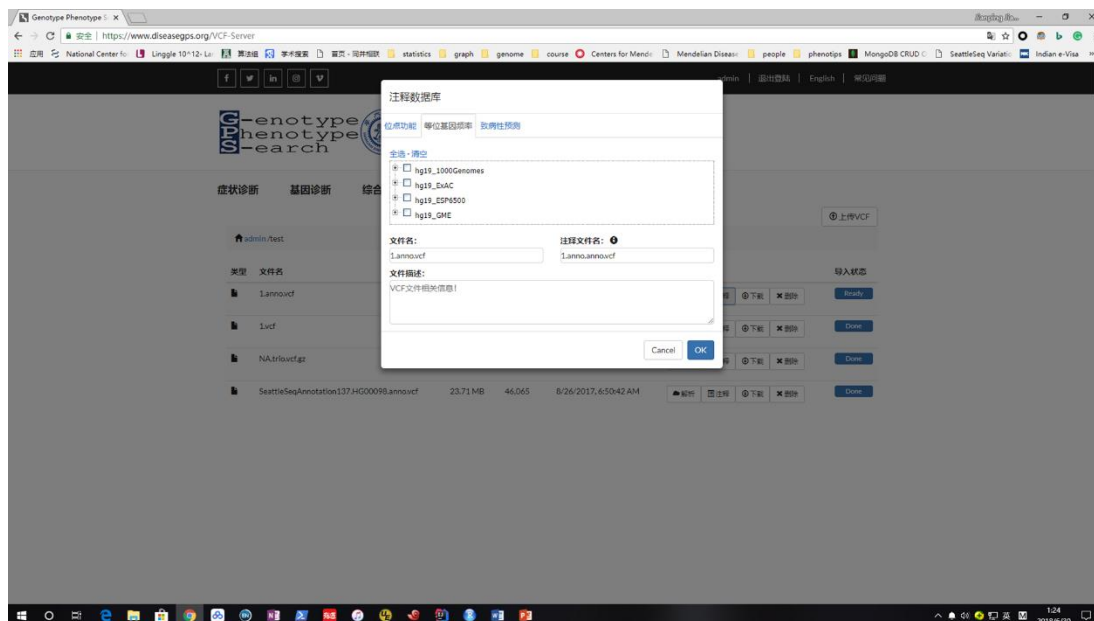


## 注释数据库选择

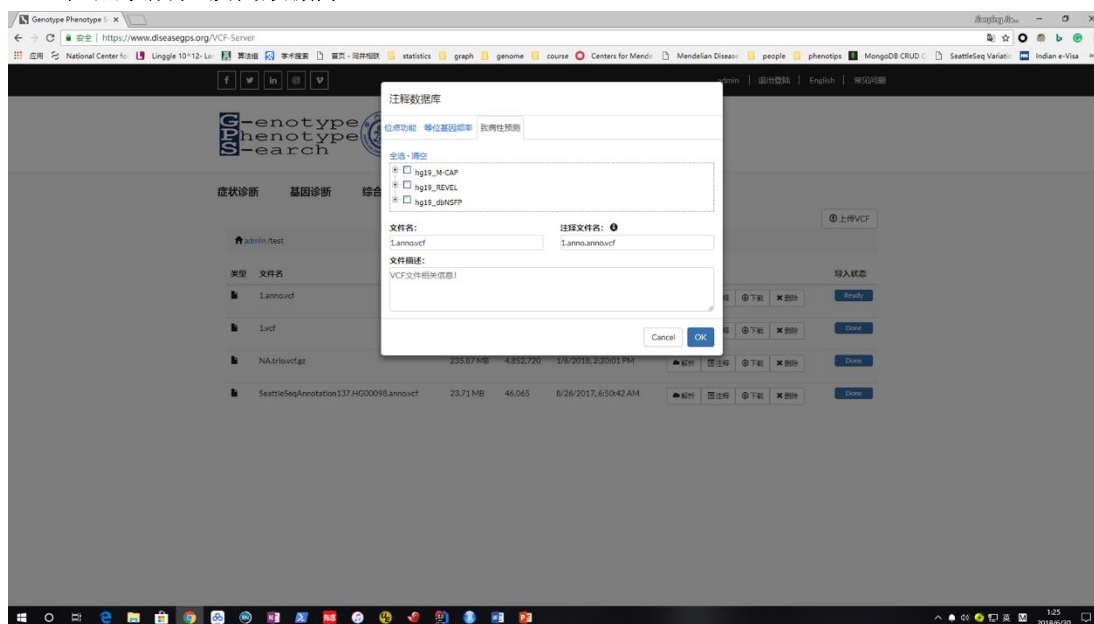
### ● 位点功能数据库



## ● 等位基因频率数据库

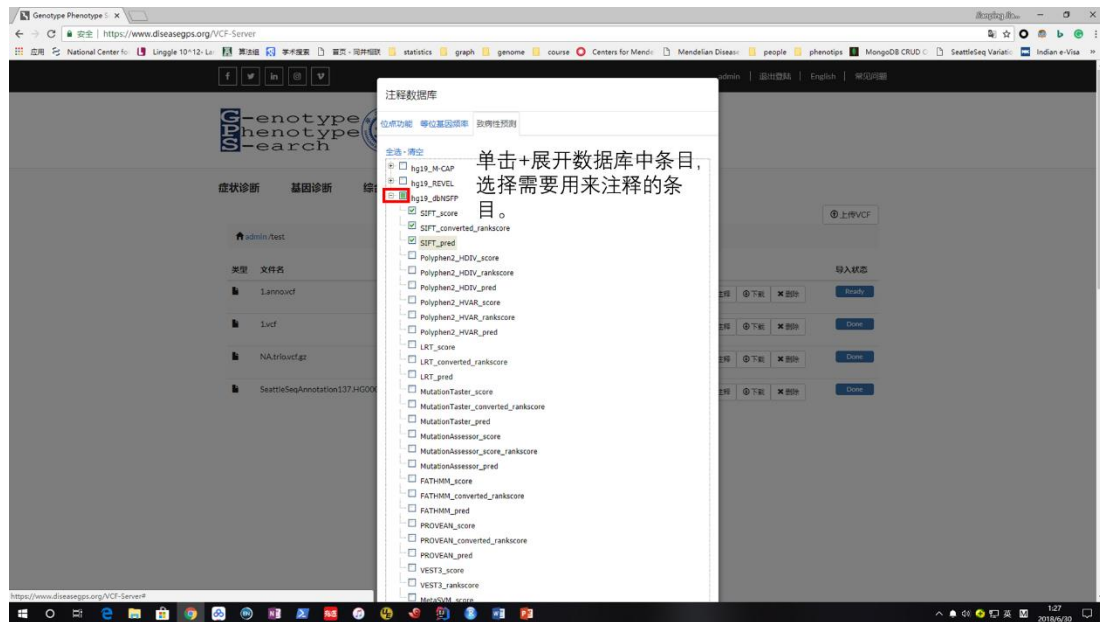


## ● 位点致病性预测数据库





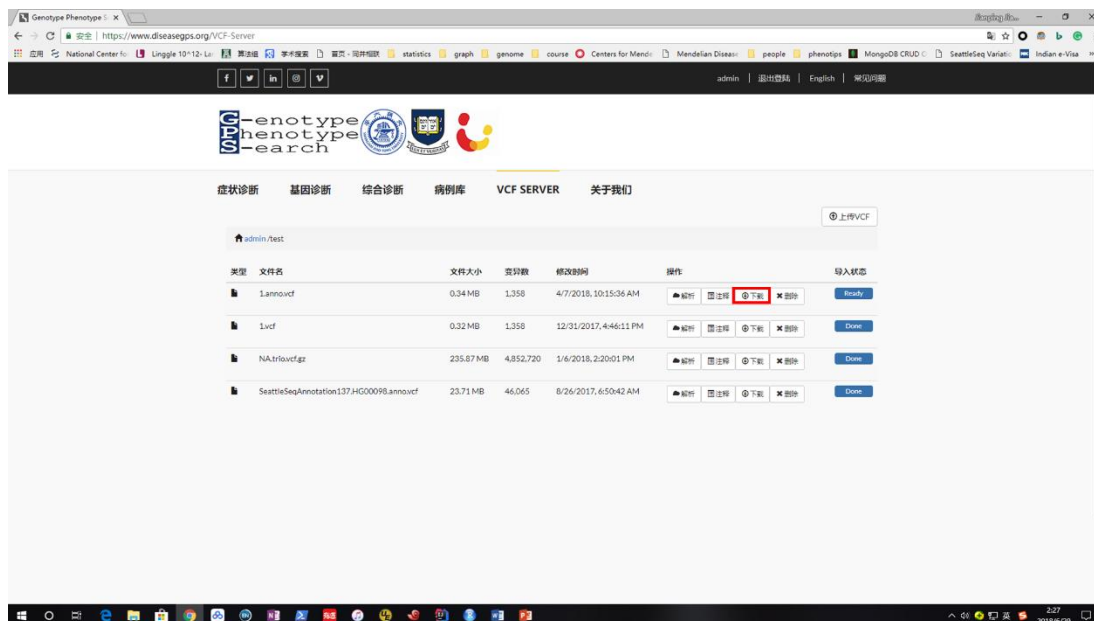
## 注释数据库中的某些条目



## 注释数据库信息

Table S1. Summary of databases used in VCF Server annotation			
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.

## VCF 文件下载



## VCF 文件删除

