GPSanno User Manual

1. Introduction

GPSanno is a software aims to search phenotypes with genetic variants. It was implemented with perl language. It takes VCF file as input file, and web pages as output.

2. Obtain and dependency

The source code and data can be found on https://github.com/kimi641/GPSanno
User need to download additional data list in the main page.
GPSanno depend on some perl packages: FindBin, LWP::Simple, utf8::all, JSON. All of these packages can be found in CPAN.

3. Runing

3.1. Single sample running:

This mode is designed for VCF with only one sample, program analysis every genetic variants without genotype information.

input file: one VCF file with one sample
output: a directory contains analysis reports

Example:
./GPSanno.pl –i example.vcf –o dirname

3.2. Multiple samples running:

This mode is designed for one VCF with different samples form one family. Program filter genetic variants with genotypes from family members first.

input file: one VCF file with samples
input parameters: sample tags in VCF file
output: a directory contains analysis reports

Example:

./GPSanno.pl –i example.vcf –o dirname –p patient1,patient2 –n normal1,normal2

3.3. Optional parameter:

MAF: Mirror Allele Frequency filter cutoff, the default value is 0.01. Any value >0.01 can be set by parameter –maf

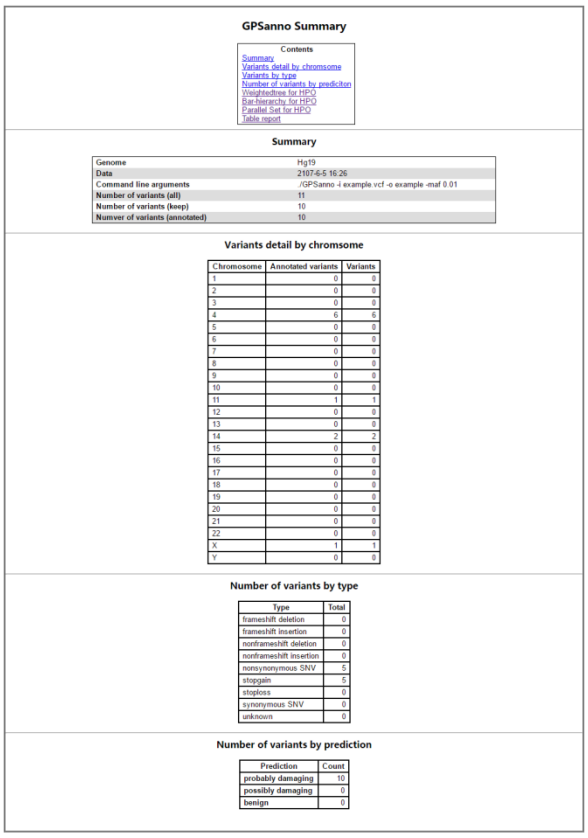
Example:
./GPSanno.pl –i example.vcf –o dirname –maf 0.05

4. Output

The output are web pages, users can visit the results in internet browsers. But by default, internet browsers forbid user to access files from local files, so in Windows OS, users can either append "--allow-file-access-from-files" in internet browsers command line and restart internet browsers. While in Mac OS, users need to restart chrome with "open /Applications/Google\ Chrome.app --args --allow-file-access-from-files" or "Run in Unsafe Mode" in safari or select "Disable Local File Restrictions" as develop version. Attention, progress should be stop totally before restart a new window of browser.

4.1. Summary page

Summary page is a page contains basic statistic information of one run. There are five sections in this page: Contents, Summary, Variants position, Variants type and Variants prediction. Links of other sections and web pages are listed in Contents. Each section shows different statistic result of variants. In addition, command line and run time are in Summary section.



GPSanno Summary

Contents

Summary

Genome	Hg19
Data	2107-6-5 16:26
Command line arguments	/GPSanno -i example.vcf -o example -maf 0.01
Number of variants (all)	11
Number of variants (keep)	10
Numver of variants (annotated)	10

Variants detail by chromsome

Chromosome	Annotated variants	Variants
1	0	0
2	0	0
3	0	0
4	6	6
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	1	1
12	0	0
13	0	0
14	2	2
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0
21	0	0
22	0	0
X	1	1
Y	0	0

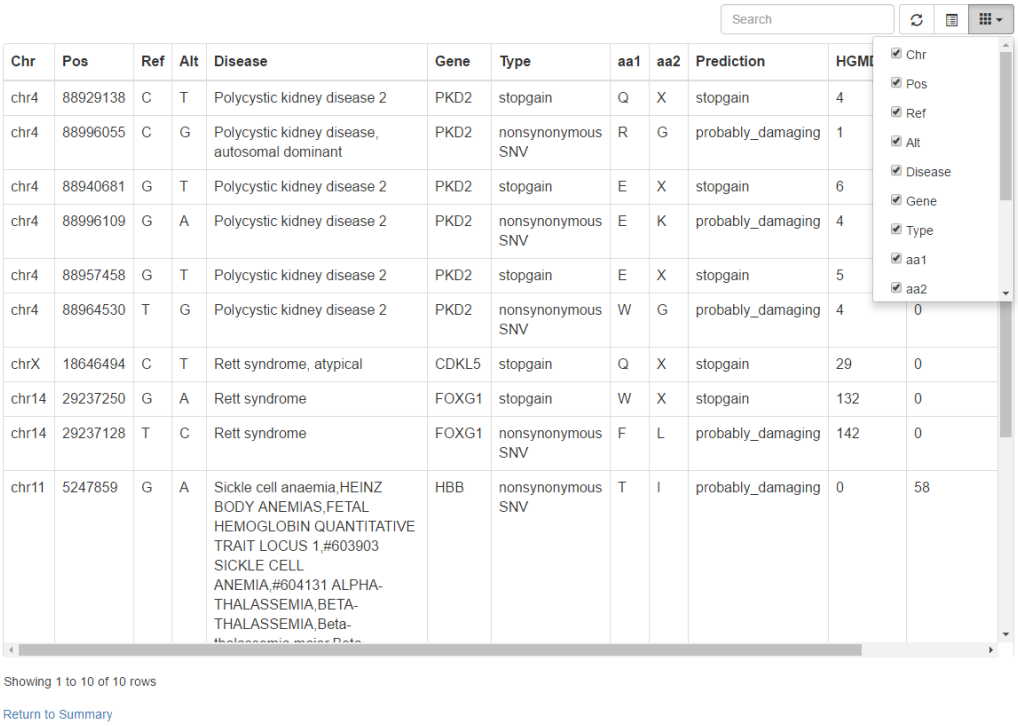
Number of variants by type

Type	Total
frameshift deletion	0
frameshift insertion	0
nonframeshift deletion	0
nonframeshift insertion	0
nonsynonymous SNV	5
stopgain	5
stoploss	0
synonymous SNV	0
unknown	0

Number of variants by prediction

Prediction	Count
probably damaging	10
possibly damaging	0
benign	0

4.2. Table page



Search

Chr Pos Ref Alt Disease Gene Type aa1 aa2 Prediction HGMD

Chr	Pos	Ref	Alt	Disease	Gene	Type	aa1	aa2	Prediction		
chr4	88929138	C	T	Polycystic kidney disease 2	PKD2	stopgain	Q	X	stopgain	4	
chr4	88996055	C	G	Polycystic kidney disease, autosomal dominant	PKD2	nonsynonymous SNV	R	G	probably_damaging	1	
chr4	88940681	G	T	Polycystic kidney disease 2	PKD2	stopgain	E	X	stopgain	6	
chr4	88996109	G	A	Polycystic kidney disease 2	PKD2	nonsynonymous SNV	E	K	probably_damaging	4	
chr4	88957458	G	T	Polycystic kidney disease 2	PKD2	stopgain	E	X	stopgain	5	
chr4	88964530	T	G	Polycystic kidney disease 2	PKD2	nonsynonymous SNV	W	G	probably_damaging	4	0
chrX	18646494	C	T	Rett syndrome, atypical	CDKL5	stopgain	Q	X	stopgain	29	0
chr14	29237250	G	A	Rett syndrome	FOXG1	stopgain	W	X	stopgain	132	0
chr14	29237128	T	C	Rett syndrome	FOXG1	nonsynonymous SNV	F	L	probably_damaging	142	0
chr11	5247859	G	A	Sickle cell anaemia,HEINZ BODY ANEMIAS,FETAL HEMOGLOBIN QUANTITATIVE TRAIT LOCUS 1,#603903 SICKLE CELL ANEMIA,#604131 ALPHA-THALASSEMIA,BETA-THALASSEMIA,Beta-	HBB	nonsynonymous SNV	T	I	probably_damaging	0	58

Showing 1 to 10 of 10 rows

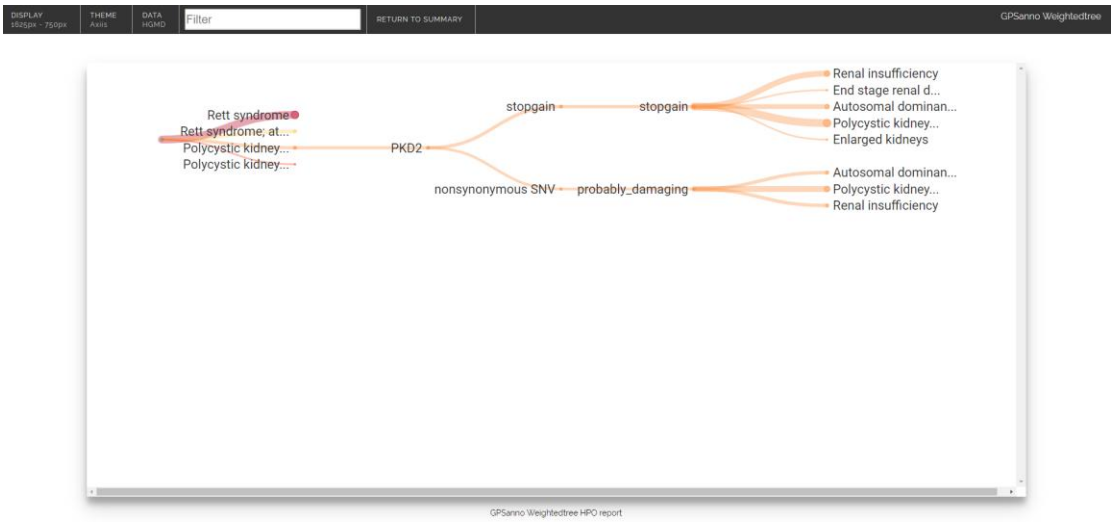
Return to Summary

Table page contains all variants annotated by GPSanno. It has 14 columns: chromosome, position, diseases, gene, variants type, reference amino acid, alternative amino acid, variants prediction,

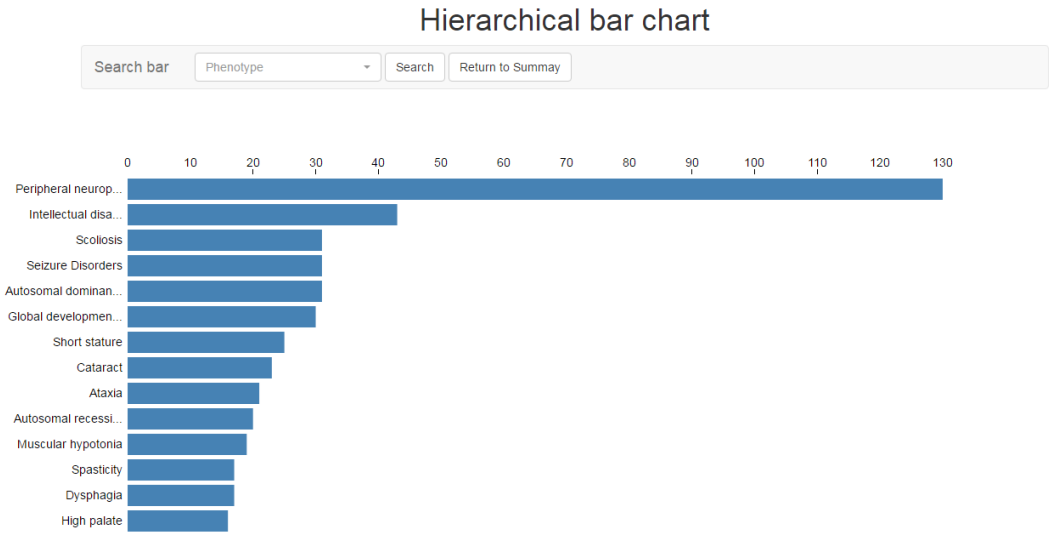
records in HGMD database, records in MedGen database, records in HPO database, PubMed ID.

4.3. Weighted tree page

Weighted tree provides a forward search function to help users find relationship among disease, gene and phenotypes, users can specific keywords first, and then click the data source button in the navigation bar to filter phenotypes.



4.4. Hierarchical bar page



Hierarchical bar page contains phenotypes with count in one run. Each bar contains three level data: Phenotypes, diseases and genes. Users can expand to next level by clicking bars and return to previous level by clicking blank space.

In addition, users can search for single or multiple phenotypes which appear in the result. Leave

search as blank will show the initial page.

4.5. Parallel dataset page

Parallel dataset shows four levels data and relationship among them. Users can specific gene, disease, phenotype in search bar to find target records.