ViGWAS User Manual

Installation

To install all required software and packages follow the instruction in **ViGWAS_Setup_Instruction.sh**

You can also use our **VirtulBox Image** or **AWS AMI** that includes an Ubuntu 18.04 with all required software and packages installed. For more information please see https://bioinformatics.csiro.au/ and look for **ViGWAS** in software list.

Input

Sample annotations

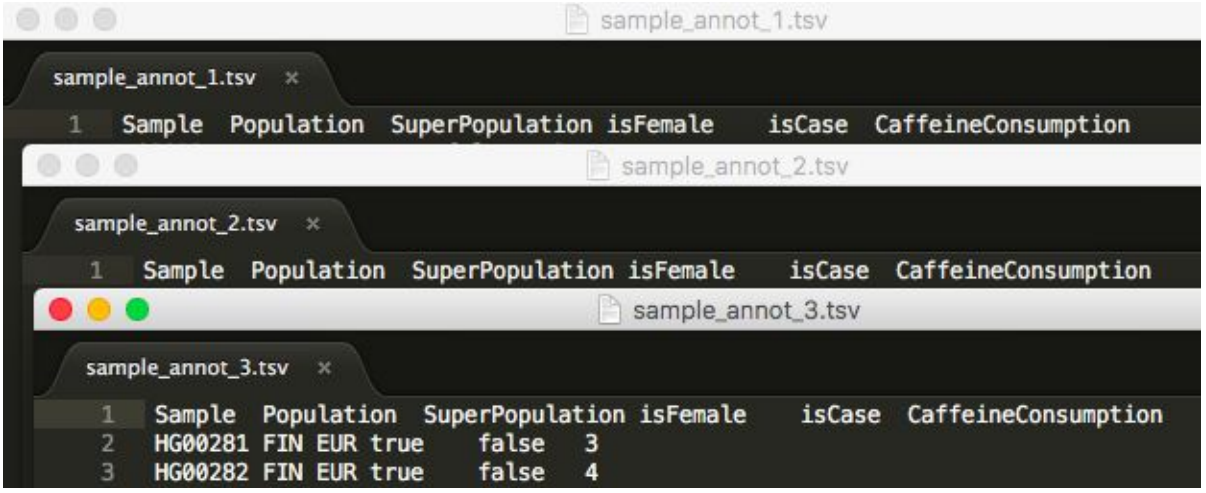
File types

TSV/CSV files. The delimiter needs to be specified in the user block.

```
sample_annot_deli = '\t' #'\t' for tsv and ',' for csv
sample_annot_file_list = ['sample_input/sample_annot_1.tsv',
                          'sample_input/sample_annot_2.tsv',
                          'sample_input/sample_annot_3.tsv']
```

Multiple files

Can take a list of continuing sample annotation files. Same header must be included in each file.



File formats

'Sample' 'isCase'(boolean) 'isFemale'(boolean) are compulsory fields required in sample annotations. Field names and value types must be exactly same for a successful read-in.

Sample	isFemale	isCase
HG00281	true	false
HG00282	true	false
HG00284	false	true
HG00285	true	true
HG00288	true	true
HG00290	false	false
HG00302	true	false
HG00303	false	true
HG00304	true	true
HG00306	true	false

Variant Information

File types
PLINK or vcf(.bgz) files. File types need to be specified.

PLink files:

variant.bed
variant.bim
variant.fam

PLink file input settings in user block:

```
mt_file_type = 'plink' # support 'vcf' or 'plink'
mt_file_list = ['sample_input/variant']
```

vcf files:

variant1.vcf.bgz
variant2.vcf.bgz

vcf file input settings in user block:

```
mt_file_type = 'vcf' # support 'vcf' or 'plink'
mt_file_list = ['sample_input/variant1.vcf.bgz',
                'sample_input/variant2.vcf.bgz']
```

Multiple Files

Can take a list of vertically/horizontally continuing files. Continuing types need to be specified as 'variant' or 'sample' for vertically or horizontally continuing files respectively.

```
mt_merge_type = 'variant'
```
⇒ same set of samples, different variants in different files

```
mt_merge_type = 'sample'
```
⇒ same set of variants, different samples in different files

Sample IDs
All samples must exist in sample annotations. Sample IDs must be exactly same as they are in sample annotations.

Other parameters

Name of analysis

```
analysis_name = 'my_analysis'
```
⇒ the name of the result directory & the output files

Graphing

```
downsample_percent = 1
graph_type = 'stack'
fields_to_plot = ['isFemale', 'Population',
                  'isCase', 'SuperPopulation']
```

⇒ downsample rate for plotting
⇒ ‘stack’ or ‘group’ graph type for complex graphs

⇒ fields to be plotted by for sample annotations, PCA

PCA and Logistic Regression

```
n_factor = 4
```

⇒ number of factors for PCA and Logistic Regression

Variant Spark

A full documentation about VariantSpark is available here:
https://docs.databricks.com/applications/genomics/variant-spark.html

```
PATH_TO_VS = '~/tools/VariantSpark/bin/variant-spark'
```

⇒ path to variant-spark

```
mtry_fraction=0.1
```

mtry_fraction is the fraction of input variants used as **mtry**. For example, if there are 20,000 variants in the input file and the **mtry_fraction** is 0.1 then the **mtry** for the Random-Forest analysis is 2,000 (=20,000*0.1). In Random-Forest **mtry** is the number of randomly selected variables which are evaluated to split each node of a tree.

```
num_of_tree=1000
```

⇒ Number of trees in random forest

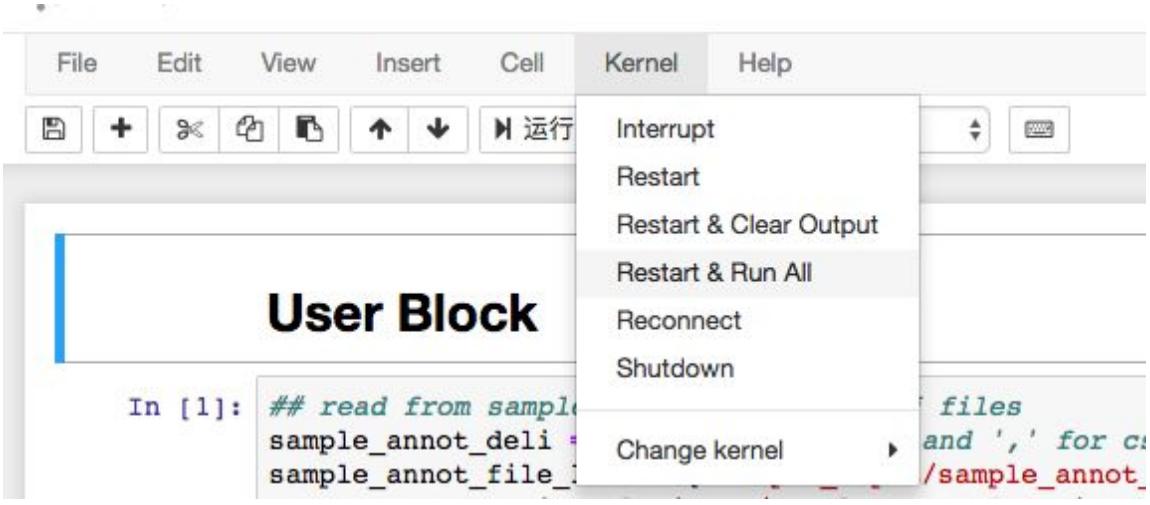
Computer Configs

```
numCPU = 32
memory = '100g'
```

⇒ number of CPUs available for the analysis
⇒ RAM available for the analysis

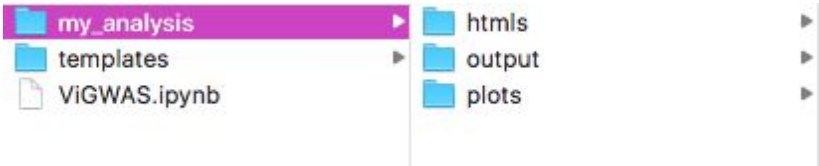
Running ViGWAS

After setting all input inside the user-block, press run all.

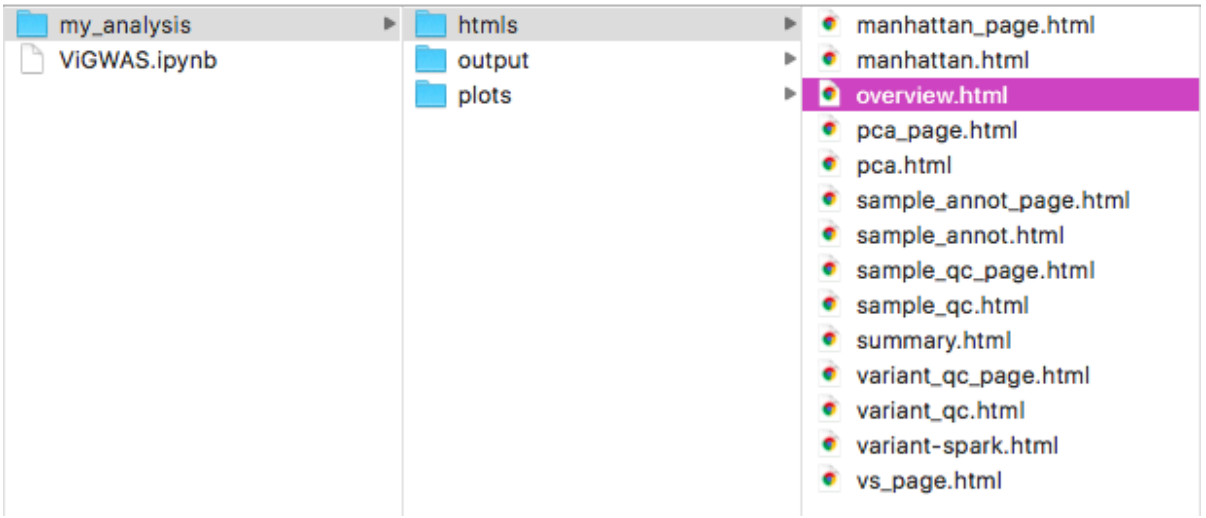


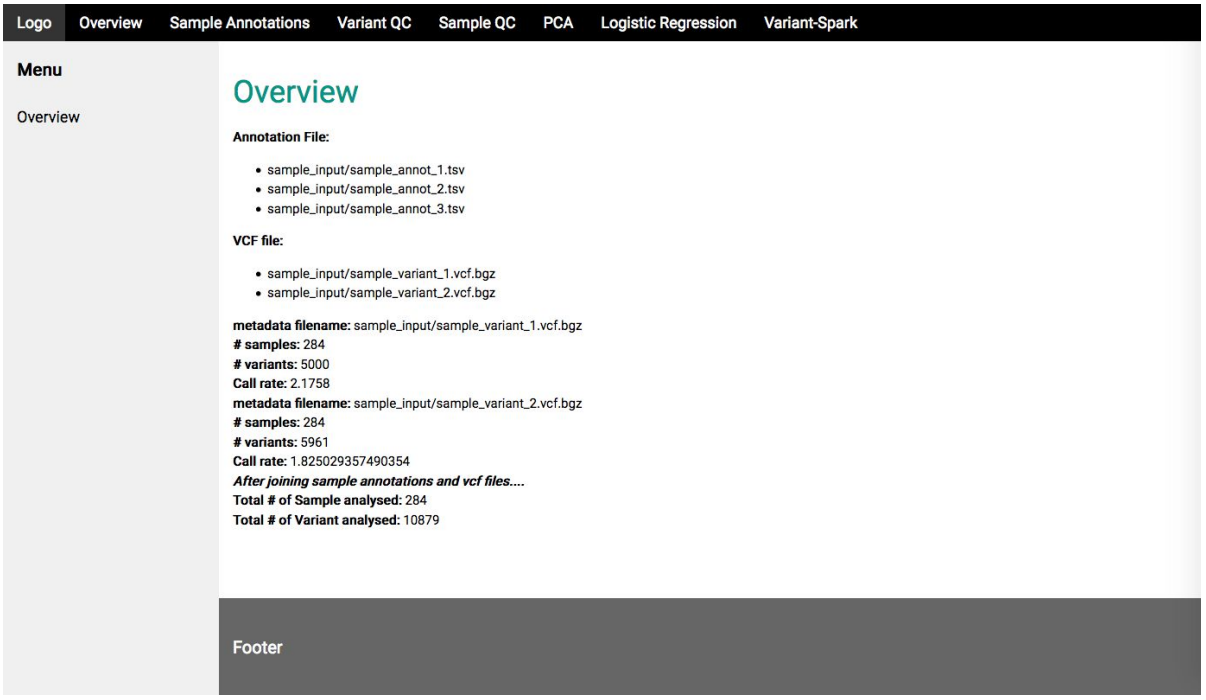
Output

A folder with user's given name created in the same directory.

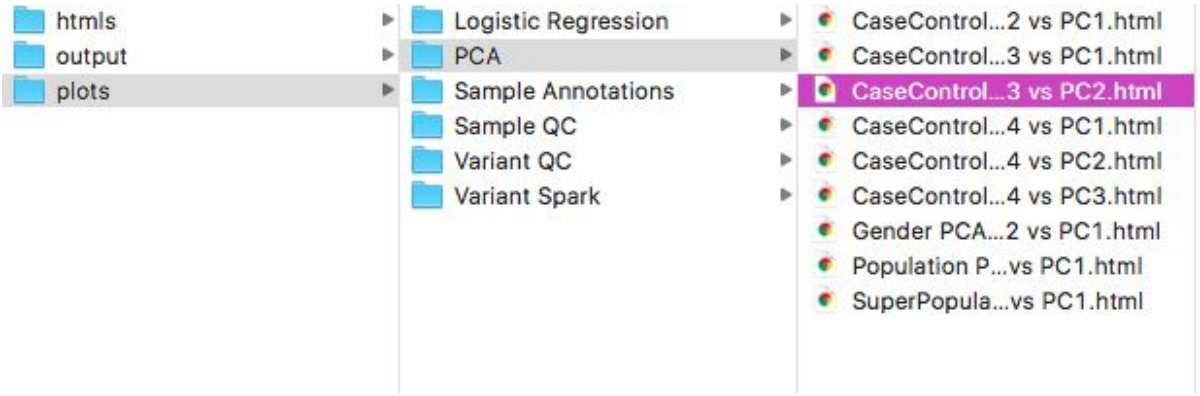


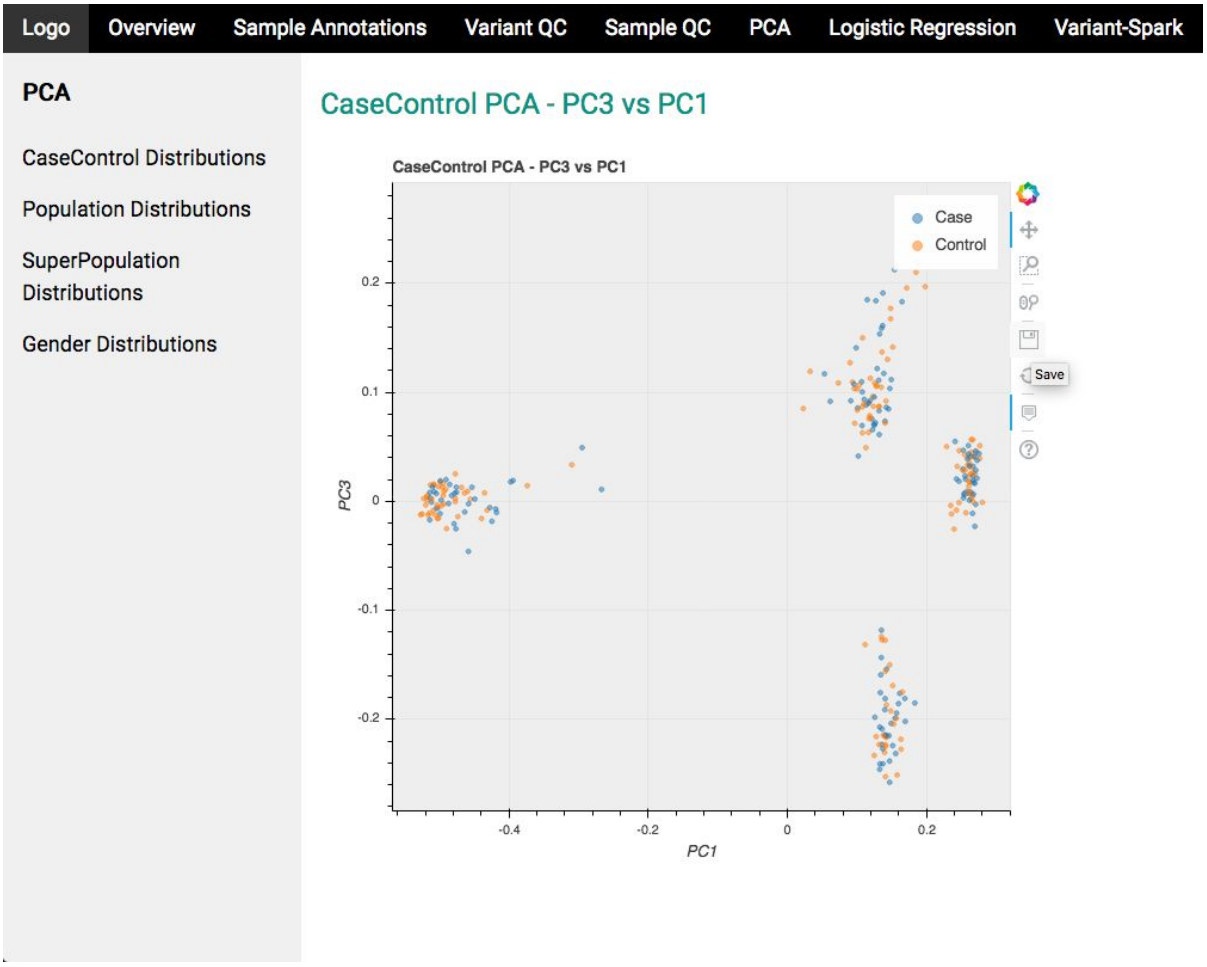
To view the result report, go to htmls folder and open overview.html with Firefox web browser. Navigate to the corresponding page for the results interested in.





To export a certain plot, go to plots folder and find the desired individual plot in html format under the analysis directory or save plot as png from the result report page.





To get the annotated variant information, go to output directory and find the annotated variant information in vcf and plink format. Hail format MatrixTable (.mt) are also available.