

User Manual for



Intelligent Prediction and Association Tool

(Version 1.3)

Last updated on Aug 14, 2017

Table of Contents

| | |
|--|----------------|
| 1. Getting start | page 3 |
| 1-1 Operation environment | page 3 |
| 1-2 Windows users | page 3 |
| 1-3 Mac OS users | page 6 |
| 2. Interface | page 7 |
| 2-1 Import files | page 7 |
| 2-2 Create a project | page 7 |
| 2-3 File formats | page 7 |
| 2.3.1 Phenotype | page 7 |
| 2.3.2 Hammap | page 8 |
| 2.3.3 Numeric | page 9 |
| 2.3.4 VCF | page 9 |
| 2.3.5 PLINK | page 9 |
| 2.3.6 Binary PLINK | page 10 |
| 2-4 Covariates and kinship | page 11 |
| 2.4.1 Add additional information to iPat | page 11 |
| 2.4.2 Covariates | page 11 |
| 2.4.3 Kinship | page 11 |
| 2-5 Define input arguments | page 12 |
| 2-6 Run an analysis | page 13 |
| 2-7 Check the results | page 14 |
| 2-8 Files remove | page 15 |
| 3. GWAS, GS | page 16 |
| 3-1 GAPIT | page 16 |
| 3-2 FarmCPU | page 16 |
| 3.3 PLINK | page 16 |
| 3.4 rrBLUP | page 17 |
| 3.5 BGLR | page 17 |
| 4. Support | page 17 |
| 5. Citation | page 17 |

1. Getting start

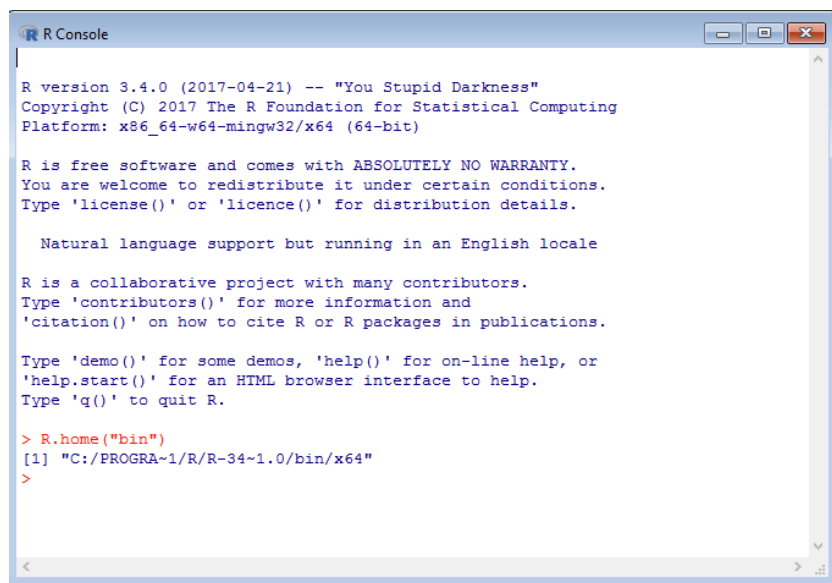
1.1 Operation environment

- The operation environment need to meet the following requirement:
 - Operation System: Windows or Mac OS X .
 - [Java Runtime Environment \(JRE\)](#): Version 8 or later.
 - [R](#): Version 3.4.0 or later.

1.2 Windows users

1.2.1 Set up R environment

- If you can call R from the commnad-line window (cmd.exe) by typing "R" or "r", then you can skip to section 1.2.2. Otherwise, please follow the instruction below to get your system compatible with iPat.
- Open R software, and type `R.home("bin")` in the console. It will return a path to the executable R. Copy this path to the clipboard.



```

R Console

R version 3.4.0 (2017-04-21) -- "You Stupid Darkness"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

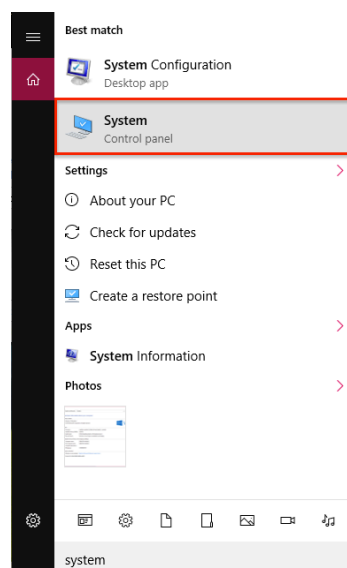
Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

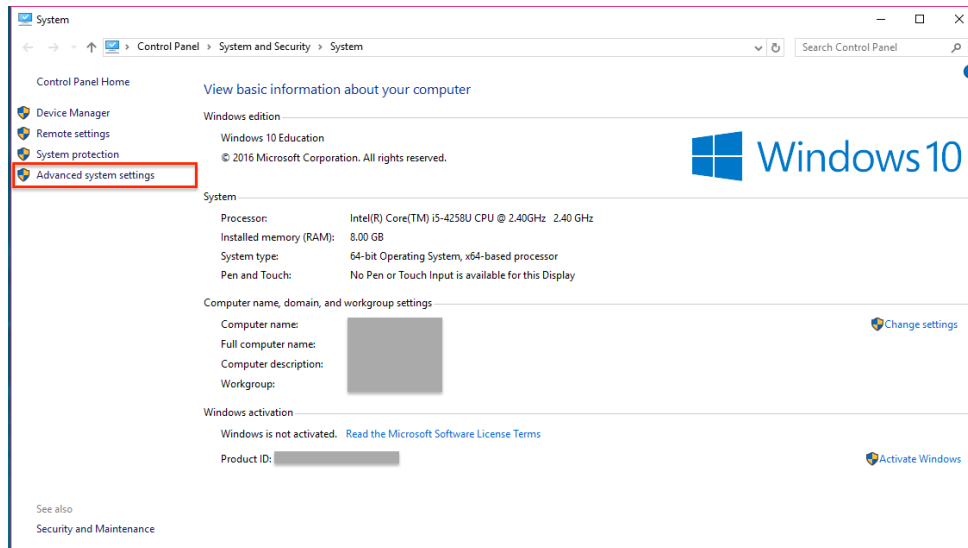
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> R.home("bin")
[1] "C:/PROGRA~1/R/R-34~1.0/bin/x64"
>
  
```

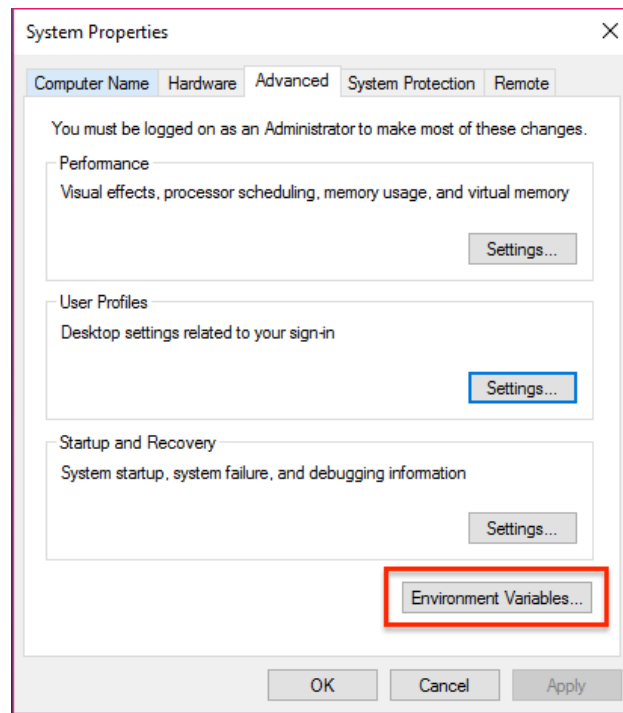
- Search keyword "system" from Windows, and open "System".



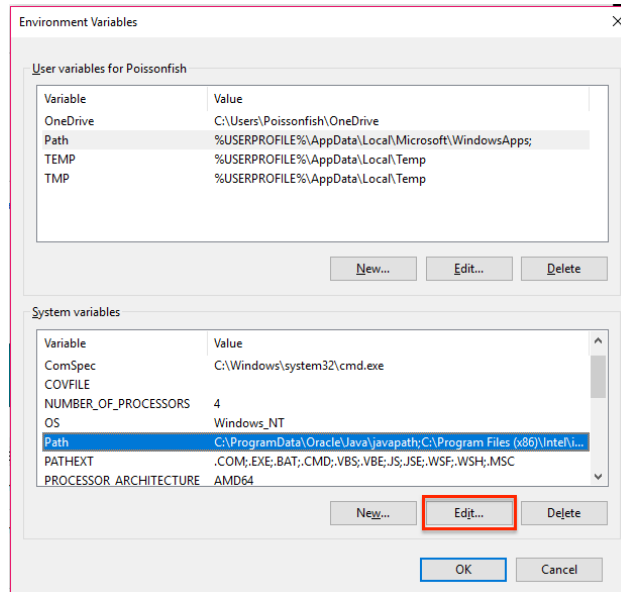
- Then select "Advanced system settings" at the left side of the panel.



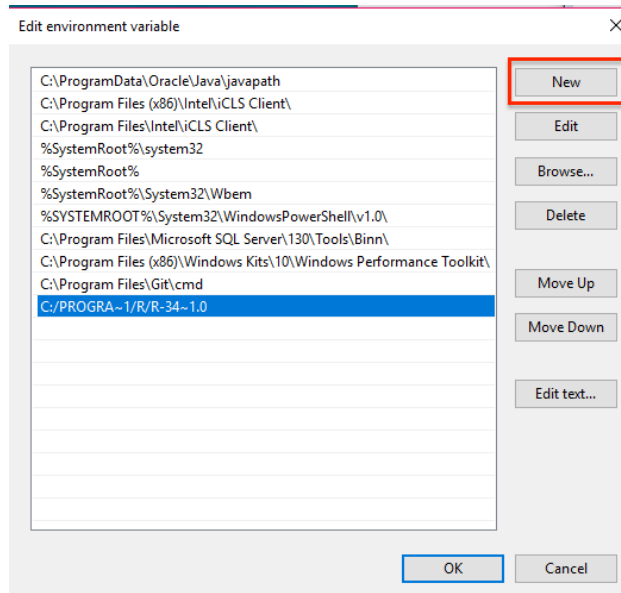
- And click "Environment Variable..." at the bottom-right area.



- The pop-up windows will display two set of system variables. Highlight the system variable "Path" at the bottom list, and click "Edit".

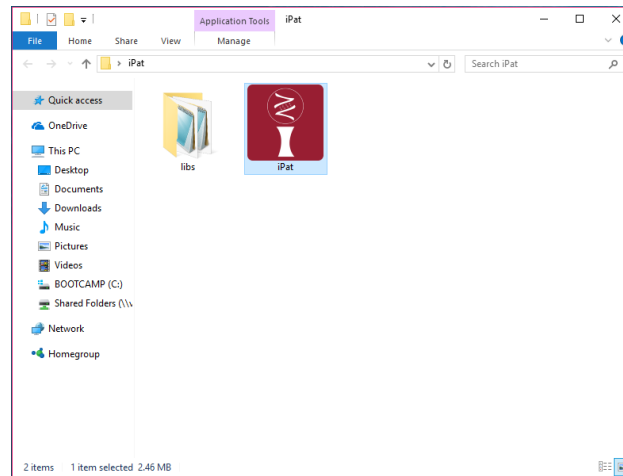


- Almost there. Click "new" and paste the path you got from the clipboard, then click "OK" to save the configuration.



1.2.2 Extract iPat.exe from iPat.zip

- Download [iPat.zip](#) and decompress it. You will then get a folder named "iPat", which contains a executable file "iPat.exe" and a folder "libs".
- It's noted that users are always required to place "iPat.exe" and the folder "libs" in the same folder (directory) so that iPat can function normally.



- Double click 'iPat.exe' to launch iPat.

1.3 Mac OS users

- Download [iPat_Installer.dmg](#) and mount it on Mac.
- Follow the instruction to install iPat.

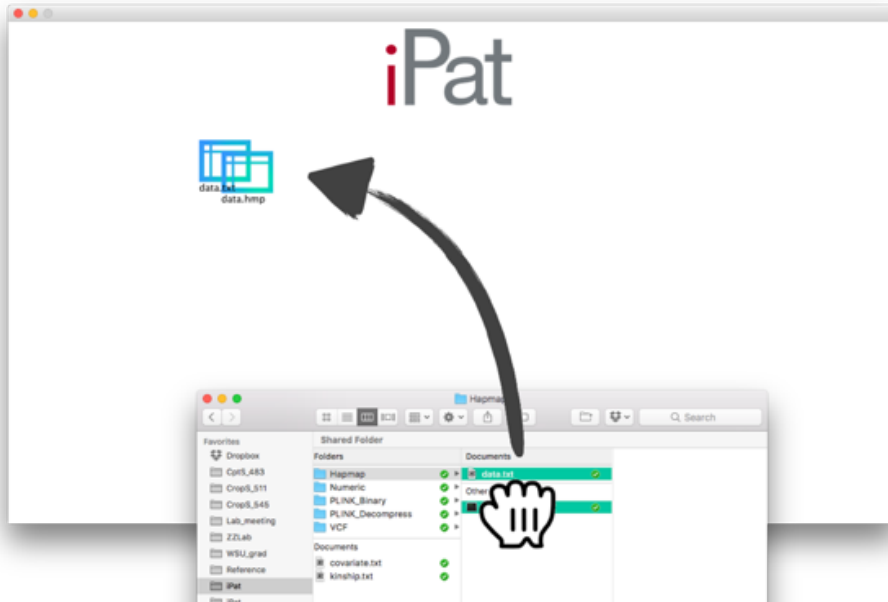


- Double click 'iPat.app' to launch iPat.

2. Interface

2.1 Import files

- At beginning, iPat will show nothing but an icon "iPat" at the top of screen.
- Users can import files simply by dragging and dropping.



2.2 Create a project

- After importing the files, double clicking on anywhere in iPat to create a new project (a gear icon).
- Build a project by dragging a files over the project icon. A dashline will be shown between the file and project, which mean this file has been already included in this project. The below two are examples for a valid project.

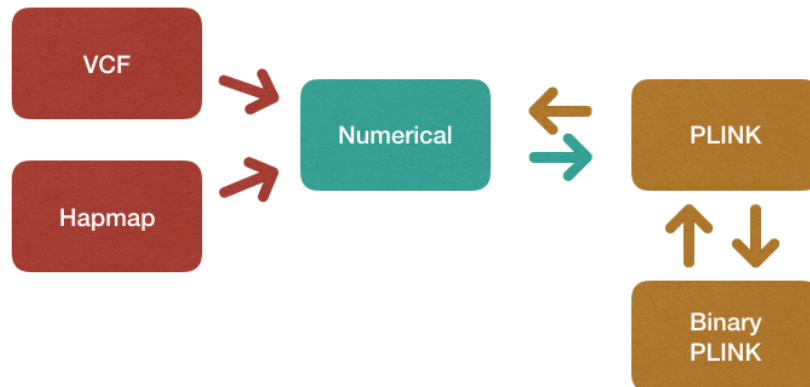


- A valid project must include a certain number of required files, **no less, no more**. Otherwise iPat won't work and will return an error message. Valid datasets for each format can be found from the table below:

| Format | File 1 (required) | File 2 (required) | File 3 (required) | File 4 (required) |
|-------------------|----------------------|----------------------|--|----------------------------------|
| Hapmap | Genotype (.hmp) | Phenotype (.txt) | None | None |
| Numeric | Genotype (.dat) | Phenotype (.txt) | Map information (.map) (Only required for GWAS) | None |
| VCF | Genotype (.vcf) | Phenotype (.txt) | None | None |
| PLINK | Genotype (.ped) | Phenotype (.txt) | Map information (.map) | None |
| PLINK (binary) | Genotype (.bed) | Phenotype (.txt) | Map information (.bim) | Individual information (.fam) |

2.3 File formats

- iPat mainly work with files in numerical format, but it can also work fine with Hapmap, VCF and PLINK format. iPat will recognize the format of input files and do a format conversion automatically if needed.



2.3.1 Phenotype

- Phenotype data for every formats except PLINK must contain **sample names** in the first column and **traits names** as the header:

| taxa | trait 1 | trait 2 |
|---------|---------|---------|
| sample1 | | |
| sample2 | | |
| sample3 | | |

- Phenotype data for PLINK must contain **sample and family names** in the first 2 columns and **traits names** as the header:

| FID | SID | trait 1 | trait 2 |
|----------|---------|---------|---------|
| family 1 | sample1 | | |
| family 2 | sample2 | | |
| family 3 | sample3 | | |

2.3.2 Hapmap

- Genotype data, the header is required to be provided:

| rs | alleles | chrom | pos | strand | assembly | center | protLSID | assayLSID | panel | QCcode | sample 1 | sample 2 | sample 3 |
|----------|---------|-------|---------|--------|----------|--------|----------|-----------|----------|--------|----------|----------|----------|
| marker 1 | A/C | 1 | 157104 | + | AGPv1 | Panzea | NA | NA | maize282 | NA | CC | CC | AA |
| marker 2 | C/G | 1 | 1947984 | + | AGPv1 | Panzea | NA | NA | maize282 | NA | GG | GG | CC |

2.3.3 Numeric

- Genotype data, samples are recorded in rows. The header and sample names can be omitted:

| taxa | marker 1 | marker 2 | marker 3 |
|---------|----------|----------|----------|
| sample1 | 0 | 0 | 1 |
| sample2 | 0 | 0 | 0 |
| sample3 | 1 | 0 | 0 |

- Map information, the header is required to be provided:

| SNP | Chromosome | Position |
|----------|------------|----------|
| marker 1 | 1 | 157104 |
| marker 2 | 1 | 1947984 |
| marker 3 | 1 | 2914066 |

2.3.4 VCF

- Genotype data, the header is required to be provided:

| CHROM | POS | ID | REF | ALT | QUAL | FILTER | INFO | FORMAT | sample 1 | sample 2 | sample 3 |
|-------|---------|----------|-----|-----|------|--------|------|--------|----------|----------|----------|
| 1 | 157104 | marker 1 | A | C | . | PASS | . | GT | 0/0 | 1/1 | 0/0 |
| 1 | 1947984 | marker 2 | C | G | . | PASS | . | GT | 0/0 | 1/1 | 1/1 |

2.3.5 PLINK (the header should be removed)

- Genotype data (.ped). Missing value can be filled as "0":

| Family ID | Sample ID | Paternal ID | Maternal ID | Sex | Affection | marker 1 | marker 2 | marker 3 |
|-----------|-----------|-------------|-------------|-----|-----------|----------|----------|----------|
| FAM1 | NA06985 | 0 | 0 | 1 | 1 | A A | T T | A A |
| FAM1 | NA06991 | 0 | 0 | 1 | 1 | A A | T T | A A |
| 0 | NA06993 | 0 | 0 | 1 | 1 | C T | C C | T T |

- Map information (.map):

| Chromosome | Marker ID | Genetic distance | Physical Position |
|------------|-----------|------------------|-------------------|
| 1 | marker 1 | 0 | 157104 |
| 1 | marker 2 | 0 | 1947984 |
| 1 | marker 3 | 0 | 2914066 |

2.3.6 Binary PLINK (the header should be removed)

- Genotype data (.bed):

Please follow the instruction from [here](#)

- FAM file:

| Family ID | Sample ID | Paternal ID | Maternal ID | Sex | Affection |
|-----------|-----------|-------------|-------------|-----|-----------|
| FAM1 | NA06985 | 0 | 0 | 1 | 1 |
| FAM1 | NA06991 | 0 | 0 | 1 | 1 |

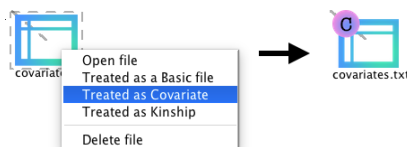
- BIM file:

| Chromosome | Marker ID | Genetic distance | Physical Position | Allele 1 | Allele 2 |
|------------|-----------|------------------|-------------------|----------|----------|
| 1 | marker 1 | 0 | 157104 | A | C |
| 1 | marker 2 | 0 | 1947984 | A | T |

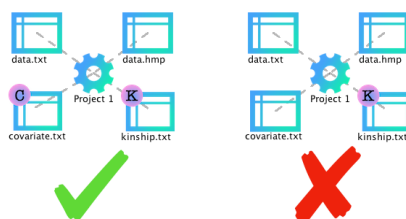
2.4 Covariates and kinship

2.4.1 Add additional information to iPat

- Covariates provided by users will be treated as **fixed effect** in the selected model except in BGLR.
- It's **optional** that users can add **user-define** covariates or kinship into the project. Right clicking on the file to tell iPat what type of file it is. (i.e. covariates, kinship or a basic required file)



- Be aware that apart from the basic required file (i.e. phenotype and genotype), optional files must be properly labeled in a project.



- The file labeled "C" stands for a covariate file, while files labeled "K" is identified as a kinship file by iPat. For the example of a valid project below (Left one), the file "covariate.txt" and "kinship.txt" are treated as covariates and a kinship in this project, respectively. Each project can contain **one** covariate files and **one** single kinship.

2.4.2 Covariates

- Demo format for a covariate file. The header is required to be provided:

| PC1 | PC2 | PC3 |
|------------|-------------|------------|
| -1.8942149 | -4.91532916 | 0.8674568 |
| 1.6858820 | -5.08378277 | -0.4069675 |
| 0.2579269 | -6.29547725 | 2.6867939 |

2.4.3 Kinship

- Demo format for a kinship file. Taxa name is required while the header can be omitted:

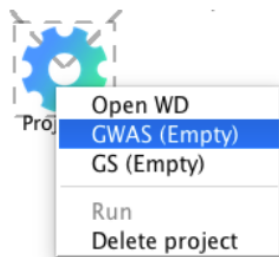
| taxa | sample 1 | sample 2 | sample 3 |
|----------|------------|------------|------------|
| sample 1 | 2.00000000 | 0.22883683 | 0.22932180 |
| sample 2 | 0.22883683 | 2.00000000 | 0.24496455 |
| sample 3 | 0.22932180 | 0.24496455 | 2.00000000 |

- If there is no user-define kinship, a kinship will be generated by the selected package:

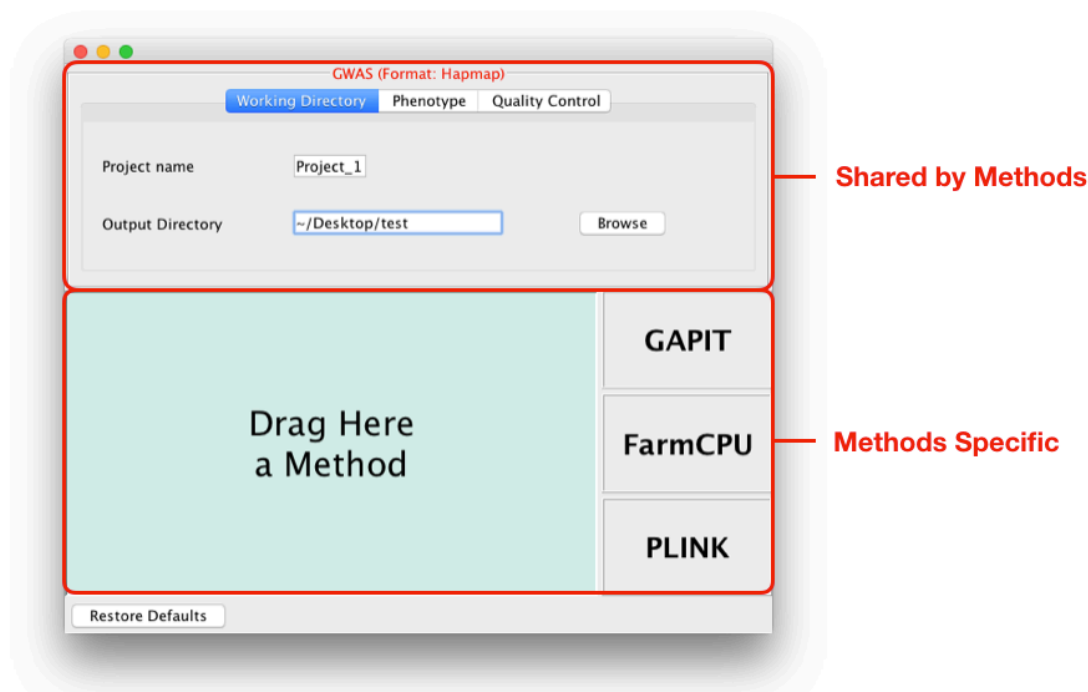
| Package | Kinship algorithm |
|---------|---|
| GAPIT | VanRaden (VanRaden, 2008), Loiselle (Loiselle <i>et al.</i> , 1995) or EMMA (Kang <i>et al.</i> , 2008) |
| FarmCPU | FARM-CPU (Liu <i>et al.</i> , 2016) |
| PLINK | User-provided |
| rrBLUP | VanRaden (VanRaden, 2008) |
| BGLR | User-provided |

2.5 Define Your Analysis

- After linking every files needed in the project, right click on the project and choose either GWAS or GS to open a configuration panel.



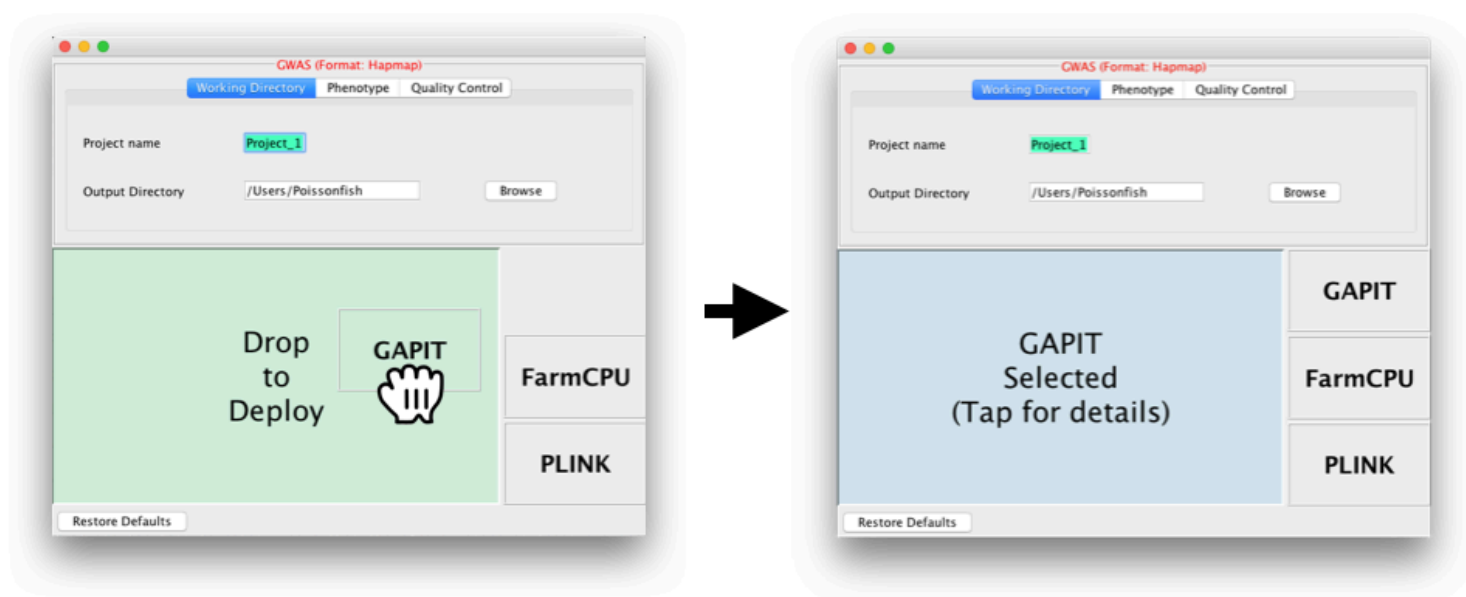
- The panel consist of two sections. The upper one presents a set of input arguments shared by all methods, while users can define method-specific arguments from the lower section.



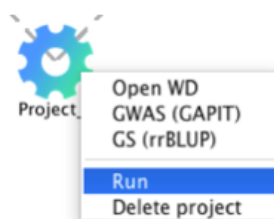
- Available parameters in the upper section :

| Tab | Parameters | Definition | Default |
|-------------------|------------------|--|---|
| Working Directory | Project name | Prefix for output files | Project_x (x is a number starts from 1) |
| Working Directory | Output Directory | A path where output files will be generated | Home directory |
| Phenotype | Trait names | Subsetting traits data | All traits are selected |
| Quality Control | By missing rate | Filtering out markers where certain rate of value is missing | No threshold |
| Quality Control | By MAF | Filtering out markers based on minor allele frequency | 0.05 |

- To select a method, simply drag a "method block" to the left-side area. And tap on this area for further defining (The information of method-specific arguments can be found in the section 3) .

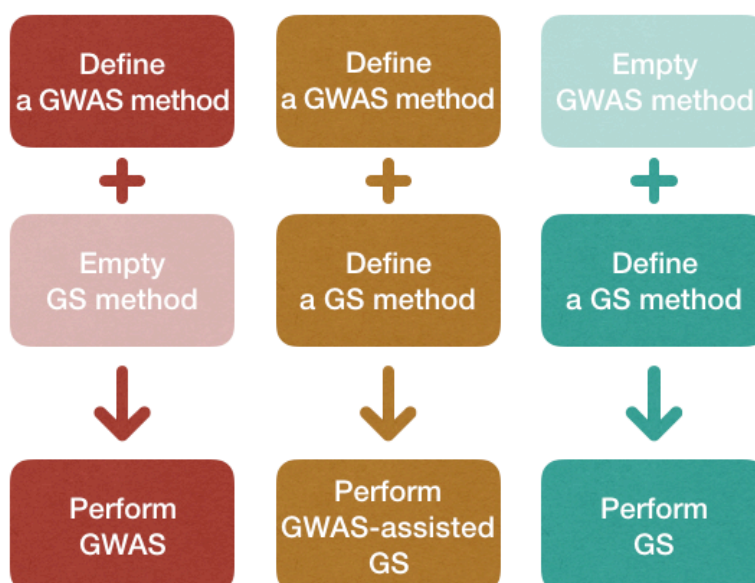


- After defining the analysis, user can start to run the procedure by clicking 'Run' at the pop-up menu of the project.

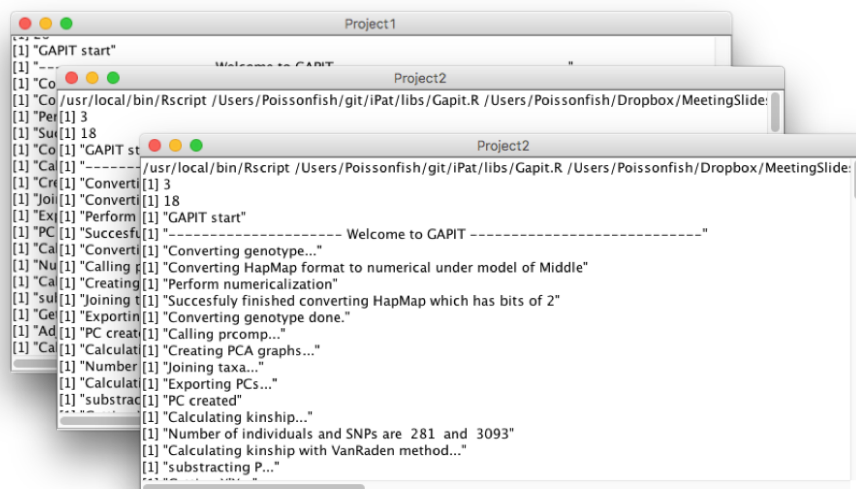


2.6 Run an analysis

- In iPat, users are allowed to do genomic studies such as GWAS, GS and GWAS-Assisted GS (Associated SNPs reported by GWAS will be treated as fixed effect in GS). iPat will detect the project configuration and decide which analysis should be implemented afterward.



- Each project will generate a console window while running the analysis. User can track the progress of the task from window messages.
- iPat also capable of multitasking. Users can arrange another project even when the previous one have not done yet.



2.7 Inspect the result

- When iPat complete a project, the gear icon will show a green dot if the task run successfully without any error occurred. Otherwise it will show a red dot at its top-left to notify users that there're existing at least one error message during the analysis.

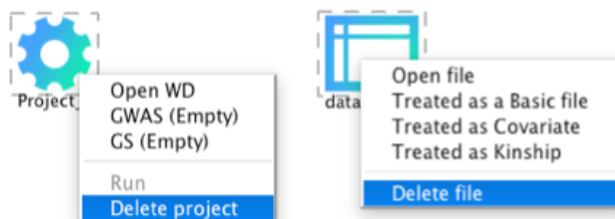


- Users can inspect the results by double clicking on the gear icon, which will direct users to the folder where output files generated.

| Shared Folder | | | | |
|---|--|---------------|-----------|-----------------|
| Name | | Date Modified | Size | Kind |
| GAPIT_dpollAllelic_Effect_Estimates.csv | | Today, 22:38 | 125 KB | Comm...t (.csv) |
| GAPIT_dpoll.Compression.multiple.group.pdf | | Today, 22:38 | 7 KB | PDF Document |
| GAPIT_dpoll.Df.tValue.StdErr.csv | | Today, 22:38 | 244 KB | Comm...t (.csv) |
| GAPIT_dpoll.GPS.BLUPvsPEV.pdf | | Today, 22:38 | 5 KB | PDF Document |
| GAPIT_dpoll.GWAS.Results.csv | | Today, 22:38 | 301 KB | Comm...t (.csv) |
| GAPIT_dpoll.Log.csv | | Today, 22:38 | 211 bytes | Comm...t (.csv) |
| GAPIT_dpoll.MAF.pdf | | Today, 22:38 | 7 KB | PDF Document |
| GAPIT_dpoll.Manhattan.Plot.Chromosomewise.pdf | | Today, 22:38 | 35 KB | PDF Document |
| GAPIT_dpoll.Manhattan.Plot.Genomewise.pdf | | Today, 22:38 | 35 KB | PDF Document |
| GAPIT_dpoll.Memory.Stage.csv | | Today, 22:38 | 22 KB | Comm...t (.csv) |
| GAPIT_dpoll.Optimum.pdf | | Today, 22:38 | 6 KB | PDF Document |
| GAPIT_dpoll.phenotype.csv | | Today, 22:36 | 3 KB | Comm...t (.csv) |
| GAPIT_dpoll.phenotype.view.pdf | | Today, 22:36 | 8 KB | PDF Document |
| GAPIT_dpoll.PRED.csv | | Today, 22:38 | 23 KB | Comm...t (.csv) |
| GAPIT_dpoll.QQ-Plot.pdf | | Today, 22:38 | 53 KB | PDF Document |
| GAPIT_dpoll.ROC.csv | | Today, 22:38 | 1 KB | Comm...t (.csv) |
| GAPIT_dpoll.ROC.pdf | | Today, 22:38 | 7 KB | PDF Document |
| GAPIT_dpoll.Timming.csv | | Today, 22:38 | 51 KB | Comm...t (.csv) |
| GAPIT_EarDiaAllelic_Effect_Estimates.csv | | Today, 22:41 | 125 KB | Comm...t (.csv) |

2.8 Remove files from iPat

- Users are allowed to remove projects or files from iPat in the pop-up menu through right clicking on them, or simply press `Backspace` (press `delete` on Mac) after selecting them.



3. GWAS and GS

Tools implemented in iPat allow users to do genome-wide associate study (GWAS) and genomic selection (GS). Currently GWAS can be performed by GAPIT, FarmCPU and PLINK, and GS can be done by GAPIT, rrBLUP and BGLR in iPat. Tables below are the input arguments available in iPat:

3.1 GAPIT

| Tab | Parameters | Definitions | Default |
|-------------|-----------------|---|-----------------------------|
| Covariates | Covariate names | Subsetting covariates data | All covariates are selected |
| GAPIT input | Model | Which linear model to be used in GWAS | GLM |
| GAPIT input | kinship.cluster | Clustering algorithm to group individuals based on their kinship | average |
| GAPIT input | kinship.group | Method to derive kinship among groups | Mean |
| Advance | SNP.fraction | Fraction of SNPs Sampled to Estimate Kinship and PCs | 1 |
| Advance | File.fragment | The Fragment Size to Read Each Time within a File | 512 |
| Advance | Model selection | Conduct Bayesian information criterion (BIC)-based model selection to find the optimal number of PCs for inclusion in the GWAS models | FALSE |

3.2 FarmCPU

| Category | Parameters | Definitions | Default |
|---------------|-----------------|---|-----------------------------|
| Covariates | Covariate names | Subsetting covariates data | All covariates are selected |
| FarmCPU input | method.bin | It uses fixed or optimized of possible QTN window size and number of possible QTNs selected into FarmCPU model. | static |
| FarmCPU input | maxLoop | Maximum number of iterations allowed | 10 |

3.3 PLINK

| Category | Parameters | Definitions | Default |
|-------------|-----------------|--|-----------------------------|
| Covariates | Covariate names | Subsetting covariates data | All covariates are selected |
| PLINK input | C.I. | The desired coverage for a confidence interval | 0.95 |

3.4 rrBLUP

| Category | Parameters | Definitions | Default |
|--------------|----------------------|--|-----------------------------|
| Covariates | Covariate names | Subsetting covariates data | All covariates are selected |
| rrBLUP input | Shrinkage estimation | Shrinkage estimation can improve the accuracy of genome-wide marker-assisted selection, particularly at low marker density (Endelman and Jannink 2012) | |
| rrBLUP input | impute.method | Imputation algorithm for missing values in markers data | mean |

3.5 BGLR

| Category | Parameter | Definitions | Default |
|----------|--|--|------------|
| Subset | Subset of traits data | Users can select all or partial of traits to be analyzed | All traits |
| BGLR | Regression model for predictor (Markers) | The regression type for the markers data | BRR |
| BGLR | response_type | Data type of the response (y) | gaussian |
| BGLR | nIter | The number of iterations of the sampler | 1200 |
| BGLR | burnIn | The number of samples discarded | 200 |
| BGLR | thin | The number of thinning | 5 |

4 Support

- If there is any difficulty on iPat, please leave your question in the page of [issue report](#).
- Or you can directly send an email to the author [James Chen](#)

5 Citation

- Bradbury,P.J. et al. (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics, 23, 2633–2635.
- Endelman,J. (2011) Ridge regression and other kernels for genomic selection in the R package rrBLUP. Plant Genome, 4, 250–255.
- Kang,H.M. et al. (2008) Efficient control of population structure in model organism association mapping. Genetics, 178, 1709–1723.
- Liu,X. et al. (2016) Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLoS Genet., 12, e1005767.
- Purcell,S. et al. (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet, 81, 559–575.
- Tang,Y. et al. (2016) GAPIT Version 2: An Enhanced Integrated Tool for Genomic Association and Prediction. Plant J., 9.