

User Manual for



Intelligent Prediction and Association Tool

(Version 1.2)

Last updated on May 26, 2017

Table of Contents

1. [Getting start](#)

- 1-1 [Operation environment](#)
- 1-2 [Windows users](#)
- 1-3 [Mac OS users](#)
- 1-4 [Launch iPat](#)

2. [Interface](#)

- 2-1 [Import files](#)
- 2-2 [Create a project](#)
- 2-3 [File formats](#)
- 2-4 [Covariates and kinship](#)
- 2-5 [Define input arguments](#)
- 2-6 [Run an analysis](#)
- 2-7 [Check the results](#)
- 2-8 [Files remove](#)

3. [GWAS, GS and BSA](#)

- 3-1 [GAPIT](#)
- 3-2 [FarmCPU](#)
- 3.3 [PLINK](#)
- 3.4 [rrBLUP](#)
- 3.5 [BGLR](#)
- 3.6 [BSA](#)

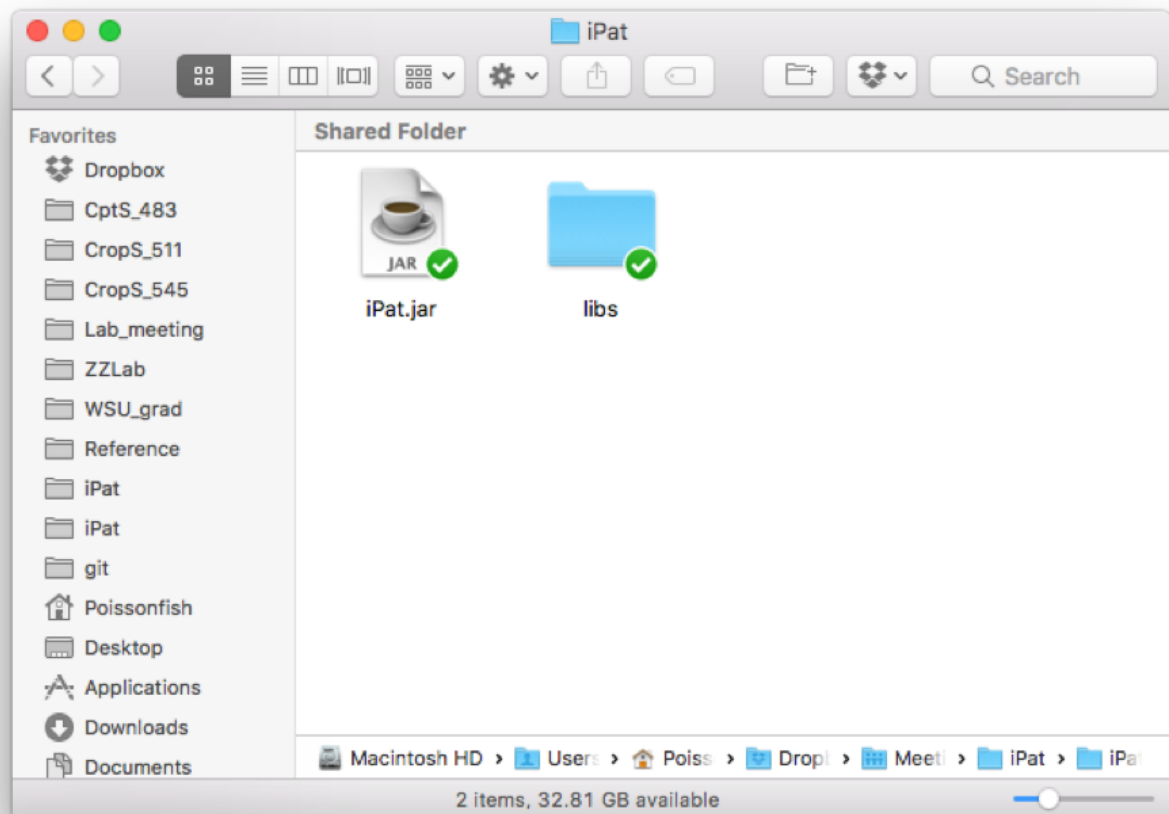
4. [Support](#)

5. Citation

1. Getting start

1.1 Operation environment

- Before launching iPat, remember to place folder 'libs' to the directory where 'iPat.jar' exists. iPat can function normally only when both 'libs' and 'iPat.jar' are located in the same folder.

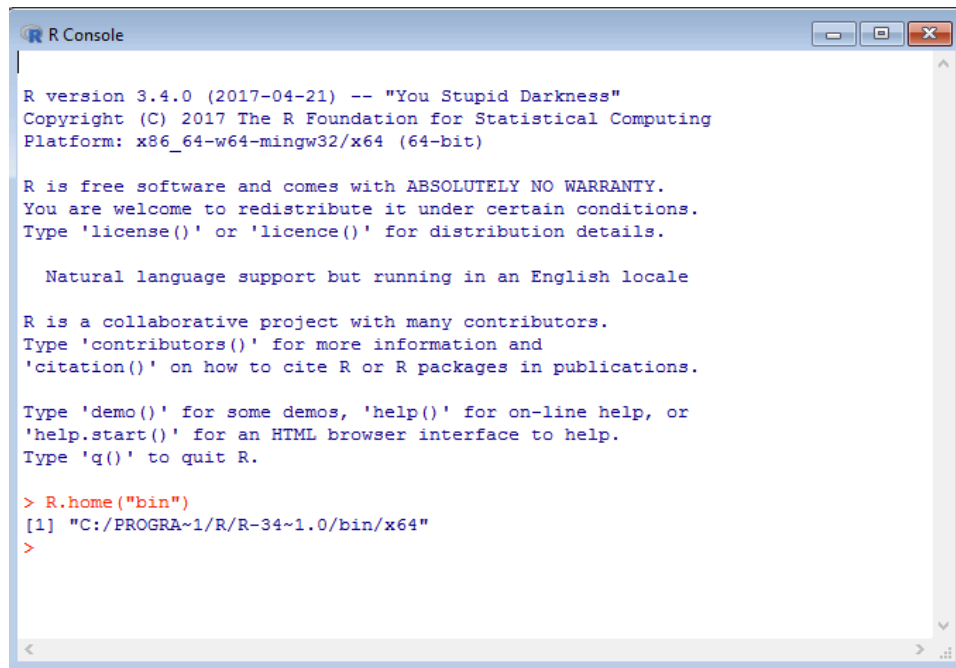


- The operation environment need to meet the following requirement:
 - Operation System: Windows and Mac OS X.
 - [Java Runtime Environment \(JRE\)](#): Version 8 or later.
 - [R](#): Version 3.4.0 or later.

1.2 Windows users

- If you can call R from the commnad-line window (cmd.exe) by typing "R" or "r", then you can skip to section 1.4. 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.

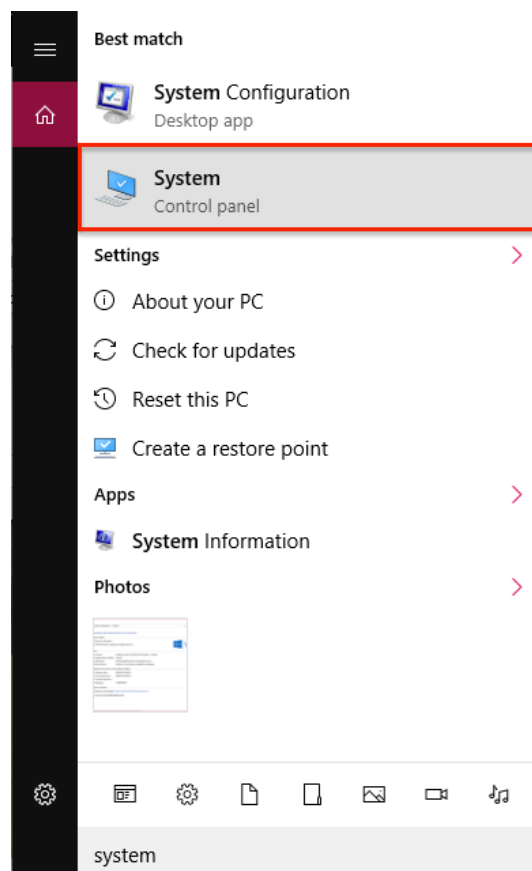
Natural language support but running in an English locale

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.

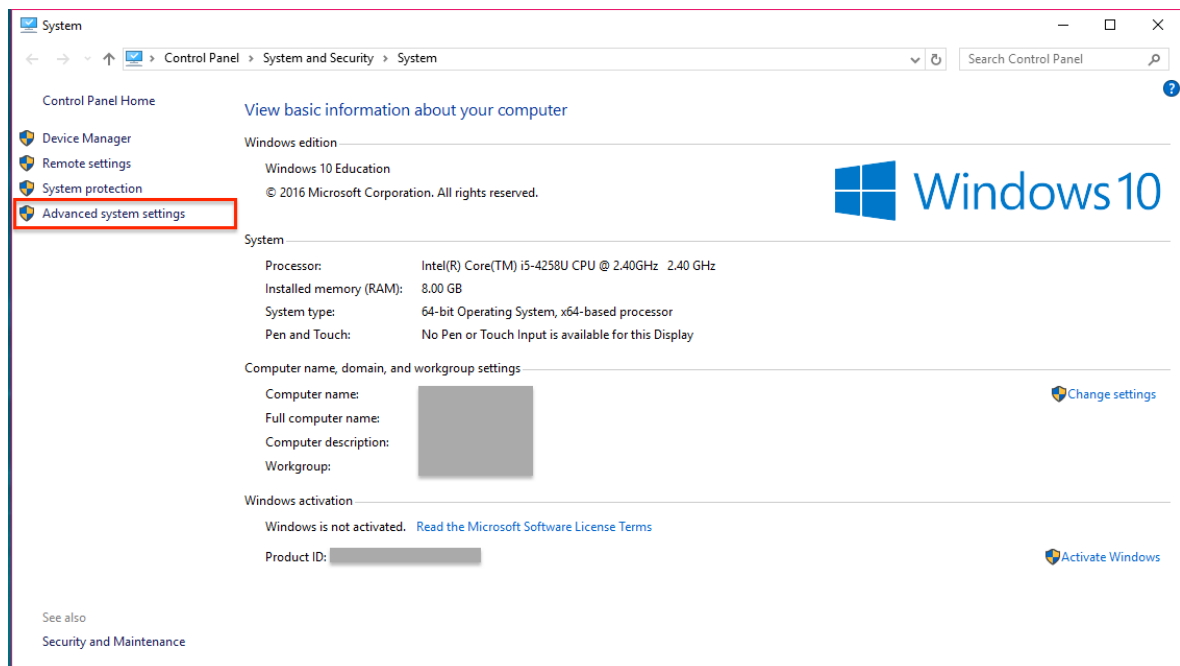
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

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

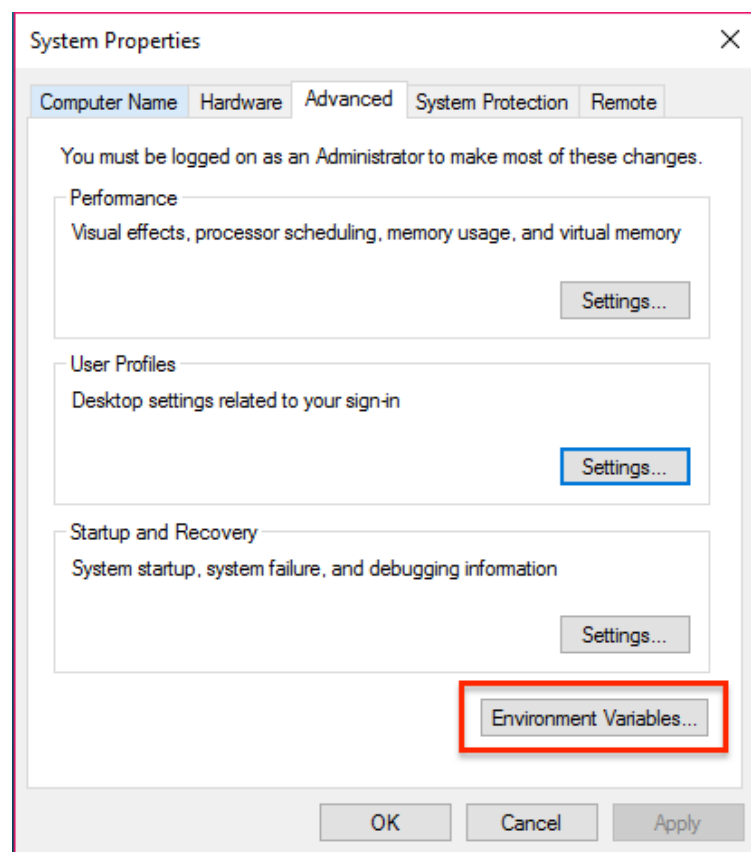
- Search keyword "system" from Windows, and open "System".



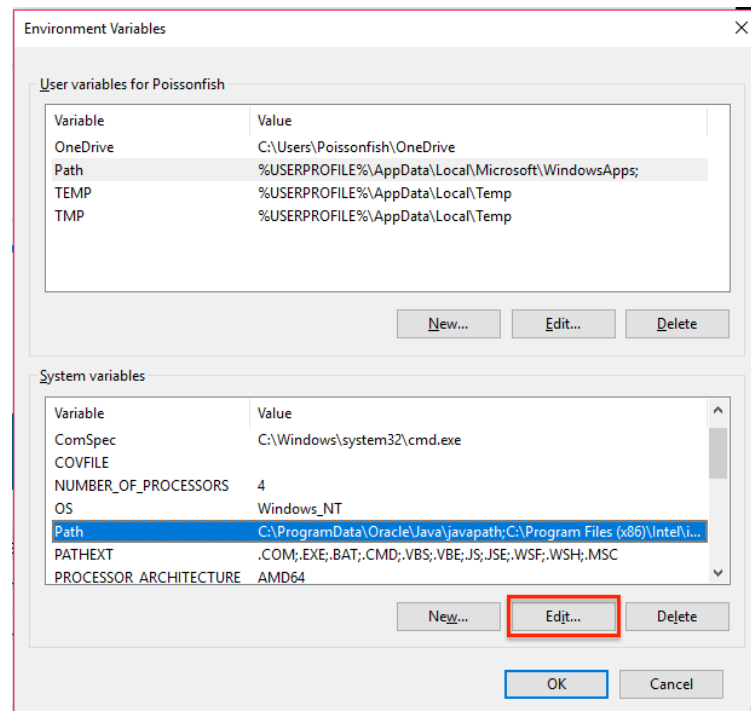
- Then select "Advanced system settings" at the left side of the panel.



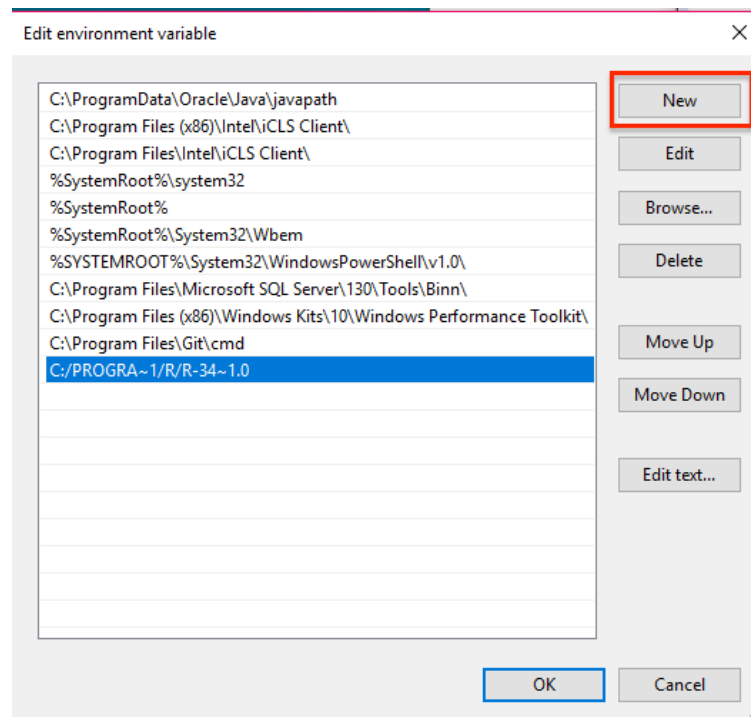
- And click "Environment Variable..." at the bottom-right area.



- 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.



1.3 Mac OS users

- For the users who run iPat on Mac OS, there's no need to do extra adjustment in your system.

1.4 Launch iPat

- After setting up the required environment mentioned above, iPat is now ready to go. Double clicking on 'iPat.jar' to launch iPat.

2. Interface

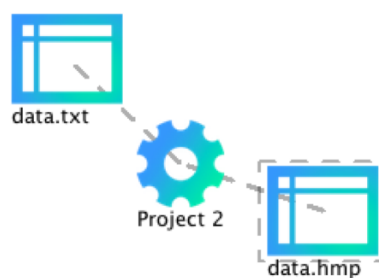
2.1 Import files

- 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.
- Build linkages between this project and all the files planned to be analyzed. A linkage can be built by dragging one object over one another.



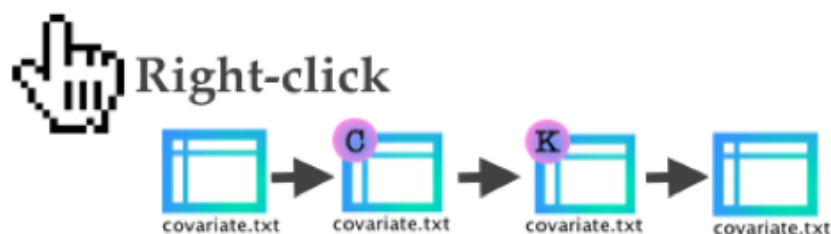
2.3 File formats

- iPat can recognize and work fine with different formats, which include hapmap, numeric, vcf and plink.
- Imported file set need to have identical names and correct extension name if they need to be converted to a proper format.** For example, if you want to perform GWAS using VCF format in FarmCPU, files set should be named as: data.vcf and data.txt. The table below shows examples of files and its extension name for the corresponded format:

Format	Genotype	Phenotype	Other information
Hapmap	.hmp	.txt	None
Numeric	.dat	.txt	.map
VCF	.vcf	.txt	None
PLINK	.bed	.txt	.fam .bim

2.4 Covariates and kinship

- Users are allowed to add covariates into the project. Right clicking on the file can change the file type.

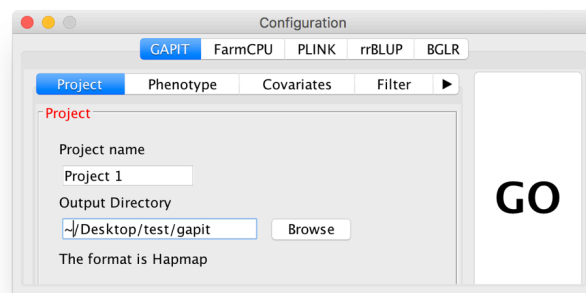


- Label “C” stands for covariate, and Label “K” is indicated as kinship.



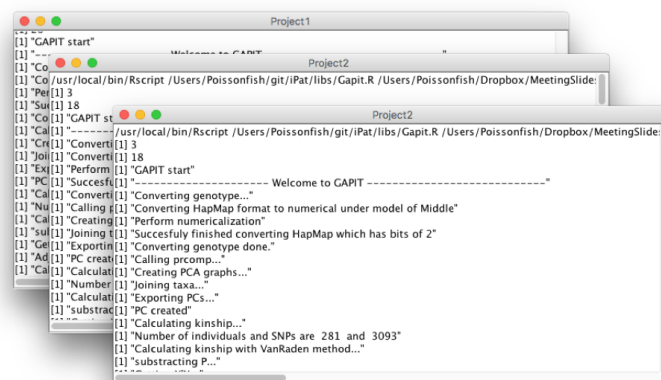
2.5 Define input arguments

- After linking every files needed in the project, right click on the project to open a configuration panel.
- At the top of the panel, users can select a tool to perform GWAS or genomic prediction.
- A hint will pop up for each argument when the cursor hovers over the arguments' name.



2.6 Run an analysis

- After defining the analysis, user can start to run the procedure by clicking 'GO' at the top of the panel.
- Each task 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 Check 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.

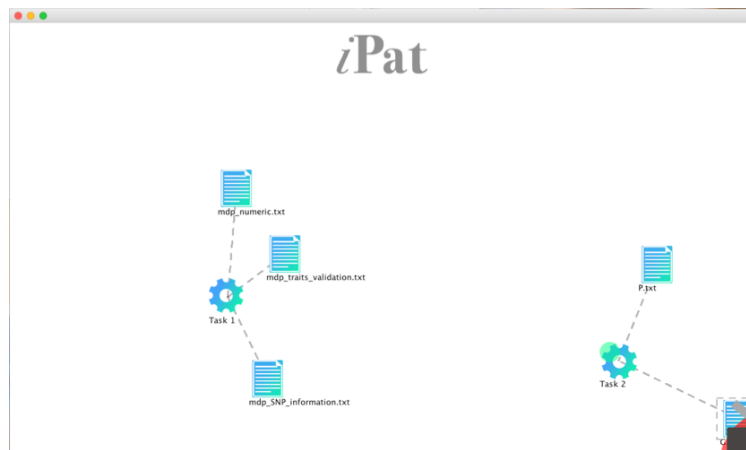


- Users can check the result by double clicking on the gear icon, which will direct users to the folder where the output files generated.

Shared Folder				
Name		Date Modified	Size	Kind
GAPIT...dpoll.Allelic_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...EarDia.Allelic_Effect_Estimates.csv		Today, 22:41	125 KB	Comm...t (.csv)

2.8 Files remove

- Users can remove objects and linkage by typing “Del” after selecting
- For linkages, the line will become solid when it's selected.
- For objects, there will be a dashed line surrounded to indicate that the object is selected.
- Users can also drag the linkages or objects to the bottom-right corner, a hidden trashcan will show up for deletion.



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 and rrBLUP in iPat. Tables below are the input arguments available in iPat:

3.1 GAPIT

Category	Parameters	Definitions	Default
Subset	Subset of traits data	Users can select all or partial of traits to be analyzed	All traits
Subset	Subset of chromosomes	Users can select all or partial of chromosomes to be analyzed	All
Covariates	PCA.count	How many of PCs should be treated as covariates	3
Covariates	Inheritable covariate	If there's a user-input covariates, users can specify how many columns are inheritable.	All
Quality control	By missing rate	Users can do a quality control on the marker set by missing rate.	NULL
Quality control	By MAF	Users can do a quality control on the marker set by minor allele frequency (MAF).	NULL
GWAS	Model	Which linear model to use in GWAS	GLM
GWAS	kinship.cluster	Clustering algorithm to group individuals based on their kinship	average
GWAS	kinship.group	Method to derive kinship among groups	Mean
GWAS	SNP.fraction	Fraction of SNPs Sampled to Estimate Kinship and PCs	1
GWAS	File.fragment	The Fragment Size to Read Each Time within a File	512

3.2 FarmCPU

Category	Parameters	Definitions	Default
Subset	Subset of traits data	Users can select all or partial of traits to be analyzed	All traits
Subset	Subset of chromosomes	Users can select all or partial of chromosomes to be analyzed	All
Covariates	PCA.count	How many of PCs should be treated as covariates	3
Covariates	Inheritable covariate	If there's a user-input covariates, users can specify how many columns are inheritable.	All
Quality control	By missing rate	Users can do a quality control on the marker set by missing rate.	NULL
Quality control	By MAF	Users can do a quality control on the marker set by minor allele frequency (MAF).	NULL
GWAS	method.bin	It uses fixed or optimized of possible QTN window size and number of possible QTNs selected into FarmCPU model.	static
GWAS	maxLoop	Maximum number of iterations allowed	10

3.3 PLINK

Category	Parameters	Definitions	Default
Subset	Subset of chromosomes	Users can select all or partial of chromosomes to be analyzed	All
Quality control	By missing rate	Users can do a quality control on the marker set by missing rate.	NULL
Quality control	By MAF	Users can do a quality control on the marker set by minor allele frequency (MAF).	NULL
GWAS	C.I.	The desired coverage for a confidence interval	0.95

3.4 rrBLUP

Category	Parameters	Definitions	Default
Subset	Subset of traits data	Users can select all or partial of traits to be analyzed	All traits

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	Number of iterations	Number of iterations	1200
BGLR	Burn-In	Iteration of burn-in	200

3.6 BSA

Category	Parameter	Definitions	Default
BSA	Windows Size	A smoothing coefficient for G statistics	5KB

4 Support

- User can download demo files from [here](#).
- 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.