



**GWAS and GS Are as Easy  
as Clicking and Dragging  
with iPat**

World Congress On Genetics Applied to  
Livestock Production (WCGALP, 11th)

**Chunpeng (James) Chen**

Ph.D. Student

Department of Crop and Soil Sciences

Washington State University



# The Challenges of Biologists

Programming skills

Big data

Formatting

Modeling

...



iPat makes it easier for biologists to analyze data and stay focused on biology

# The Challenges of Biologists

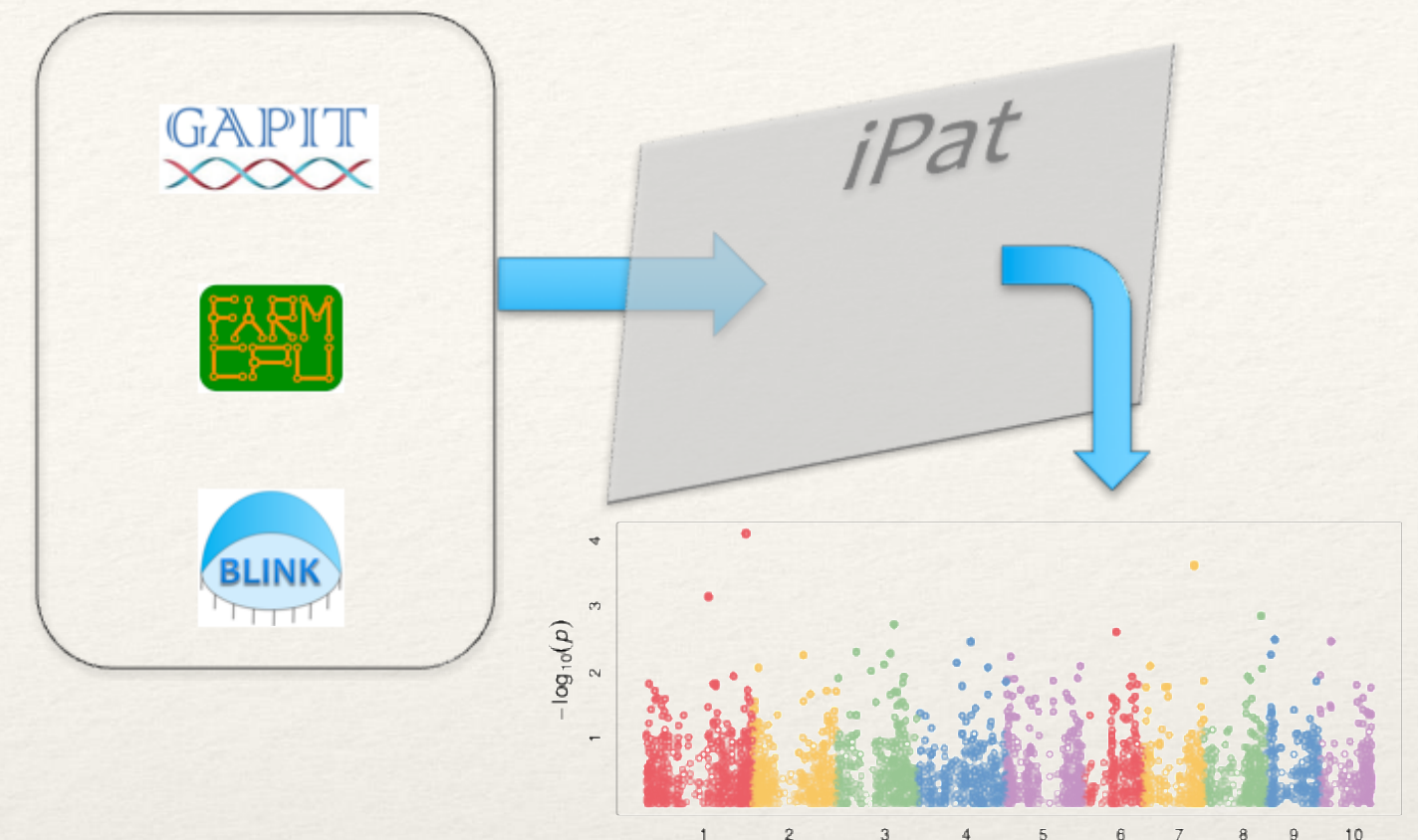
Programming skills

Big data

Formatting

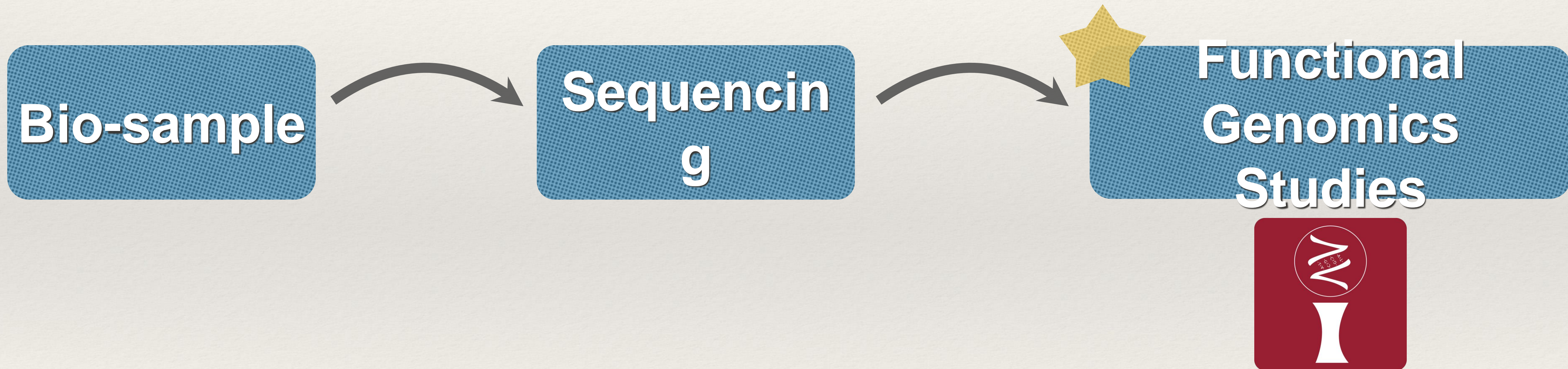
Modeling

...

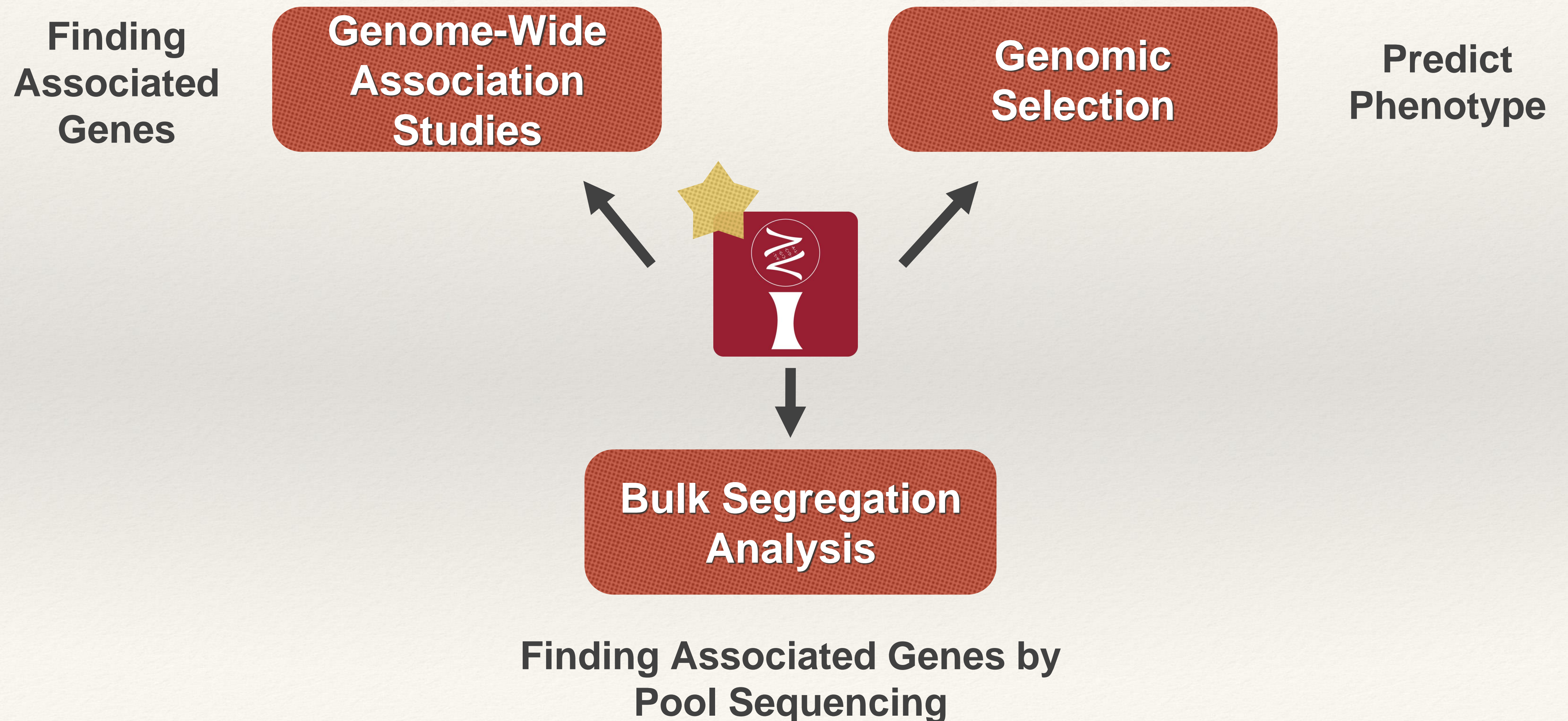


iPat makes it easier for biologists to analyze data and stay focused on biology

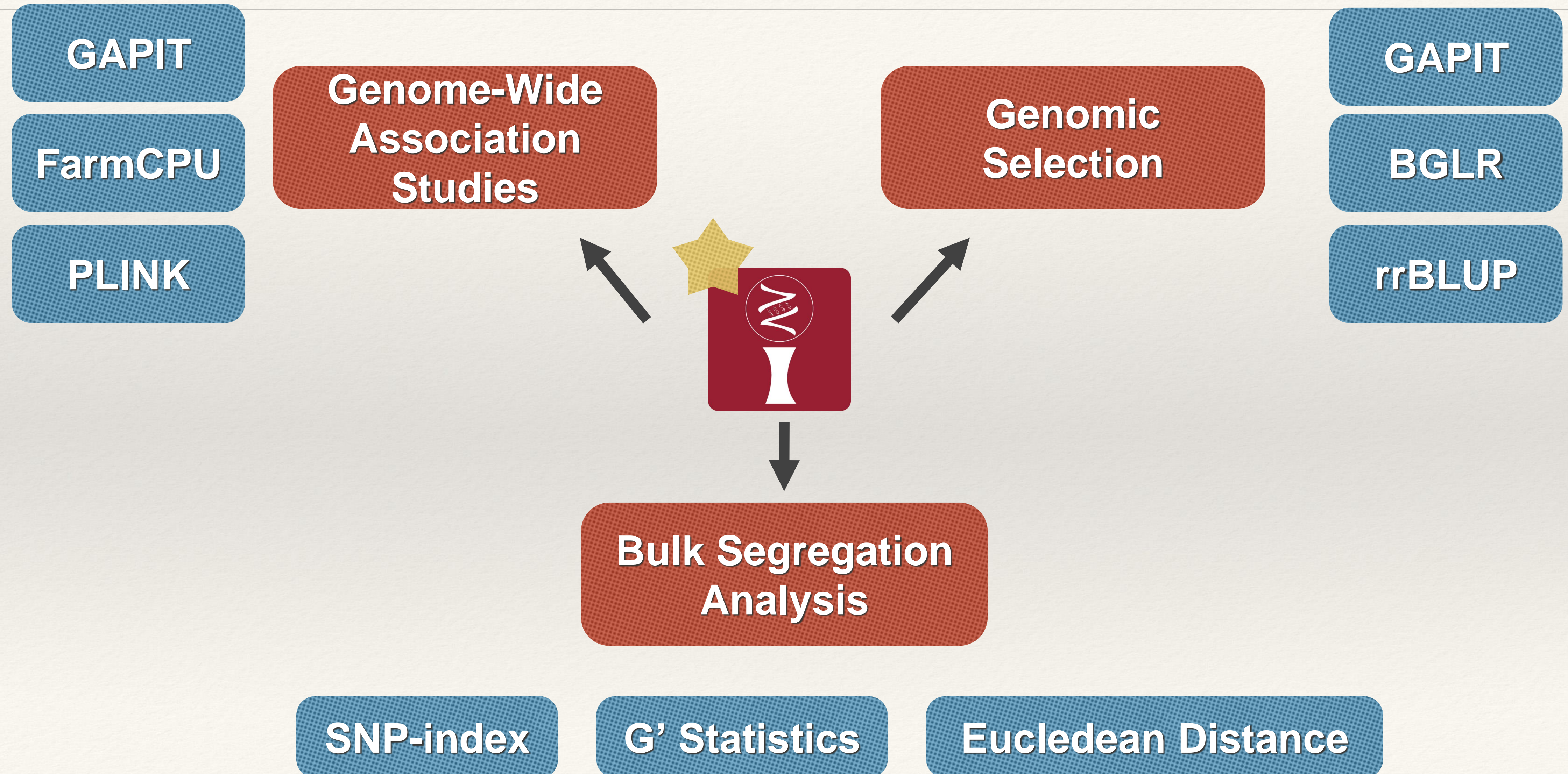
# The Focus of iPat



# The Focus of iPat



# Embedded Tools in One Place!



# Unfortunately... not all formats work on them

Tool	Supported Format	
GAPIT	Numeric	Hapmap
FarmCPU	Numeric	
BGLR	Numeric	
rrBLUP	Numeric	
PLINK	PLINK	Binary

# Unfortunately... not all formats work on them

Tool	Supported Format	
GAPIT	Numeric	Hapmap
FarmCPU	Numeric	
BGLR	Numeric	
rrBLUP	Numeric	
PLINK	PLINK	Binary

?

VCF



# Inspect the format

Hapmap

```
rs alleles chrom pos strand assembly center protLSID assayLSID panel QCcode 33-16 38-11 4
CML157Q CML158Q CML218 CML220 CML228 CML238 CML247 CML254 CML258 CML261 CML264 CML277
KI43 KI44 KY21 KY226 KY228 L317 L578 M14 M162W M37W MEF156-55-2 M017 M018W M01W M024W M
PA880 PA91 R109B R168 R177 R229 R4 SA24 SC213R SC357 SC55 SD40 SD44 SG1533 SG18 T23
PZB00859.1 A/C 1 157104 + AGPv1 Panzea NA NA maize282 NA CC CC CC CC AA CC AA AA
CC CC CC CC CC AA CC AA NN AA CC CC CC CC CC AA CC CC CC CC AA CC CC AA
CC CC CC CC CC CC AA CC CC AA CC CC AA AA AA AA AA CC CC CC CC CC CC CC CC
PZA01271.1 C/G 1 1947984 + AGPv1 Panzea NA NA maize282 NA CC GG CC GG CC CC CC CC
GG GG NN GG GG GG GG CC CC GG CC GG GG CC GG CC CC CC CC GG CC CC CC CC
GG GG NN GG GG CC GG CC GG GG GG NN CC CC CC GG GG GG GG GG GG CC CC CC
```

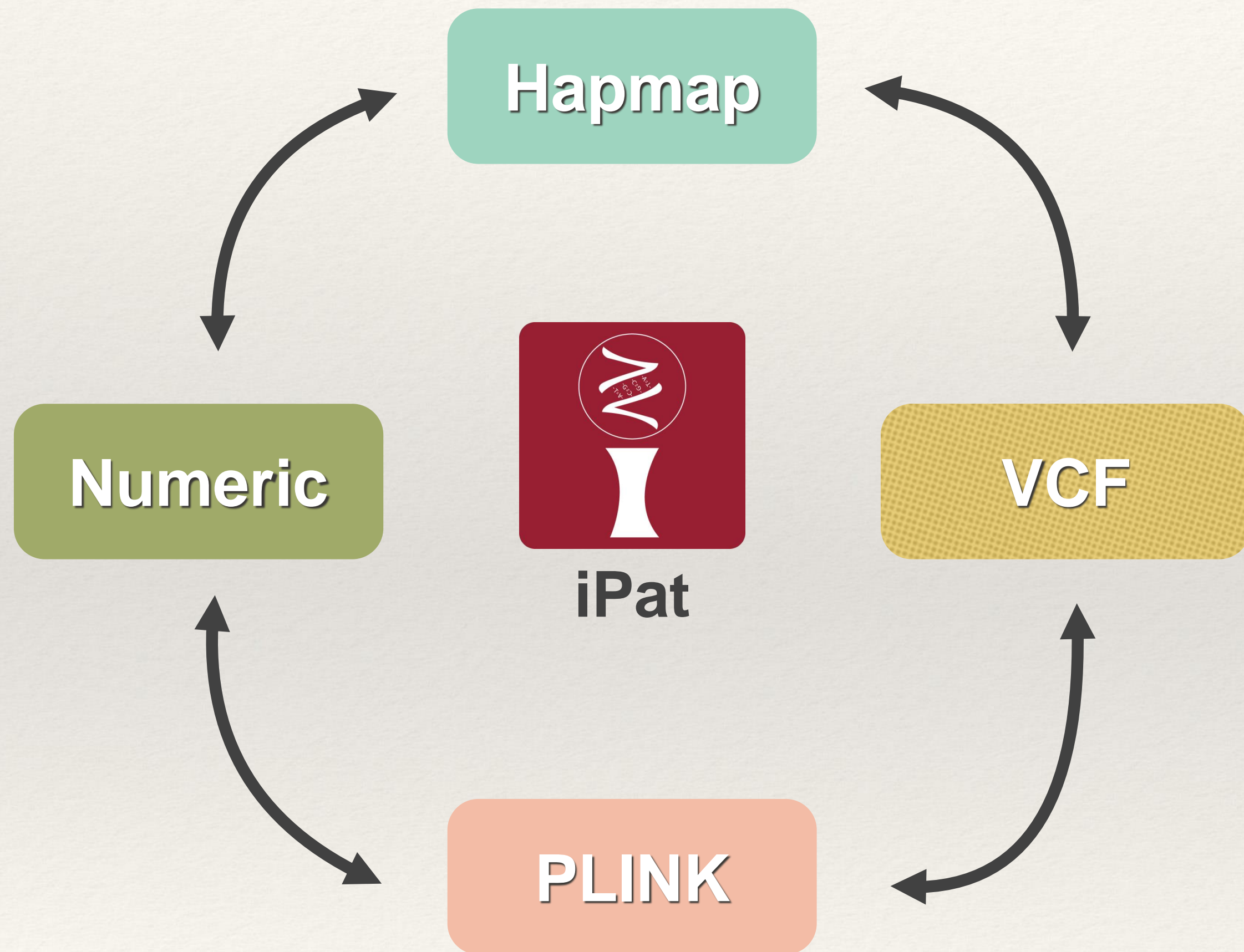
PLINK

```
abc1 abc1 0 0 1 1 T T T T T T A A A A C C T T G G C C T T G G A
A A G G T T G G C C A A C C G G A A T T T T C C A A C C G A T T G
A A T T C C C C A A A A T T C C C C T T C C A A G T A A C C C C C
A A T T C C G T C C G G A A C C A A C C C C G G A A T C C A C C C
A A C C G G T T G G G A A A G G C A A A G G A A C C C T C C T T C
T T A A G G T T T T G G T T T T C T C C C C T T G G C C C C G G G
C C G G T T A A C C C C C C C C A A G G C C C C G G G G G G T T A
```

VCF

```
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT 33-16 38-11 4226 4722 A188 A214N A239 A272 A441-5 A554 A556 A6 A619 A6
CML228 CML238 CML247 CML254 CML258 CML261 CML264 CML277 CML281 CML287 CML311 CML314 CML321 CML322 CML323 CML328 CML3
L317 L578 M14 M162W M37W MEF156-55-2 M017 M018W M01W M024W M044 M045 M046 M047 MOG MP339 MS1334 MS153 MS71 MT42 N192 N2
R229 R4 SA24 SC213R SC357 SC55 SD40 SD44 SG1533 SG18 T232 T234 T8 TX303 TX601 TZI10 TZI11 TZI16 TZI18 TZI25 TZI8 TZI9
1 157104 PZB00859.1 C A . PASS . GT 0/0 0/0 0/0 0/0 1/1 0/0 1/1 1/1 0/0 0/0 0/0 1/1 0/0 0/0 0/0 1/1 0/0 0/0 0/0 1/1 ./ 1/1 0/0 0/0 0/0 0/0
0/0 1/1 ./ 1/1 0/0 0/0 0/0 0/0 0/0 1/1 0/0 0/0 0/0 0/0 1/1 0/0 0/0 1/1 1/1 0/0 1/1 0/0 1/1 1/1 0/0 1/1 0/0 0/0
1/1 0/0 0/0 1/1 0/0 0/0 1/1 1/1 1/1 1/1 0/0 0/0 0/0 0/0 0/0 0/0 0/0 0/0 1/1 0/0 0/0 0/0 0/0 0/0 0/0 0/0 0/0
1 1947984 PZA01271.1 G C . PASS . GT 1/1 0/0 1/1 0/0 1/1 1/1 1/1 1/1 1/1 0/0 1/1 1/1 0/0 1/1 1/1 1/1 1/1 0/0 0/0 0/0 1/1 1/1 1/1
0/0 1/1 1/1 0/0 1/1 0/0 0/0 1/1 0/0 1/1 1/1 1/1 1/1 0/0 1/1 1/1 1/1 1/1 0/0 0/0 1/1 1/1 1/1 1/1 1/1 1/1 1/1
0/0 1/1 0/0 0/0 0/0 ./ 1/1 1/1 1/1 0/0 0/0 0/0 0/0 0/0 0/0 1/1 1/1 1/1 1/1 0/0 0/0 1/1 0/0 0/0 0/0 0/0 1/1 0/0
1 2914066 PZA03613.2 T G . PASS . GT 1/1 1/1 1/1 1/1 1/1 0/0 0/0 0/0 1/1 0/0 1/1 0/0 1/1 0/0 0/0 0/0 0/0 0/0 0/0 0/0 0/0 0/0 0/0 0/0 0/0
0/0 0/0 0/0 0/0 0/0 0/0 1/1 0/0 0/0 1/1 0/0 1/1 1/1 0/0 0/0 0/0 0/0 1/1 0/0 0/0 0/0 0/0 0/0 0/0 0/0 0/0 1/1 1/1
0/0 0/0 1/1 0/0 0/0 ./ 1/1 ./ 0/0 0/0 1/1 0/0 0/0 0/0 1/1 0/0 0/0 0/0 1/1 0/0 0/0 0/0 0/0 ./ 0/0 0/0 0/0 1/1 0/0 1/1
```

# Format Free



**No need to spend effort  
on format conversion**

---

# Let's work with iPat step by step!

---

## Task : GWAS



---

# Import Data

---

```
genotype = read.table("data.hmp", header = T, sep = "\t")  
phenotype = read.table("data.txt", header = T, sep = "\t")
```

---

# Import Data

---

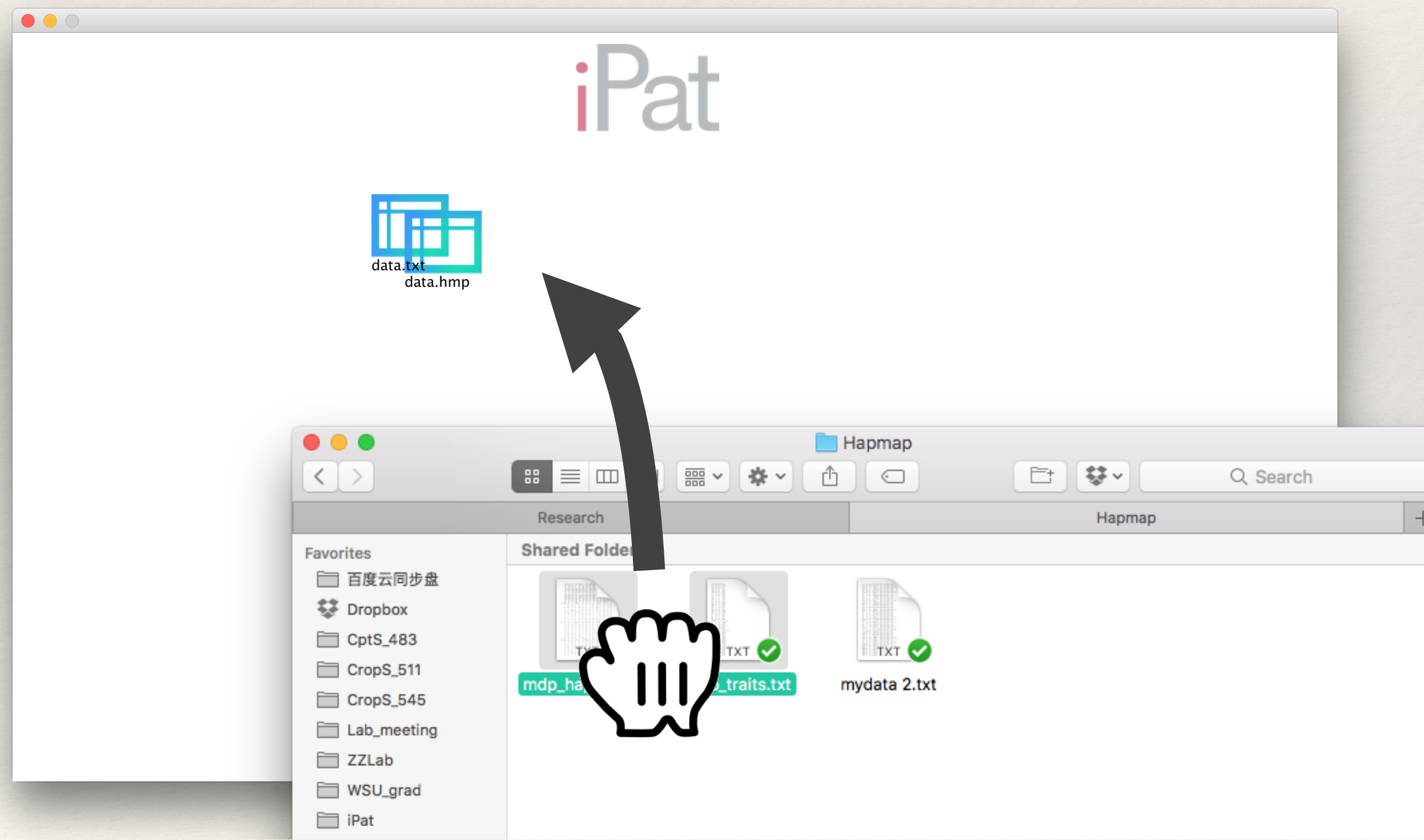
```
genotype = read.table("data.hmp", header = T, sep = "\t")  
phenotype = read.table("data.txt", header = T, sep = "\t")
```

**Does data contain a header?**

**What's the delimiter? Tab? Comma? Space?**

# Drag and Drop To Import

**Detect the header and delimiter automatically**

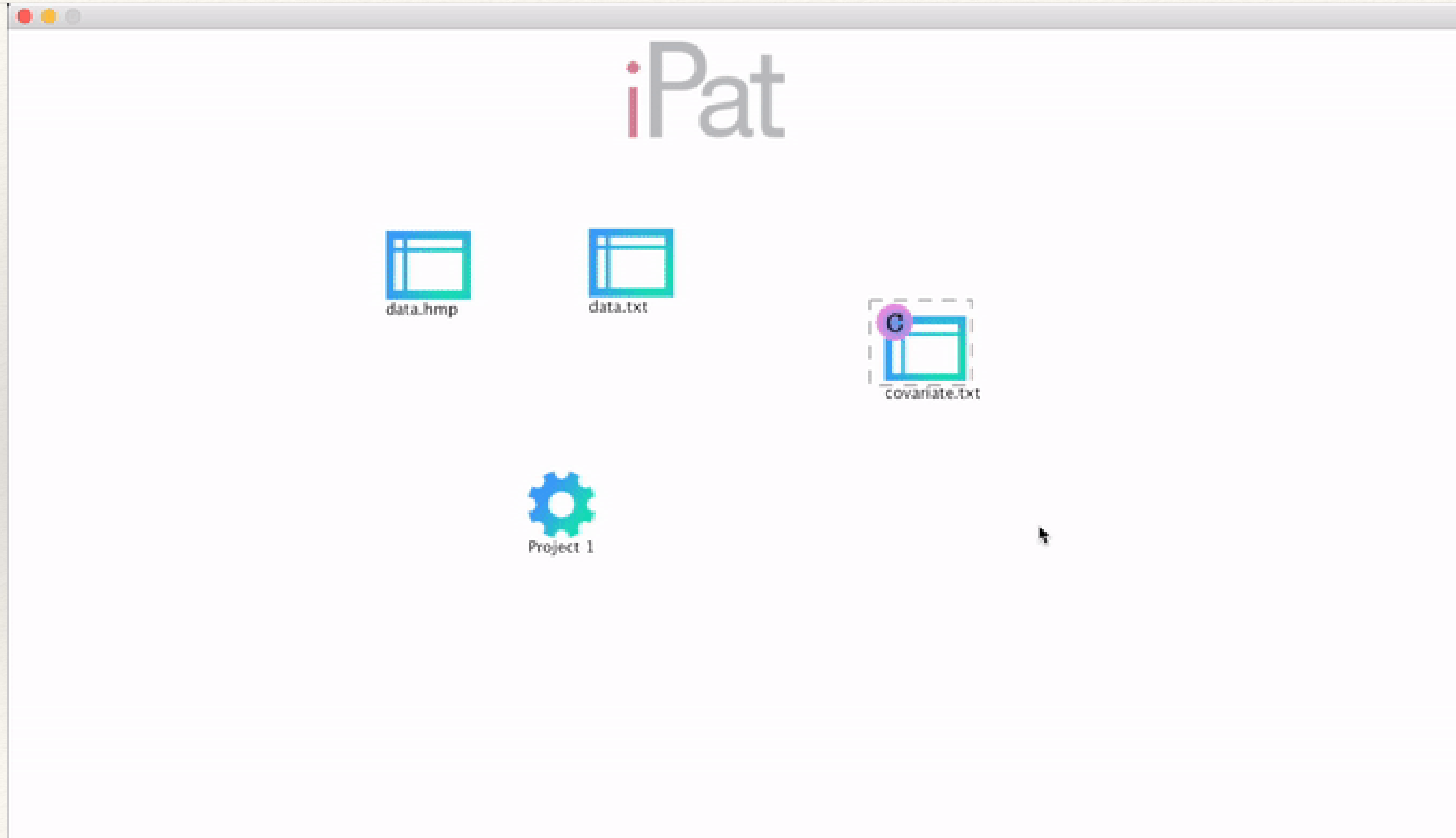


# Inspect

The screenshot shows the iPat software interface. In the center, there is a file explorer window with three files: data.txt, data.hmp, and covariate.txt. A hand cursor is pointing at the covariate.txt file, with the text "Double-click" next to it. To the right, a terminal window titled "covariate.txt" displays a table of data. The table has four columns: Trait, PC1, PC2, and PC3. The data is organized into a table with 20 rows of trait names and their corresponding PC values.

Trait	PC1	PC2	PC3
207	-24.61260272	2.929758052	2.378974655
764	19.01517623	35.99961149	-3.534235921
778	-18.75445031	5.755455338	1.626373633
779	-21.25806895	1.568326634	-0.075703399
787	-21.22114006	-8.02247851	2.041586307
790	19.77723042	37.5455684	-3.211485814
792	14.40888008	34.14099124	-1.297498297
793	26.343461	40.85397514	-4.159001978
794	27.63342477	42.8150698	-4.210617699
807	-20.87368399	5.337665874	4.869219547
991	-14.45796695	16.64651691	3.938890296
4226	-25.71476283	-2.722481341	1.680169305
4722	-28.01049541	-0.793563748	-2.837014414
5707	-15.09953578	-19.12766408	-1.835536914
6103	-8.426793982	32.0256389	3.89467021
11430	-24.27496719	-3.252969147	5.993197571
78004	15.9254961	39.6820946	-2.055236403
78010	16.8062231	41.36296889	-3.696416377
2MA22	-26.99434312	-0.938612674	4.539695279
33-16	-23.85302995	-0.979564292	-1.109431878
38-11	-24.41975043	1.659113068	7.534738524
4676A	-16.55117283	15.50632366	-0.025563177
78002A	23.43588599	41.78904767	-4.373400834
78371A	-16.24248123	-14.90567141	7.229065726
A188	-25.58528519	-3.818950151	-1.820976303
A214N	-15.20327431	13.69027633	-4.515143144
A239	-24.57799874	2.693263403	17.96394809
A272	-31.245357	-4.388834441	-10.89999527
A340	-28.22410628	-1.792121188	-2.57761375

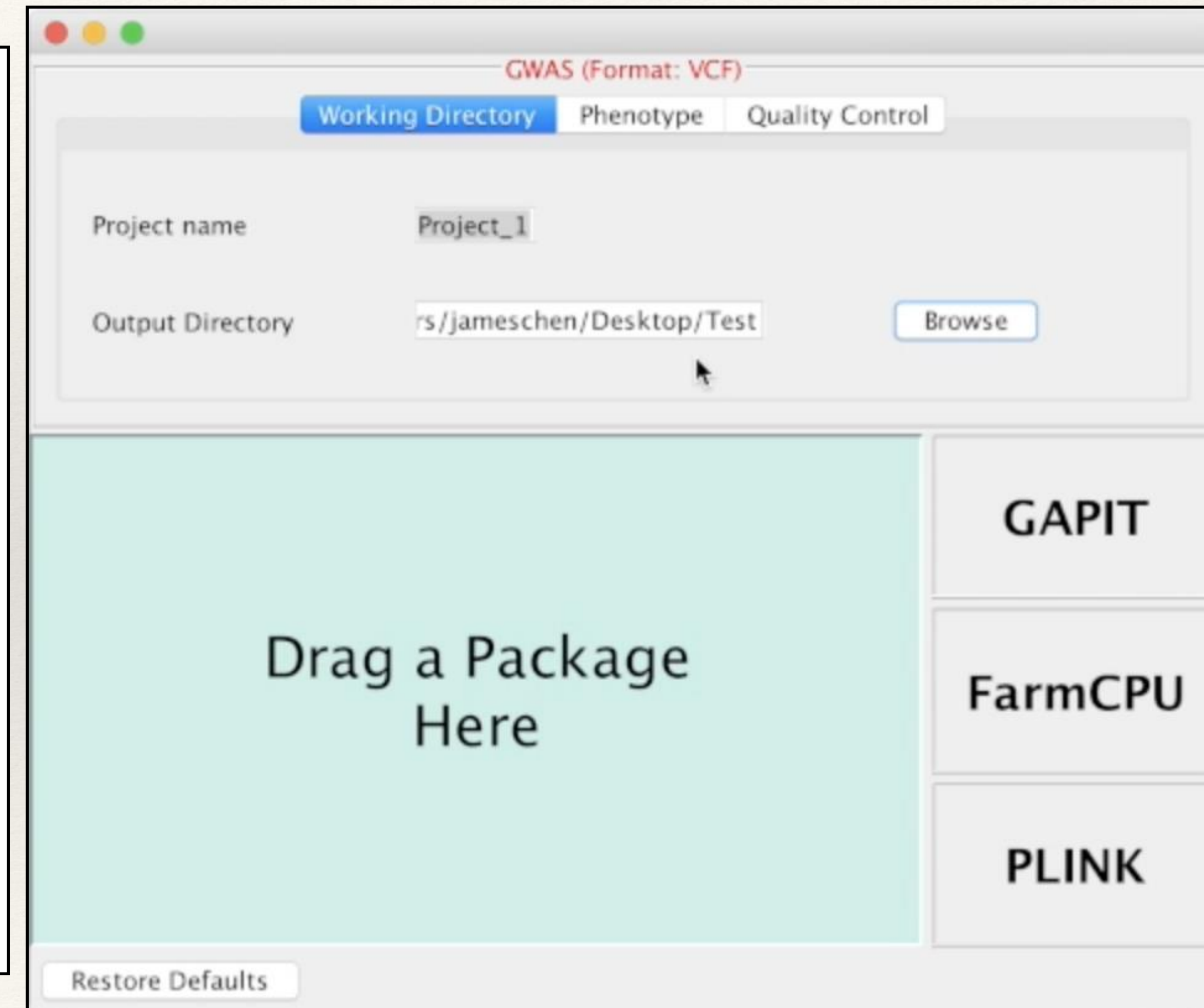
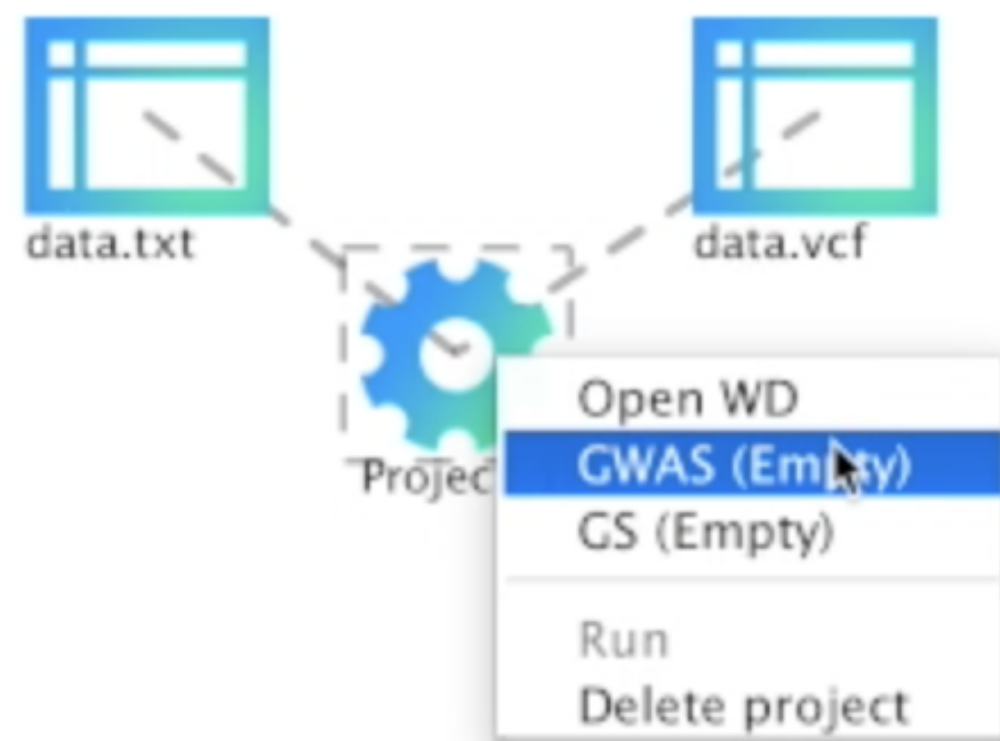
# Build a Module





# Define Your Analysis

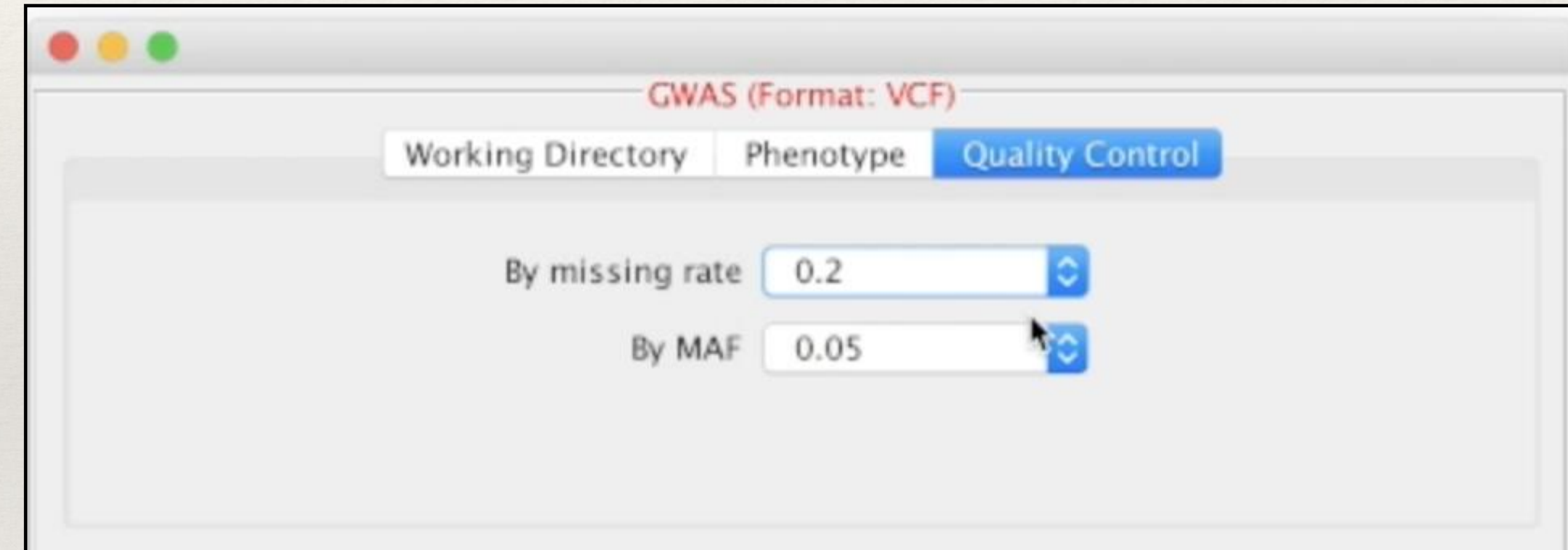
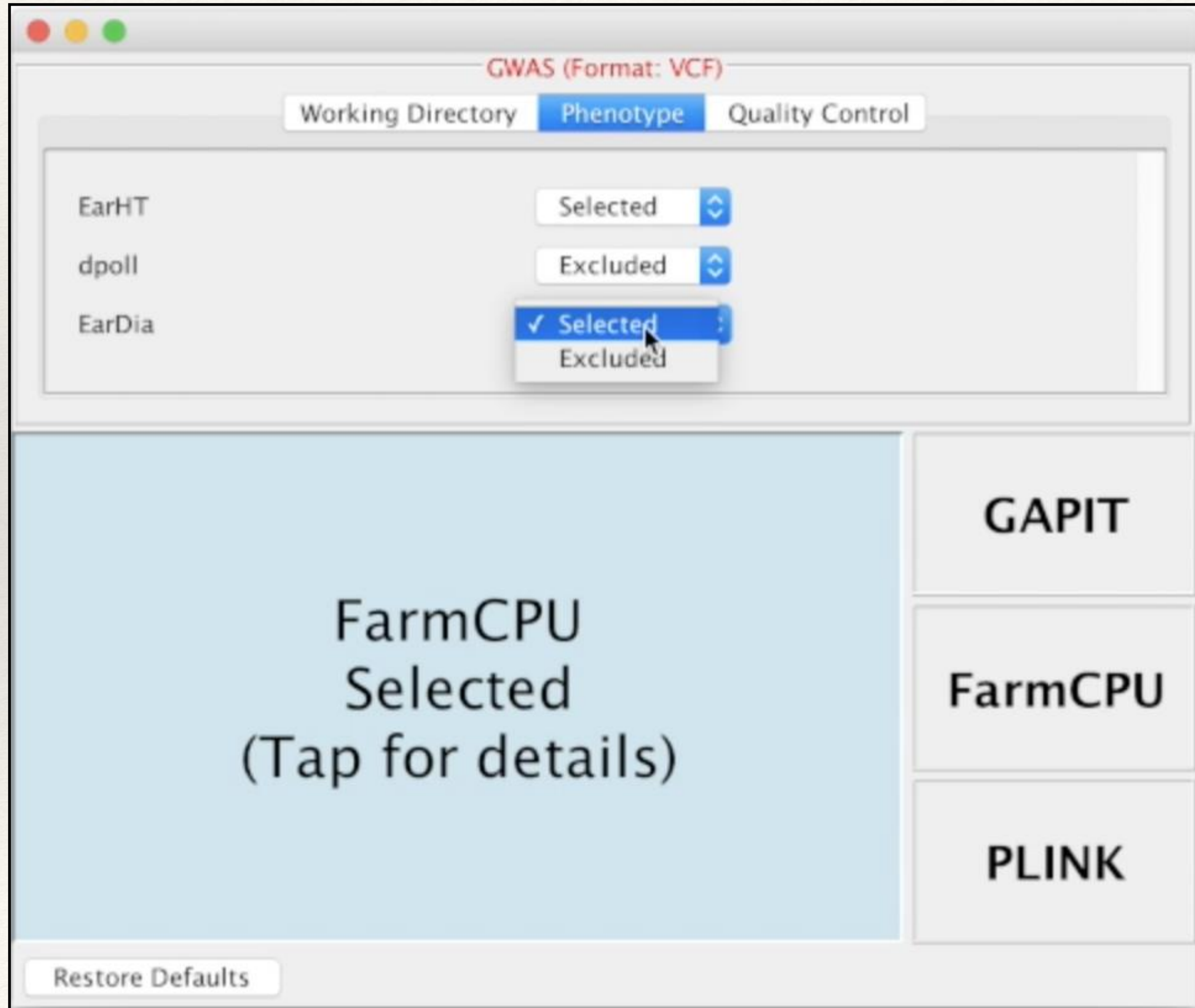
iPat



# Missing Rate? Minor Allele Frequency?

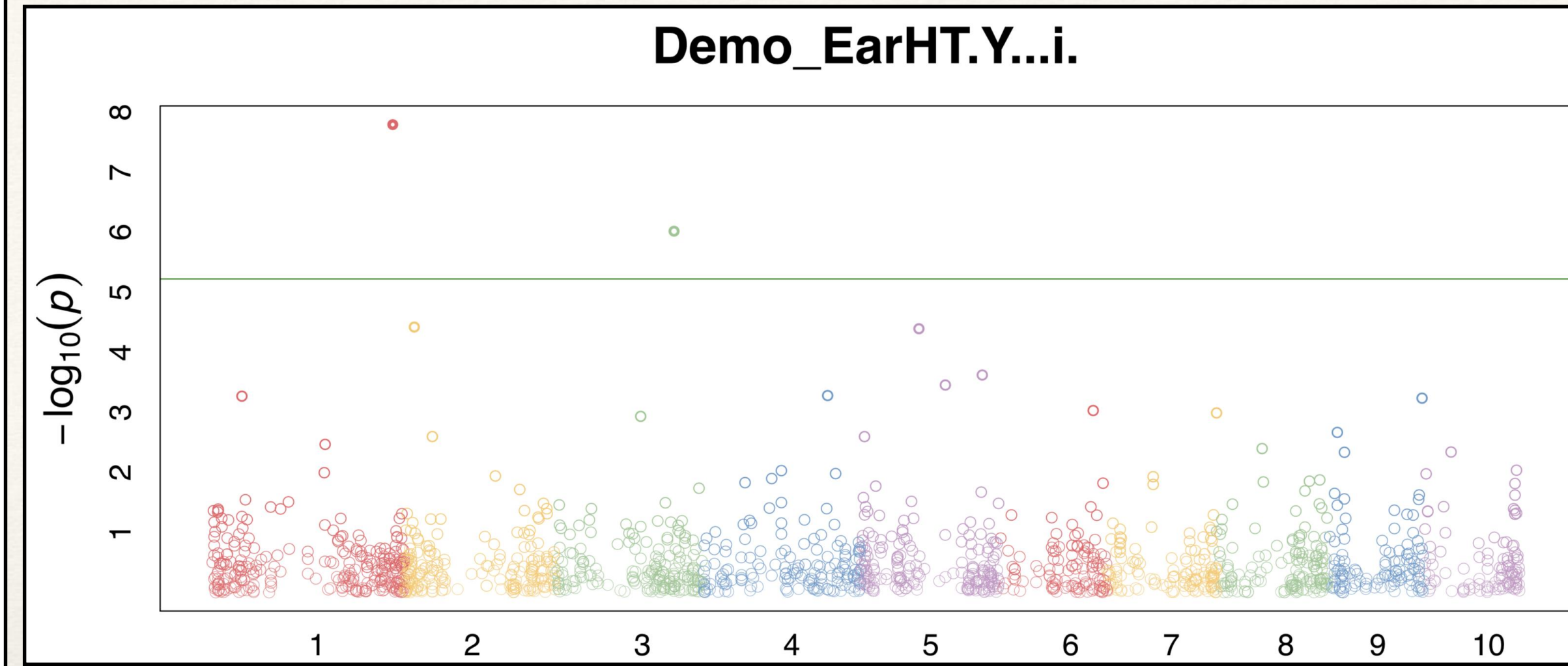
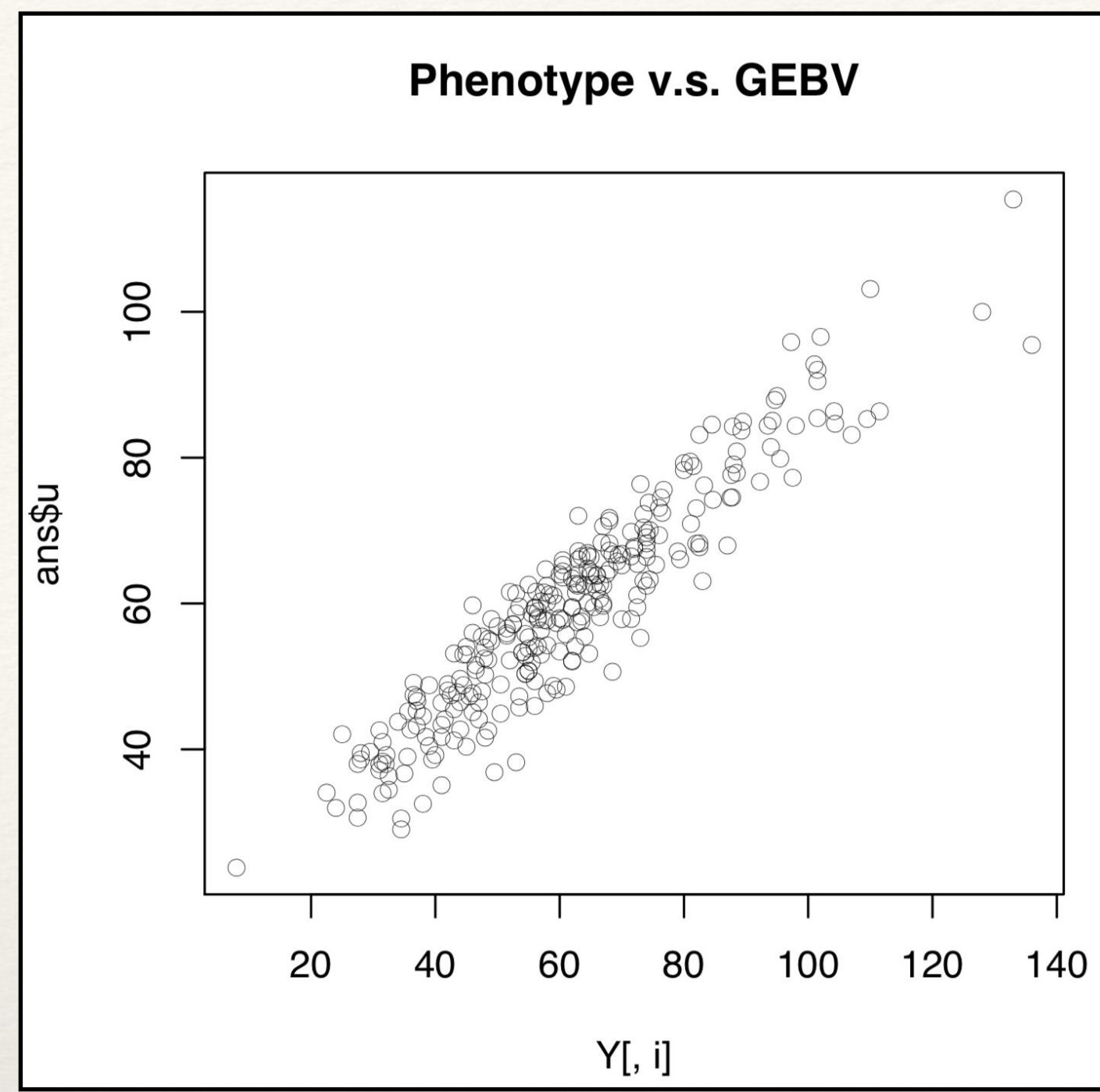
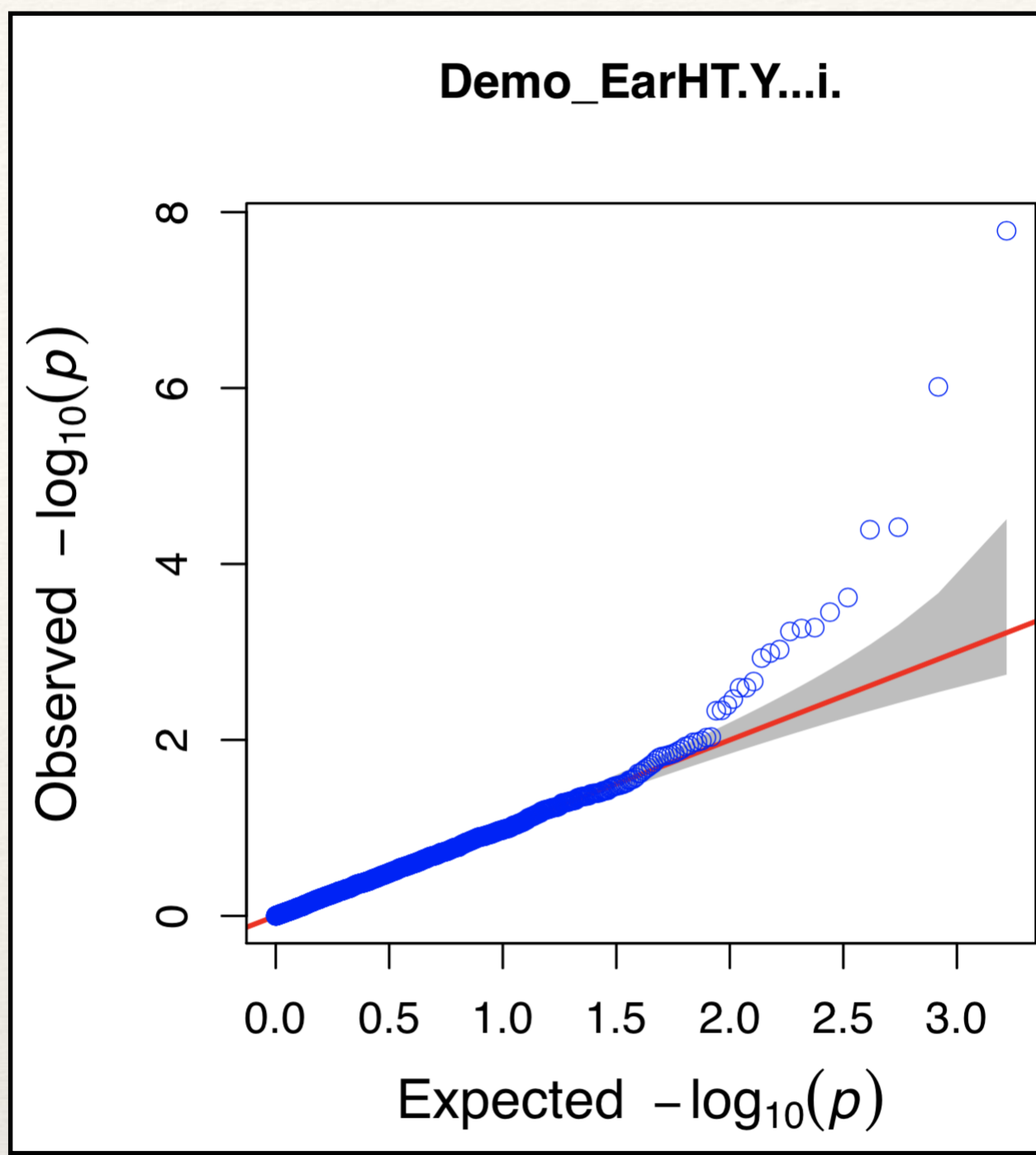
```
phenotype = phtnoeyp[,c(2, 3)]
MS = is.na(genotype) %>%
  apply(2, function(x) sum(x)/length(x))
MAF = apply(genotype, 2, mean) %>%
  as.matrix() %>%
  apply(1, function(x) min(1 - x/2, x/2))
```

# iPat Got You Back!





iPat



### Breeding value

taxa	u	u.SE	
33-16	53.1213677206192		8.05706131458929
38-11	76.7092223273507		8.11027035267517
4226	62.6168677588951		7.92749161327708
4722	70.9413420726807		7.37336167756384
A188	30.6209637249398		7.97636416276851
A214N	63.8441664303389		7.40571054863867
A239	52.4632605177171		8.3795399083574
A272	45.2143214735658		8.38647699841112
A441-5	59.5710810597331		8.45442977368395
A554	41.7105684119449		7.7820324109983
A556	39.477390280773	8.11221141057064	
A6	85.3066022321695		8.17941245867077
A619	42.6893547823443		6.96544922422308
A632	53.4090273261965		6.43925558329717
A634	53.3302784522467		7.06167844955625

### P value of association test

SNP	Chromosom	Position	P.value	MAF
PZB00859.1	1	157104	0.315789212723918	0.240213523131673
PZA01271.1	1	1947984	0.0436995168679747	0.48932384341637
PZA03613.2	1	2914066	0.16552837227778	0.290035587188612
PZA03613.1	1	2914171	0.598855533795033	0.250889679715303
PZA03614.2	1	2915078	0.0678701534537391	0.469750889679715
PZA03614.1	1	2915242	0.240188905155287	0.469750889679715
PZA00258.3	1	2973508	0.118528522798418	0.284697508896797
PZA02129.1	1	3706018	0.856262621133709	0.480427046263345
PZA00393.1	1	4175293	0.80813438581583	0.247330960854093
PZA02869.4	1	4429927	0.0974914108116822	0.215302491103203
PZA02869.2	1	4430055	0.559008736584389	0.231316725978648
PZB00919.1	1	5353319	0.998059668362255	0.227758007117438
PZA03093.10	1	8075572	0.485393283284251	0.491103202846975
PZA00181.2	1	8366411	0.262385243401303	0.302491103202847
PZA00528.1	1	8367944	0.74639129220144	0.250889679715303
PZA00175.2	1	8510027	0.0459860426759082	0.414590747330961
PZA00447.6	1	9023947	0.223056121959639	0.494661921708185

# Availability

The screenshot shows the article page for 'iPat: intelligent prediction and association tool for genomic research' on the Bioinformatics journal website. The page includes a navigation bar with 'Bioinformatics' and 'iSCBS INTERNATIONAL SOCIETY FOR COMPUTATIONAL BIOLOGY'. The article title is 'iPat: intelligent prediction and association tool for genomic research' by Chunpeng James Chen and Zhiwu Zhang. It is a corrected proof published on 11 January 2018. The abstract states: 'The ultimate goal of genomic research is to effectively predict phenotypes from genotypes so that medical management can improve human health and molecular breeding can increase agricultural production. Genomic prediction or selection (GS) plays a complementary role to genome-wide association studies'. The page also features a 'View Metrics' button showing 3 views and an 'Email alerts' section with options for 'New issue alert', 'Advance article alerts', and 'Article activity alert'. A sidebar on the left lists 'Article Contents' with sections: 1 Introduction, 2 GWAS-assisted genomic prediction, 3 GUI, data and third party CLI packages, 4 Implementation, 5 Conclusions, Acknowledgement, and References. There are also buttons for 'Views', 'PDF', 'Cite', 'Permissions', and 'Share'.

Published on Bioinformatics

The Zhiwu Zhang Laboratory logo is 'Zhiwu Zhang Laboratory for Statistical Genomics'. The navigation menu includes: Home, People, Publication, Research, Teaching, Software, Outreach, and Jobs.

The screenshot shows the iPat software interface with a workflow diagram. It includes a sidebar with tools like GAPIT, FarmCPU, PLINK, rBLUP, and BGLR. The main area shows a workflow from 'data.txt' to 'covariates.txt' to 'Project\_1' to 'Modeling' to 'Drop to Deploy' to 'GART iPatLUP' and 'BGLR iPatLUP'. There are also plots for 'frequency' and 'GBEY'.

**EASY WAY TO GWAS AND GS**

Logos for Apple, Windows, and Ubuntu operating systems are shown at the bottom of the advertisement.

<http://zzlab.net/iPat/>

# Acknowledgement : Z.Z. Lab

PI:  
Dr. Zhiwu Zhang

It's me  
James Chen

