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

Chunpeng (James) Chen
Ph.D. Student
Department of Crop and Soil Sciences
Washington State University

World Congress On Genetics Applied to Livestock Production (WCGALP, 11th)
The Challenges of Biologists

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

Programming skills
Big data
Formatting
Modeling
...
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

Bio-sample → Sequencing → Functional Genomics Studies
The Focus of iPAt

Finding Associated Genes

Genome-Wide Association Studies

Genomic Selection

Predict Phenotype

Bulk Segregation Analysis

Finding Associated Genes by Pool Sequencing
Embedded Tools in One Place!

- GAPIT
- FarmCPU
- PLINK
- Genome-Wide Association Studies
- Genomic Selection
- BGLR
- rrBLUP
- Bulk Segregation Analysis
- SNP-index
- G' Statistics
- Euclidean Distance
Unfortunately… not all formats work on them

<table>
<thead>
<tr>
<th>Tool</th>
<th>Supported Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPIT</td>
<td>Numeric, Hapmap</td>
</tr>
<tr>
<td>FarmCPU</td>
<td>Numeric</td>
</tr>
<tr>
<td>BGLR</td>
<td>Numeric</td>
</tr>
<tr>
<td>rrBLUP</td>
<td>Numeric</td>
</tr>
<tr>
<td>PLINK</td>
<td>PLINK, Binary</td>
</tr>
</tbody>
</table>
Unfortunately… not all formats work on them

<table>
<thead>
<tr>
<th>Tool</th>
<th>Supported Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPIT</td>
<td>Numeric, Hapmap</td>
</tr>
<tr>
<td>FarmCPU</td>
<td>Numeric</td>
</tr>
<tr>
<td>BGLR</td>
<td>Numeric</td>
</tr>
<tr>
<td>rrBLUP</td>
<td>Numeric</td>
</tr>
<tr>
<td>PLINK</td>
<td>PLINK, Binary</td>
</tr>
</tbody>
</table>

VCF
Inspect the format

**Hapmap**
- rs alleles
- chrom pos strand assembly center protLSID assayLSID panel Qcode

**PLINK**
- abc1

**VCF**
- #CHROM POS ID REF ALT QUAL FILTER INFO FORMAT
Format Free

No need to spend effort on format conversion
Let’s work with iPat step by step!

Import Data → Quality Control → Define a Method → Inspect Results

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

Double-click
Build a Module
Define Your Analysis
Missing Rate? Minor Allele Frequency?

```r
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))
```
¡Pat Got You Back!

FarmCPU
Selected
(Tap for details)
Published on Bioinformatics

http://zzlab.net/iPat/
Acknowledgement: Z.Z. Lab

PI: Dr. Zhiwu Zhang

It's me James Chen