Association whap

Benjamin Neale
19th International Workshop on Twin Methodology
Practical--Whap

- Whap implements both Case/control and Family-based designs
- We will be looking at an example TDT from whap, as we have been dealing with families all week long
- Whap is found at http://pngu.mgh.harvard.edu/~purcell/whap/
- Copy all files from ben/association/practical
Reminder of TDT

Under the null, we expect equal chance of either allele transmitting from a heterozygote parent to the offspring.

Though Whap only does biallelic markers (SNPs).
Testing in Whap

- Whap demands you specify the alternative hypothesis.
- The null hypothesis can be supplied, but the default is that no marker has any effect on the trait.
- When you specify an alternative hypothesis you are testing that marker/haplotype is associated with disease.
Significance in Whap

• Whap uses permutation analysis to assess significance
• Permutation analysis assumes that all data come from the null distribution
• Thus, under TDT, the null distribution is that it is equally likely allele 1 or allele 2 is transmitted.
• We ‘mix up’ the transmitted and non-transmitted allele within family
Visualization of permutation

Original data:
2: allele 1 trans
0: allele 2 trans

Permutation 1:
1: allele 1 trans
1: allele 2 trans
Visualization of permutation

Permutation 2:
0: allele 1 trans
2: allele 2 trans

Permutation 3:
2: allele 1 trans
0: allele 2 trans
Whap input

• 3 files
  – Ped file
  – Map file
  – Dat file
Ped files

• Pedigree file looks like:

<table>
<thead>
<tr>
<th>Family</th>
<th>ID</th>
<th>Pa</th>
<th>Ma</th>
<th>sex</th>
<th>Aff</th>
<th>Al1</th>
<th>Al2</th>
</tr>
</thead>
<tbody>
<tr>
<td>PED054</td>
<td>430</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>PED054</td>
<td>412</td>
<td>430</td>
<td>431</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>PED054</td>
<td>431</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

The DATA file
Must have .ped extension
## Map Files

<table>
<thead>
<tr>
<th>Chr</th>
<th>Marker name</th>
<th>cM</th>
<th>bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>IGR1118a_1</td>
<td>100.00</td>
<td>274044</td>
</tr>
<tr>
<td>5</td>
<td>IGR1119a_1</td>
<td>100.00</td>
<td>274541</td>
</tr>
<tr>
<td>5</td>
<td>IGR1143a_1</td>
<td>100.00</td>
<td>286593</td>
</tr>
<tr>
<td>5</td>
<td>IGR1144a_1</td>
<td>100.00</td>
<td>287261</td>
</tr>
<tr>
<td>5</td>
<td>IGR1169a_2</td>
<td>100.00</td>
<td>299755</td>
</tr>
</tbody>
</table>

Tells Whap where the markers are on the genome
Must have .map extension
## Dat file

<table>
<thead>
<tr>
<th>Type</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>disease</td>
</tr>
<tr>
<td>M</td>
<td>IGR1118a_1</td>
</tr>
<tr>
<td>M</td>
<td>IGR1119a_1</td>
</tr>
<tr>
<td>M</td>
<td>IGR1143a_1</td>
</tr>
<tr>
<td>M</td>
<td>IGR1144a_1</td>
</tr>
</tbody>
</table>

Tells Whap what is in each column
Must have .dat extension

- A = affection
- B = disease
- T = trait
- C = covariate
- M = marker
- S = marker skip
- X = trait skip
Running Whap!

• Whap is a command line program which means we have to use the command prompt in Windows.
• To run command prompt click start, then run…
• In the run box type cmd, then hit return
• Change to your network drive by typing ‘H:’ (not necessary if you are in H:)
• Change directory with cd ‘directory name’
• Dir brings up a list of files and directories.
Command prompt

Click start, then run
Running command prompt

Type cmd then hit return
You should have this up now

This is a command prompt
Changing to H:

H: then enter
may not be necessary if you are in H: already
Dir shows you what is in your directory

Dir shows what directories and files you have
My directory is...

My directory for the practical is:
H:\2006\Association\practical so I have typed:
cd 2006\Association\practical

Now I am in the directory where Whap and my datafiles are

You need to navigate to your directory, where you saved the files and the program
Now what?

• Hopefully you have successfully navigated into the directory (you can check with dir!)
• We can now start to run Whap:
• Command is:
• `whap --file chr5q31 --alt 1 --window --cond --prev 0.01 --model w --wperm 500`
What are all these bits?

- **Whap**: program name
- **--file chr5q3**: whap requires your map, dat, and ped files to have the same name. File supplies this
- **--alt 1**: this is the alternative hypothesis
- **--window** means that whap will look at a sliding window of SNPs
  - Whap begins with the alt, then slides across all remaining markers if window is active
What are all these ‘and pieces’?

- --cond: conditional analysis—necessary for family data
- --prev 0.01: prevalence of the trait
- --model w: within model of association
- --wperm 500: number of within permutations
A picture of the command to be run

```
H:\>cd 2006\Association\practical
H:\2006\Association\practical>whap --file chr5q31 --alt 1 --window --cond --prev 0.01 --model w --wperm 500
```

- **Program**: Whap
- **File name**: chr5q31
- **Alternative hypothesis**: --alt 1
- **Sliding window analysis**: --window
- **Conditional analysis**: --cond
- **Prevalence**: --prev 0.01
- **Within analysis**: --model w
- **500 within permutations**: --wperm 500
P max is a comparison of the best statistic you observed in your data against the ‘best’ statistic in each permutation.
P sum is a comparison of the sum total of the statistics you observed in your data against the sum total in each permutation.
Final notes

• P-sum and P-max are corrected for multiple testing, as long as you include all markers tested
• We can use Whap to do haplotype tests as well as single markers
• Much more on association at the advanced workshop