ACCURACY AND ANALYSIS OF GWAS (GENOME WIDE ASSOCIATION STUDIES)

Karen Chau

CSC 334

Fall 2015

GWAS BACKGROUND

- GWAS identifies genetic variations associated with diseases
  - Benefits
    - Detection, treatment, prevention
    - Personalized medicine
  - Ethics
    - Prenatal testing
  - Refinement
    - Increase efficiency and decrease cost

1 in 11

Americans have

DIABETES

*American Diabetes Association (2012)

1

million

Americans have PARKINSON'S DISEASE

*Parkinson's Disease Foundation (2015)

1

American dies from **HEART DISEASE** every

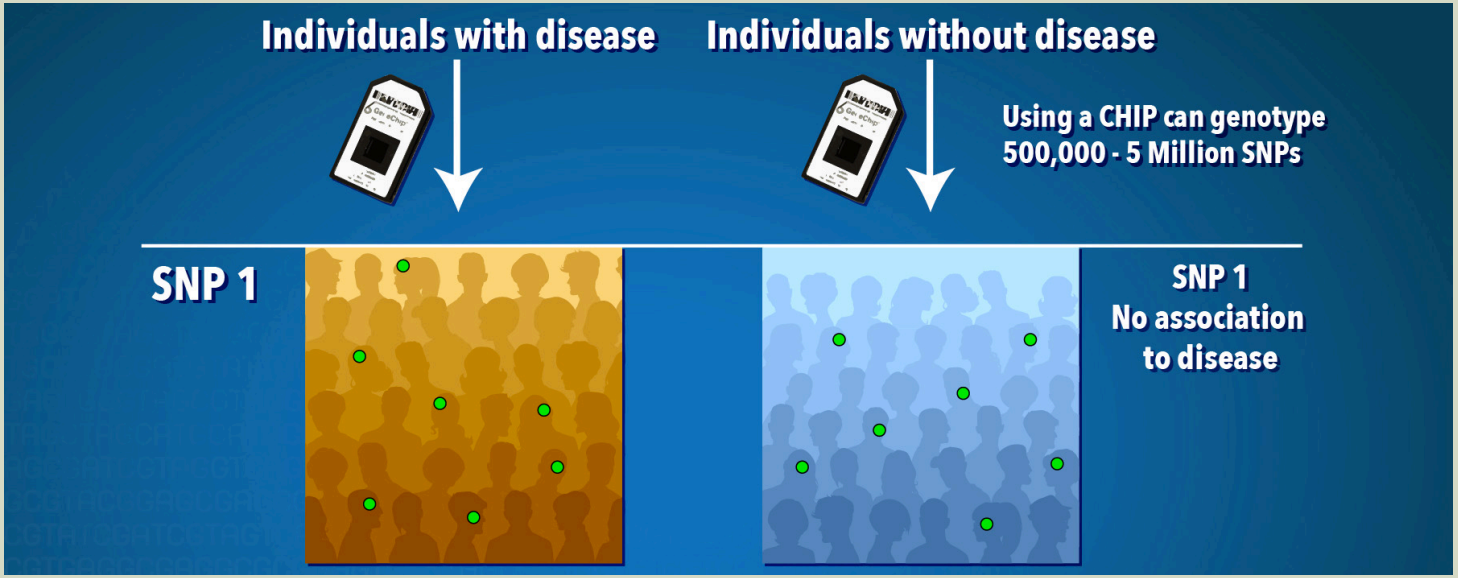
90

seconds

*American Heart Association (2015)

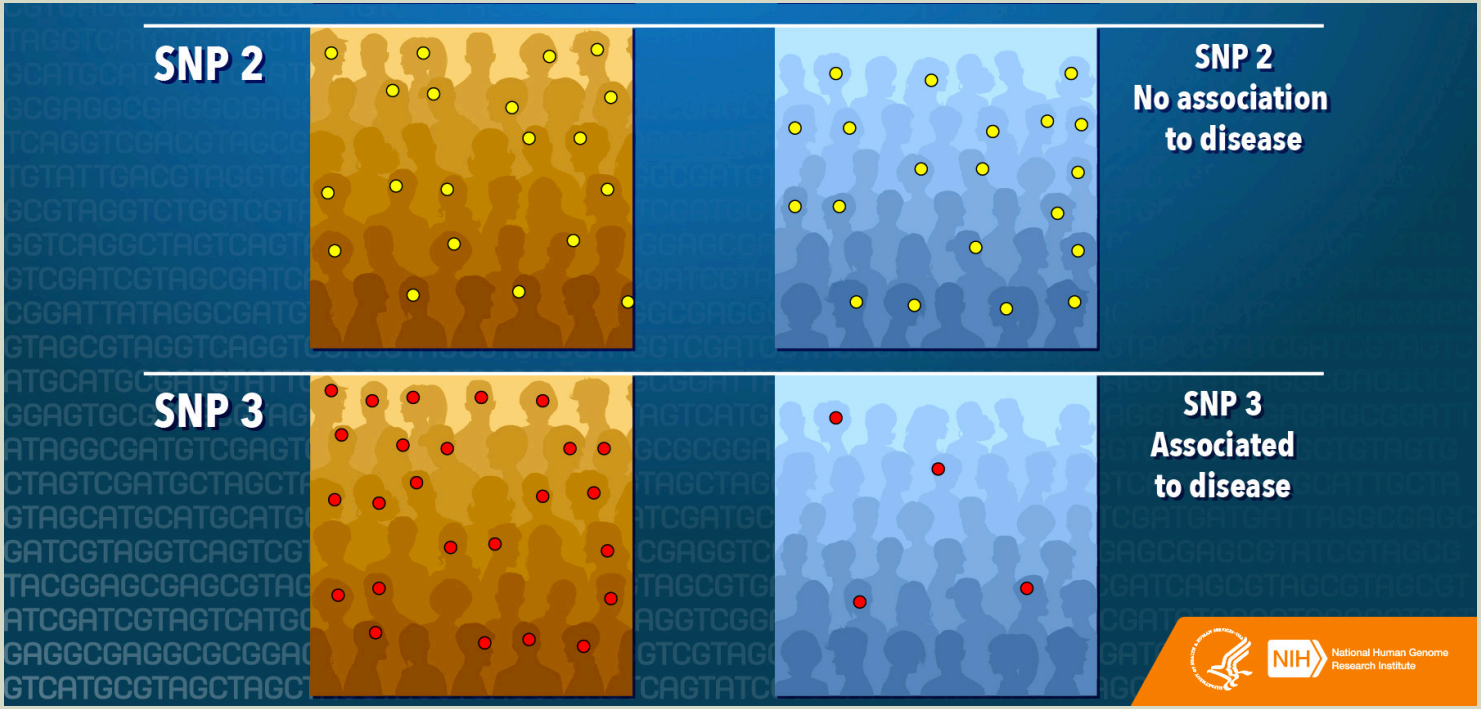
GWAS METHOD

- Compares SNPs between those with and without the disease



GWAS METHOD

- Compares SNPs between those with and without the disease

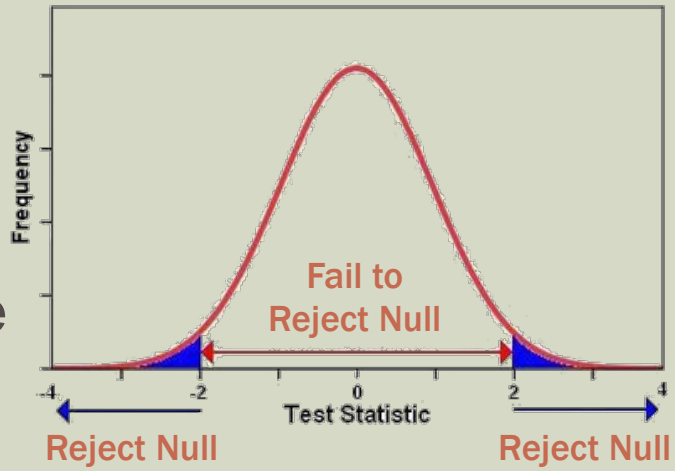


GWAS METHOD

- Compares SNPs between those with and without the disease
- Each disease
  - Many experiments
  - Each experiment
    - Series of associated SNP
      - Each SNP has a p-value

STATISTICS BACKGROUND

- Hypothesis testing
  - Null hypothesis
    - Observations are due to chance
  - Alternative hypothesis
    - Observations are influenced
  - Test statistic
    - Standardized value calculated from sample
  - P-value
    - Pr(observing the test statistic, assuming H_0 is true)



WHAT DOES THE DATA MEAN?

- Accuracy
  1) How accurate were the studies compared to each other?
- Analysis
  1) Given a SNP:
     - What disease(s) are associated with it?
  2) Given a disease:
     - Which SNP(s) are associated with it?
     - Which chromosome(s) are more associated with it?

DETERMINING ACCURACY OF GWAS DATA

- Consolidate SNPs from all studies and filter out >0.05 p-values

[First experiment]

DETERMINING ACCURACY OF GWAS DATA

- Consolidate SNPs from all studies and filter out >0.05 p-values

Data set HGVRS3891

Top **35** associations for Coronary Artery Disease (HGVPM3648) from data set HGVRS3891(35 in database)

Export as Microsoft Excel Go

Experiment: Association analysis experiment for coronary artery disease (HGVE3733)
Data set: Coronary artery disease analysis (HGVRS3891)

Rank	Identifier	Accession	Region	p-value	-log p-value	Effect size	Risk allele	Related data	Links
1	HGVM1880766	rs1558902	chr16:53803574..53803574	0.04317	1.365	Not supplied	Not supplied	Details	Not supplied
2	HGVM5089460	rs12444979	chr16:19933600..19933600	0.06869	1.163	Not supplied	Not supplied	Details	Not supplied
3	HGVM3473352	rs10767664	chr11:27725986..27725986	0.1325	0.878	Not supplied	Not supplied	Details	Not supplied
4	HGVM1405127	rs2531995	chr16:4013467..4013467	0.1373	0.862	Not supplied	Not supplied	Details	Not supplied
5	HGVM118737	rs887912	chr2:59302877..59302877	0.1735	0.761	Not supplied	Not supplied	Details	Not supplied
6	HGVM2862188	rs4929949	chr11:8604593..8604593	0.1765	0.753	Not supplied	Not supplied	Details	Not supplied
7	HGVM1872195	rs1514175	chr1:74991644..74991644	0.1856	0.731	Not supplied	Not supplied	Details	Not supplied
8	HGVM759097	rs543874	chr1:177889480..177889480	0.2111	0.676	Not supplied	Not supplied	Details	Not supplied
9	HGVM3675423	rs10968576	chr9:28414339..28414339	0.2118	0.674	Not supplied	Not supplied	Details	Not supplied
10	HGVM498801	rs571312	chr18:57839769..57839769	0.2363	0.627	Not supplied	Not supplied	Details	Not supplied

DETERMINING ACCURACY OF GWAS DATA

- Consolidate SNPs from all studies and filter out >0.05 p-values

[Second experiment]

DETERMINING ACCURACY OF GWAS DATA

- Consolidate SNPs from all studies and filter out >0.05 p-values

Data set HGVRS3892

Top **35** associations for Coronary Artery Disease (HGVPM3648) from data set HGVRS3892(35 in database)

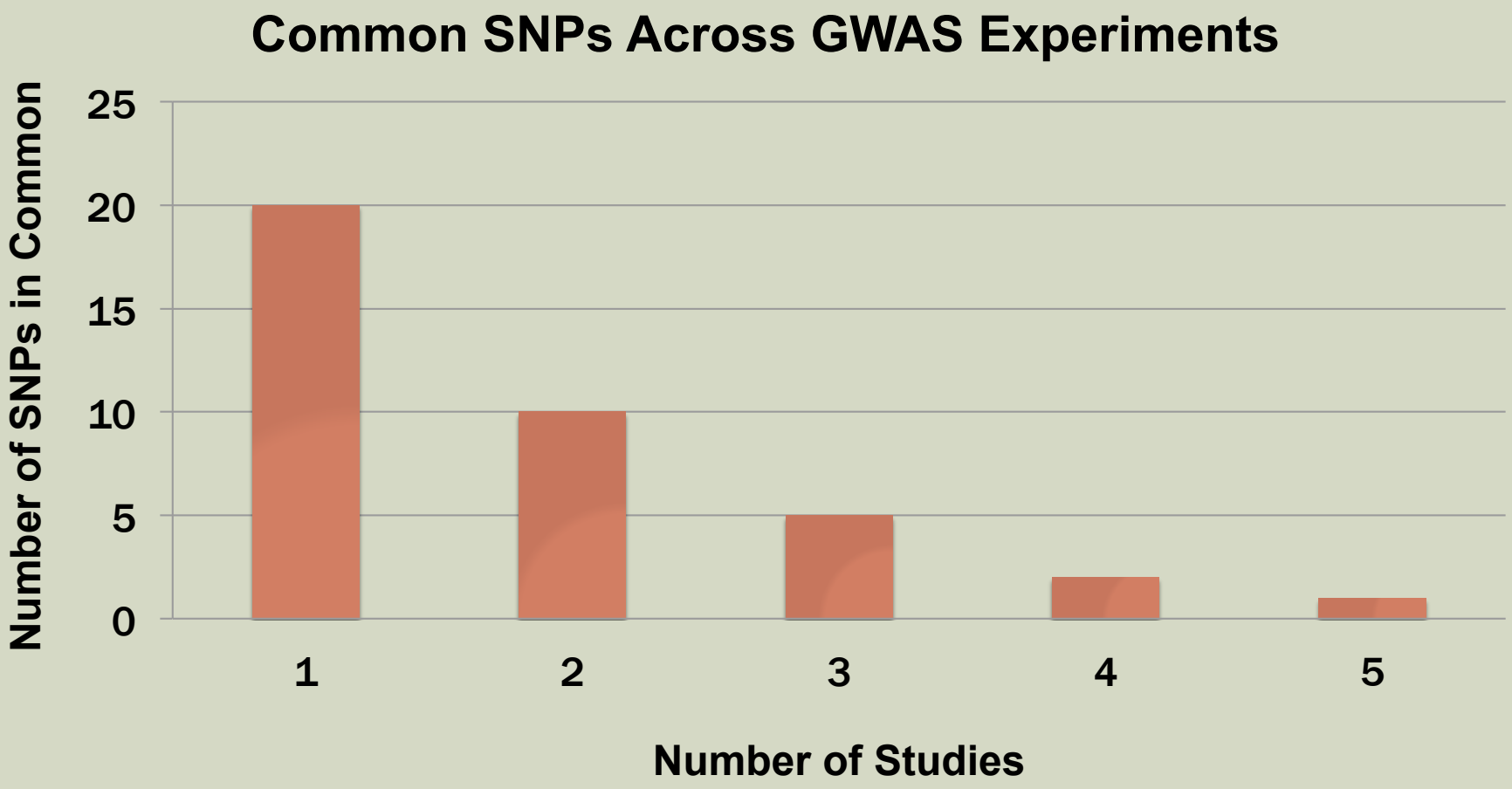
Export as --choose a format -- Go

Experiment: Association analysis experiment for BMI (HGVE3734)
Data set: BMI analysis (HGVRS3892)

Rank	Identifier	Accession	Region	p-value	-log p-value	Effect size	Risk allele	Related data	Links
1	HGVM498801	rs571312	chr18:57839769..57839769	0.0002495	3.603	Not supplied	Not supplied	Details	Not supplied
2	HGVM1405127	rs2531995	chr16:4013467..4013467	0.002647	2.577	Not supplied	Not supplied	Details	Not supplied
3	HGVM1880766	rs1558902	chr16:53803574..53803574	0.002729	2.564	Not supplied	Not supplied	Details	Not supplied
4	HGVM1682064	rs2867125	chr2:622827..622827	0.00452	2.345	Not supplied	Not supplied	Details	Not supplied
5	HGVM673226	rs987237	chr6:50803050..50803050	0.005968	2.224	Not supplied	Not supplied	Details	Not supplied
10	HGVM498801	rs571312	chr18:57839769..57839769	0.2363	0.627	Not supplied	Not supplied	Details	Not supplied

- Compute frequency of same SNPs

EXPECTED ACCURACY OF GWAS DATA



ANALYZING GWAS DATA

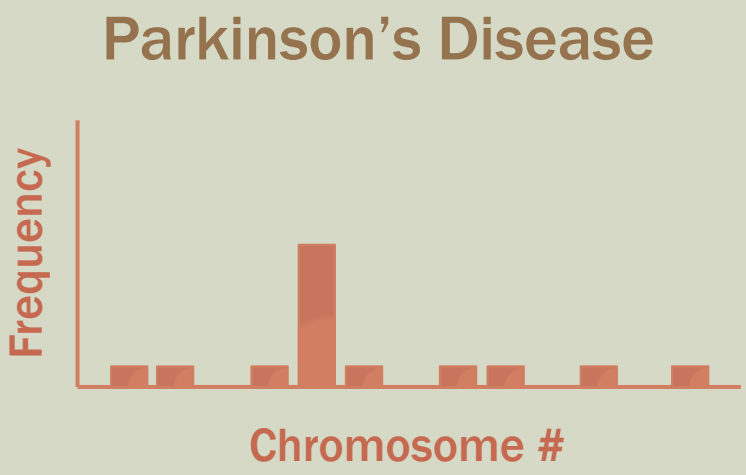
- Given a SNP, what disease(s) are associated with it?
- Hash Table
  - <KEY>,<VALUE>
    - Example: dictionary with <word>, <definition>
    - O(1) retrieval and insertion
  - Question
    - <SNP#>,<DISEASE(S)>

ANALYZING GWAS DATA

- Given a disease, which SNP(s) are associated with it?
  - Hash Table
    - <DISEASE>,<SNP#(s)>
- Given a disease, which chromosome(s) are more associated with it?
  - Hash Table
    - <DISEASE>,<CHROMOSOME#(s)>

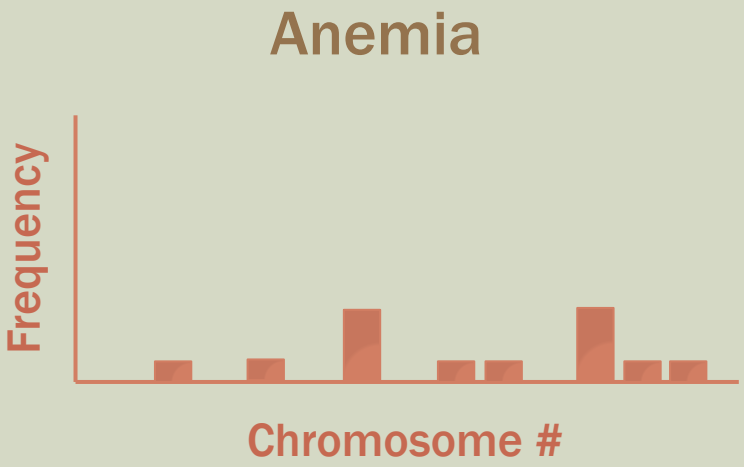
EXTENSIONS

- Use R to plot the results visually
  - For particular disease
    - Look at the frequency of SNPs on each chromosome



EXTENSIONS

- Use R to plot the results visually
  - For particular disease
    - Look at the frequency of SNPs on each chromosome



EXTENSIONS

- Explore more diseases
  - Anorexia
  - Sickle Cell Anemia
  - Parkinson's
  - Schizophrenia
  - and more!

REFERENCES

1. Hodgson, David. Emotions in Parkinson's disease. *Medical Xpress*. Dec 2013. <http://medicalxpress.com/news/2013-12-emotions-parkinson-disease.html>.
2. Levy, Sandra. Type 1 Diabetes May Cut Life Span, but Intensive Treatment Can Help Close the Gap. *Healthline*. Jan 2015. <http://www.healthline.com/health-news/type-1-diabetes-may-cut-life-span-but-intensive-treatment-can-help-010615#1>.
3. Norrgard, Karen. Genetic Variation and Disease: GWAS. *Nature Education*. 1(1):87. 2008. <http://www.nature.com/scitable/topicpage/genetic-variation-and-disease-gwas-682#>.
4. 1000 Genomes. *European Bioinformatics Institute*. 2015. <http://browser.1000genomes.org/index.html>.
5. American Diabetes Association. 2012. <http://www.diabetes.org/diabetes-basics/statistics/>.
6. American Heart Association. 2015. <https://www.heart.org/idc/groups/ahamah-public/@wcm/@sop/@smd/documents/downloadable/ucm_470704.pdf>.
7. Genome-Wide Association Studies. *National Human Genome Research Institute.* Aug 2015. <http://www.genome.gov/20019523>.
8. GWAS Central. 2015. <http://www.gwascentral.org/index>.
9. Heart Attack - Correct Treatment. *Just Do It Guides*. <https://www.youtube.com/watch?v=Vts09VkK6Q8>.
10. Parkinson’s Disease Foundation. 2015. <http://www.pdf.org/en/parkinson_statistics>.
11. The IPCC and The Null Hypothesis. *Inform the Pundits*. Sept 2013. <https://informthepundits.wordpress.com/2013/09/20/the-ipcc-and-the-null-hypothesis/>.

THANK YOU!

Questions/
Comments?