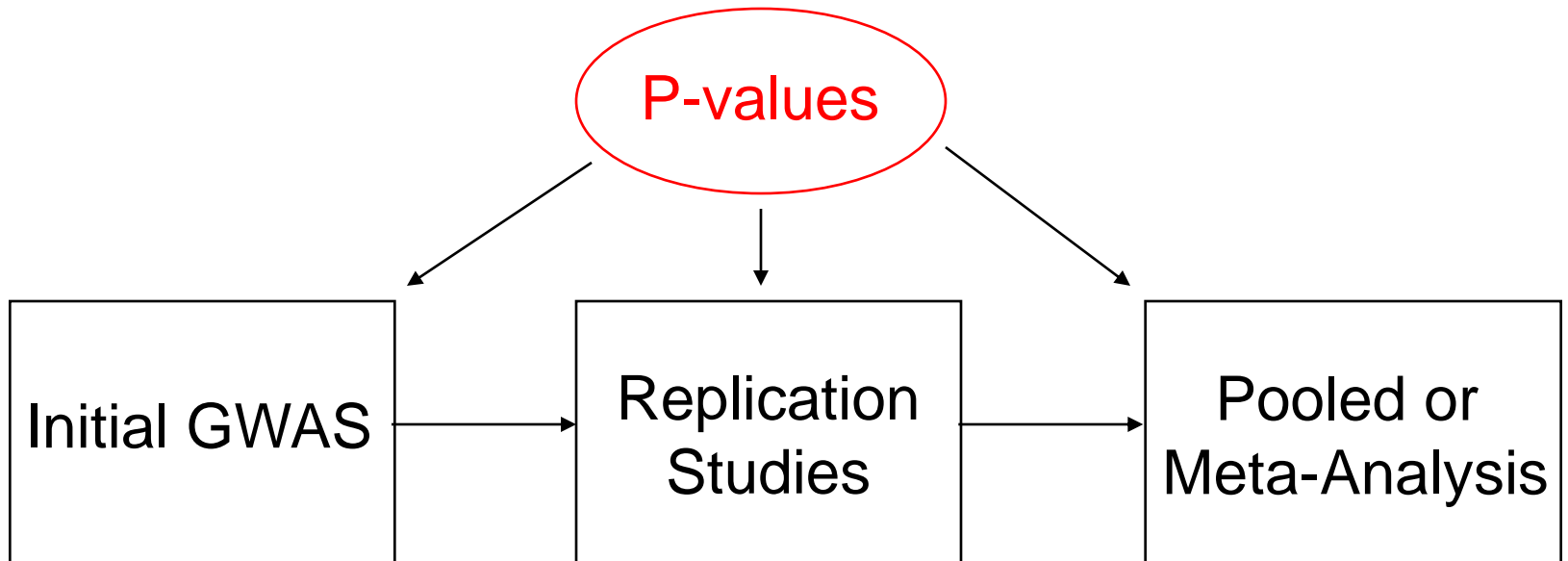# Genome Wide Meta-Analysis: Promise and Pitfalls

John Witte: jwitte@ucsf.edu

GWAs Findings

P-values

Initial GWAS → Replication Studies → Pooled or Meta-Analysis

For genome profiles need relative risk estimates. Where should these come from?

Goals of Meta-Analysis

1. Combine findings across studies to obtain 'best' estimates of association (p-value / RR).

2. Determine if and why differences exist across studies.

- Extensive work on methods and controversies in statistics and epidemiology literature.

Meta-Analysis in Assoc. Studies

- Straightforward in early days (1 SNP).
- HuGENet portal: www.cdc.gov/genomics/hugenet

- More complicated in GWAs era .

Meta-Analysis of GWAs

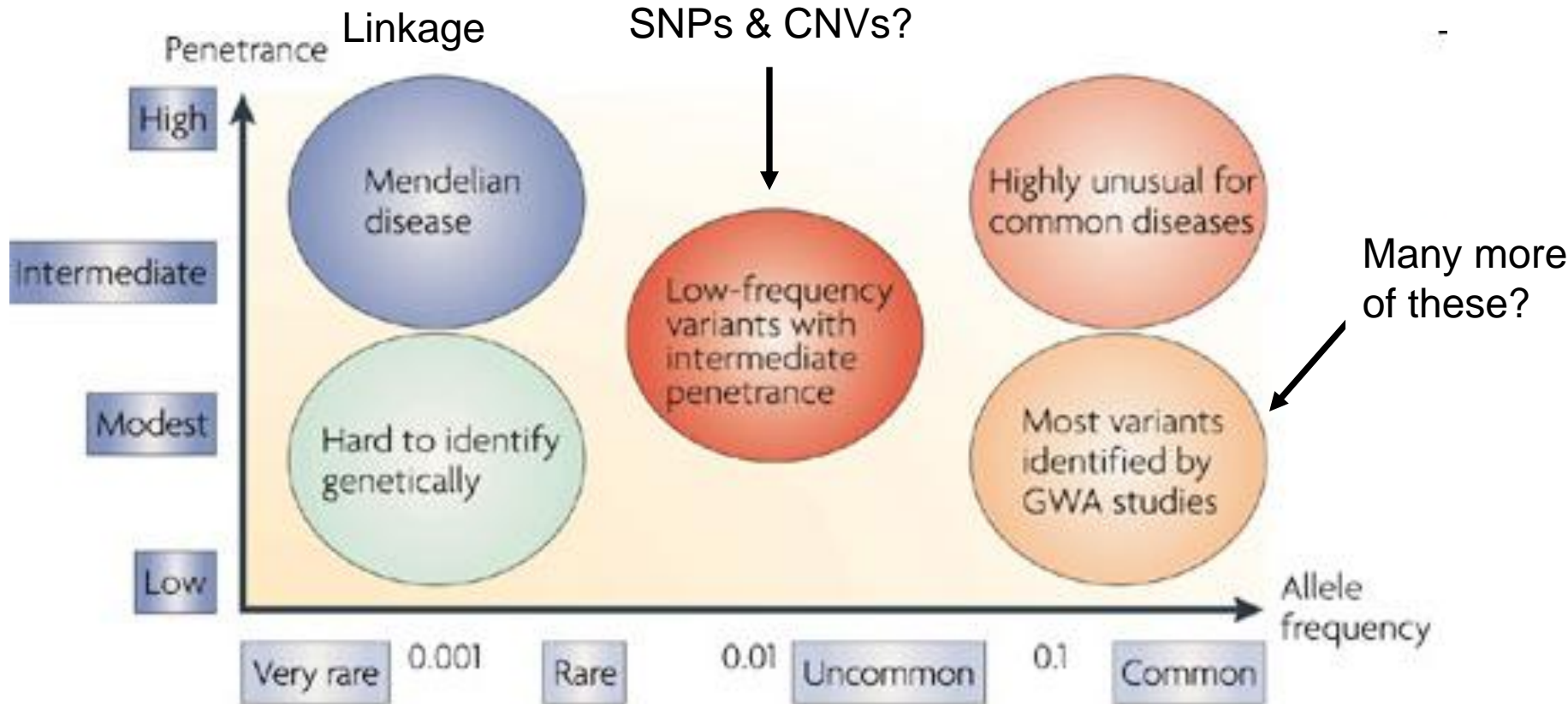| Initial GWAS	→	Replication Studies	→	Pooled or Meta Analysis

Two situations:
1. GWAs data alone (still in 'discovery phase');
2. GWAs data and focused replications.

In 1) probably know about all GWAs.
In 2) need to search for all studies.

Publication bias

Meta-Analyses & Heritability

Nature Reviews | Genetics

McCarthy et al., 2008

Need for GWAS Meta-Analyses

Locus			A Freq		Association		Nearby Genes / Fcn
Chr Reg	**SNP**		**Cntrl**	**Case**	**OR**	**p value**	
2p15	rs721048	G/A	0.19	0.21	1.15	7.7×10^{-9}	*EHBP1:* endocytic trafficking
3p12	rs2660753	C/T	0.10	0.12	1.30	2.7×10^{-8}	Intergenic
6q25	rs9364554	C/T	0.29	0.33	1.21	5.5×10^{-10}	*SLC22A3:* drugs and toxins.
7q21	rs6465657	T/C	0.46	0.50	1.19	1.1×10^{-9}	*LMTK2:* endosomal trafficking
8q24 (2)	rs16901979	C/A	0.04	0.06	1.52	1.1×10^{-12}	Intergenic
8q24 (3)	rs6983267	T/G	0.50	0.56	1.25	9.4×10^{-13}	Intergenic
8q24 (1)	rs1447295	C/A	0.10	0.14	1.42	6.4×10^{-18}	Intergenic
10q11	**rs10993994**	**C/T**	**0.38**	**0.46**	**1.38**	**8.7×10^{-29}**	***MSMB:* suppressor prop.**
10q26	rs4962416	T/C	0.27	0.32	1.18	2.7×10^{-8}	*CTBP2:* antiapoptotic activity
11q13	rs7931342	T/G	0.51	0.56	1.21	1.7×10^{-12}	Intergenic
17q12	rs4430796	G/A	0.49	0.55	1.22	1.4×10^{-11}	*HNF1B:* suppressor properties
17q24	rs1859962	T/G	0.46	0.51	1.20	2.5×10^{-10}	Intergenic
19q13	rs2735839	A/G	0.83	0.87	1.37	1.5×10^{-18}	*KLK2/KLK3:* PSA
Xp11	rs5945619	T/C	0.36	0.41	1.29	1.5×10^{-9}	*NUDT10, NUDT11:* apoptosis

24,223 smallest P-value!

Witte, Nat Rev Genet, to appear

Pooled Analysis of GWAs

Initial GWAS	→	Replication Studies	→	**Pooled Analysis**

Optimal study:
   Pooled analysis of individual-level data.
   Can look at independent, interacting
   and multi-phenotypic effects.

Reality: only get results

Meta-Analysis of GWAs

```
┌─────────────┐     ┌─────────────┐     ┌─────────────┐
│             │     │ Replication │     │             │
│ Initial GWAS│ ──► │   Studies   │ ──► │Meta-Analysis│
│             │     │             │     │             │
└─────────────┘     └─────────────┘     └─────────────┘
```

Key aspects:
- Imputing across different platforms.
- How to combine results.
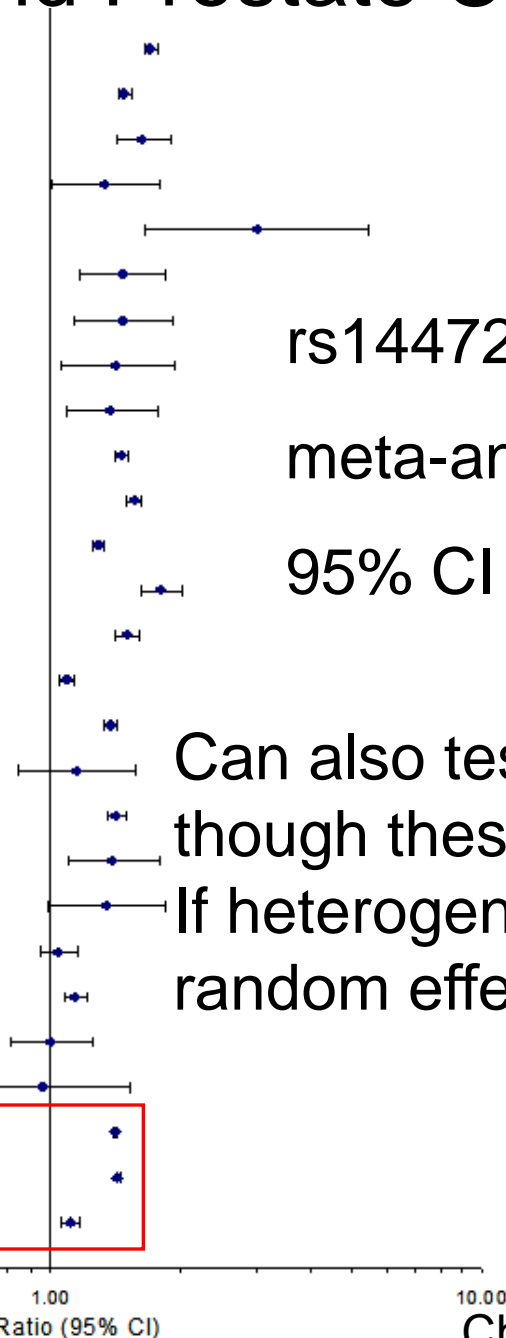
Imputation for Meta-Analysis

- **Issue**: Different platforms across studies.
  - Need to impute up to hapmap (and eventually 1,000 genomes).
  - Population stratification adjustment and analyses need to distinguish between observed and imputed data.

  - Logistical: Make sure SNPs correctly oriented, etc. (de Bakker et al., HMG 2008).

Combining GWAs Results

- Standard approaches:
  - Z scores weighted by sample size
  - Inverse variance weighted odds ratios

- Assumes no variation between studies (fixed effects model).

- Recall that one goal is to determine if and why differences exist across studies!

8q24 SNP and Prostate Cancer

Admundadottir*	European, Iceland
Admundadottir*	European, Sweden
Admundadottir*	European American, Chicago
Freedman	European American, Hawaii-California
Freedman	Native Hawiian, Hawaii-California
Freedman	Japanese American, Hawaii-California
Freedman	Latino, Hawaii-California
Gudmundson	European, Spain
Gudmundson	European, Netherlands
Yaeger*	European American, PLCO
Yaeger*	European American, ACS
Yaeger*	European, ATBC
Yaeger*	European, FPCC
Yaeger*	European American, HPFS
Schumacher*	European, EPIC
Schumacher*	European American, PHS
Wang*†	European American, Mayo
Severi*	European descent, Australia
Suuriniemi	European American, Washington
Cheng	European American, Ohio
Freedman	African American, Hawaii-California-Michigan
Admundadottir*	African American, Chicago
Gudmundson	African American, Baltimore
Cheng	African American, Ohio
Meta-analysis	All
Meta-analysis	European ancestry
Meta-analysis	African ancestry

Odds Ratio (95% CI)

rs1447295 A vs. C allele

meta-analysis OR =1.42

95% CI =1.40–1.44

Can also test for heterogeneity, though these have poor power. If heterogeneous, need to use random effects model…

Cheng…Witte, EJHG 2008

Which is Used?

- To date GWAs meta-analyses focus on fixed effects models.

- "GWA studies likely to suffer from winners curse…causing variability in effect estimates by chance. Therefore, a random-effects model may well be too conservative compared with a fixed effects model"

  De Bakker et al., HMG 2008, p. R126

ElectionCenter2008

September 26, 2008 -- Updated 2350 GMT (0750 HKT)

Presidential polling

STORY HIGHLIGHTS
- CNN's national general election poll of polls consists of multiple surveys
- It reflects registered voters' choice for president
- There is no sampling error for the poll of polls

Next Article in Politics »

- Date released:
  November 3, 2008
- Ten surveys:
  - CNN/ORC (October 30-November 1)
  - Pew (October 29-November 1)
  - CBS (October 31-November 2)
  - Fox/Opinion Dynamics (November 1-2)
  - NBC/Wall Street Journal (November 1-2)
  - ABC/Washington Post (October 30-November 2)
  - Reuters/C-SPAN/Zogby (October 31-November 2)
  - Gallup (October 31-November 2)
  - Diageo/Hotline (October 31-November 2)
  - IBD/TIPP (October 30-November 2).

- Final prediction: McCain 44%
  - Obama: 51%
  - Other: 5%

FAQ   Today's Polls   Pollster Ratings   Senate Polls and Projections   Contact   Electoral History

FiveThirtyEight   Politics Done Right

OBAMA ⭐ MCCAIN

Electoral Votes (Provisional)

365   173

FINAL PROJECTION

OBAMA ⭐ MCCAIN

ELECTORAL VOTE

348.6   189.4

WIN PERCENTAGE

98.9%   1.1%

POPULAR VOTE

52.3%   46.2%

Monday, December 15, 2008

On Making Mitch McConnell Wet His Pants

User B.R. writes in:

In your discussion of the Bayh coalition today, you used the usual 60 vote marker as the line to break filibusters.

But why is that all that folks mention? I ask because everyone (media and blogs alike) are treating the filibuster as a far more commonplace occurrence than it should be. Part of the reason for this is that the GOP has learned to use the procedural filibuster (as allowed for them by Senate Rule 22 from 1975) for everything, and Reid gives in and calls for cloture.

Reid has the power as majority leader to require actual filibusters - you know, reading the phone book for 19 hours on the senate floor. It'd also make it clear to the media and the public who stopped a piece of legislation - meaning that the story line changes from "Reid unable to find 60 votes" to "McConnell reads phone book for 19 hours to stop vote."

Why does nobody talk about that option? Why doesn't Reid exercise that option? Everyone is so stuck on "60 votes" as if it means something, but 51 votes is all that matters to pass legislation.

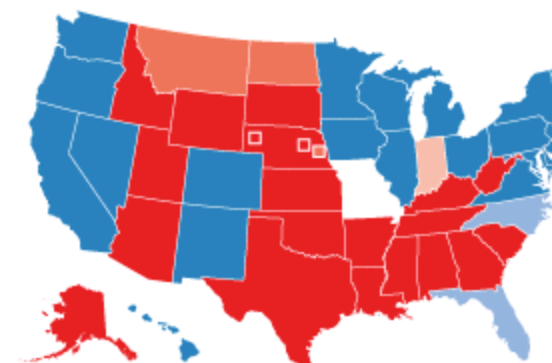**OBAMA VS MCCAIN, 2008**

Obama 365, McC

OBAMA VS. MCCAIN: FINAL PRE-ELECTION PROJECT

© FiveThirtyEight.com   11/4/08

Obama 348.6, Mc

ELECTORAL VOTE DISTRIBUTION

1200

1000

Who was closer?

	Prediction Oba	McC	Absolute Difference
CNN Poll of Polls:	51	44	3.6
Pollster:	52.0	44.4	2.2
FiveThirtyEight:	52.3	46.2	1.2
Final Results	52.9	45.7	

THE INTERESTED AMERICAN

Google™ Custom Search
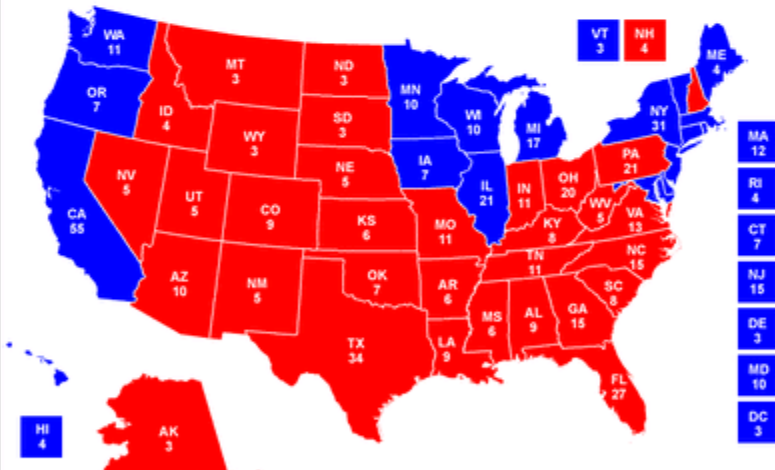
Search

TUESDAY, NOVEMBER 4, 2008

Final Electoral Vote Projection: McCain/Palin 304, Obama/Biden 234

Obama/Biden 234   McCain/Palin 304

234 Solid   0 Leaning   304 Solid   0 Leaning

Subscribe!

Ads by Google

Free Government Grant Kit

1. Averaging out the state by state polls.
2. Correcting rampant oversampling of Democrats
3. Factoring in which states the candidates are campaigning in.
4. Including an increased Black turnout.
5. An energized youth vote/an apathetic youth vote.
6. Voter fraud.
7. Throwing in a modest Bradley Effect...

Summary

```
┌──────────────┐      ┌──────────────┐      ┌──────────────┐
│              │      │              │      │              │
│ Initial GWAS │─────▶│ Replication  │─────▶│  Pooled or   │
│              │      │  Studies     │      │ Meta-Analysis│
│              │      │              │      │              │
└──────────────┘      └──────────────┘      └──────────────┘
```

- Meta-analysis estimates for genome profiles.

- Incorporate all studies: repository for these.

- Individual-level data best.

- Impute data.

- Fixed vs random effects.