Meta-Analysis For Genome-Wide Association Studies

Gonçalo Abecasis
Center for Statistical Genetics
University of Michigan

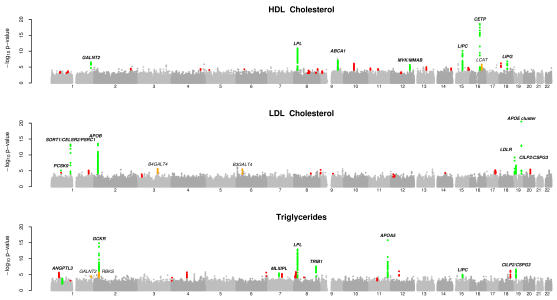
Coming Up

- Combining results across studies
- Illustration of why the approach is useful
- Discussion of newer methods for rare variant studies

Combined Lipid Scans

- SardiNIA (Schlessinger, Uda, et al.)
  - ~4,300 individuals, cohort
- FUSION (Mohlke, Boehnke, Collins, et al.)
  - ~2,500 individuals
- DGI (Kathiresan, Altshuler, Orho-Mellander, et al.)
  - ~3,000 individuals
- Individually, 1-3 hits/scan, mostly known loci
- Analysis:
  - Impute genotypes so that all scans are analyzed at the same "SNPs"
  - Carry out meta-analysis of results across scans

Combined Lipid Scan Results



First Insights…

- Meta-analysis builds on imputation methods!
  - Without imputation, each study would work from a different list of variants.
- Meta-analysis leads to lots of new signals!
- Meta-analysis does not require sharing of participant level data.

Meta-Analysis Under The Hood

	Analytical strategy	
	Sample size based	Inverse variance based
Inputs	N_i - sample size for study i $P_i - P$-value for study i Δ_i - direction of effect for study i	β_i- effect size estimate for study i se_i - standard error for study i
Intermediate Statistics	$Z_i = \Phi^{-1}(P_i/2) * \text{sign}(\Delta_i)$ $w_i = \sqrt{N_i}$	$w_i = 1/\text{SE}_i^2$ $se = \sqrt{1/\sum_i w_i}$ $\beta = \sum_i \beta_i w_i / \sum_i w_i$
Overall Z-Score	$Z = \frac{\sum_i Z_i w_i}{\sqrt{\sum_i w_i^2}}$	$Z = \beta / SE$
Overall P-value	$P = 2\Phi(\lvert -Z \rvert)$	

But Wait…

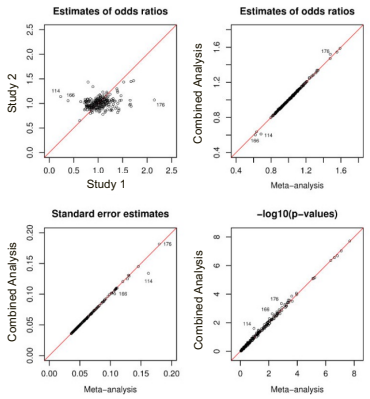
Surely, you must be losing signal by not combining raw data directly?

Genetic Epidemiology 34: 60–66 (2010)

Meta-Analysis of Genome-Wide Association Studies: No Efficiency Gain in Using Individual Participant Data

D. Y. Lin* and D. Zeng

Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina



Gene Based Burden Tests

- Most coding variants are very rare
- Testing the effect of each variant may often be impossible
- Geneticists currently favor “gene burden tests”
- These tests evaluate the combined effect of rare variants in a gene
- Can meta-analysis help reach large sample sizes with these gene based tests?

Challenges in Meta-Analysis of Burden Tests

- Each study has a different list of variants
- Variants vary in frequency between study
  - Rare variant lists will also differ!
- Similar burden test p-values could mean contradictory things in different studies!

Rare Variant Analysis in Large Samples

Seunggeung Lee

Dajiang Liu

- The insight …
- Simple burden tests can be calculated as a linear function of single variant score statistics

$$T \propto \vec{w}^T \vec{U}$$

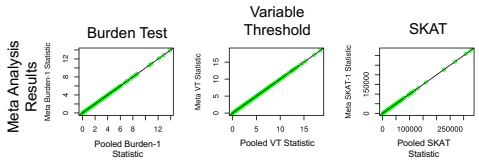
- More advanced burden tests (like SKAT) are quadratic function of single variant statistics

$$T \propto \vec{U}^T \mathbf{K} \vec{U}$$

The Solution

- For each study …
- Calculate components of single variant association score statistics
- Calculate variance covariance matrix for score statistics
- If we share these two pieces of information (single variant statistics, disequilibrium matrix) …
- … we can conveniently calculate many gene level test statistics and evaluate significance across large samples

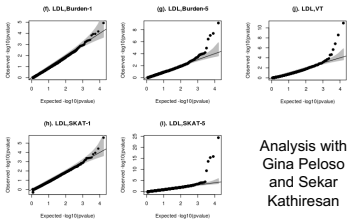
In the absence of heterogeneity, meta and pooled statistics match



Pooled Data Results

Rare Variant Meta-Analysis Example

LDL cholesterol, 15,000 individuals



Analysis with Gina Peloso and Sekar Kathiresan

QQ-plots well behaved

Rare Variant Meta-Analysis Example

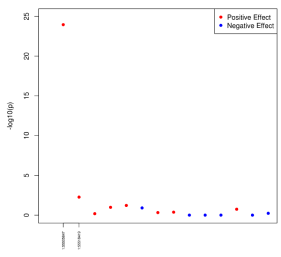
LDL cholesterol, 15,000 individuals

Gene	Burden-5	SKAT-5	Variable Threshold	VT cut-off
PCSK9	3×10^{-7}	7×10^{-25}	2×10^{-12}	.015
APOB	3×10^{-3}	2×10^{-14}	.046	.041
LDLR	.071	9×10^{-3}	2×10^{-5}	.00074

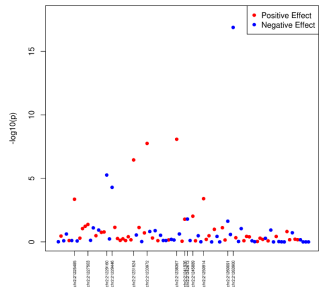
Analysis with Gina Peloso and Sekar Kathiresan

Results shown for Mendelian hypercholesterolemia genes with gene level p-value < .001 in 15,000 individuals (12 genes tested, 3 tests)

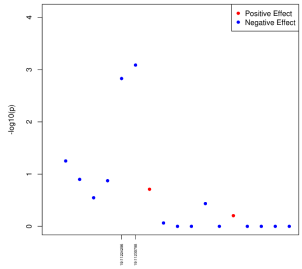
PCSK9: Marker-by-Marker



APOB: Marker-by-Marker



LDLR: Marker-by-Marker



METAL

- www.sph.umich.edu/csg/abecasis/Metal
  - Meta-analysis of three regions for fasting glucose
- genome.sph.umich.edu/wiki/Tutorial: RAREMETAL
  - Simple rare variant meta-analysis with two studies

Recommended Reading

- Willer et al (2008) *Nature Genetics* **40:**161-169
- Willer et al (2010) *Bioinformatics* **26:**2190-1
- Teslovich et al (2010) *Nature* **466:**707-713
- Liu et al (2014) *Nature Genetics* **46:**200-214