Meta-analysis

Heather Cordell
Newcastle University

heather.cordell@ncl.ac.uk

Acknowledgement: thanks to Andrew Morris who originally created many of these slides

Introduction

- We can increase power (particularly to detect rarer variants of more modest effect) by collecting larger and larger cohorts (sets of samples).
- Alternatively, we can combine the results of previously conducted GWAS of a trait through *meta-analysis*, without direct exchange of genotype data.
  - Exchange summary statistics for each SNP including "risk" allele, *p*-value, $\log_e$ odds ratio or β (effect) and 95% confidence interval (standard error).
- GWAS results can be combined for meta-analysis even if genotyped for different sets of SNPs, through the use of *imputation*.

Fixed-effects meta-analysis

- Let β_i denote the estimated allelic effect (aligned relative to a fixed baseline allele) of the *i*th study, with its estimated sampling variance denoted v_i.
- An estimate of the allelic effect over all *N* studies is then given by

$$B = \frac{\sum_i w_i \beta_i}{\sum_i w_i}$$

where $w_i = 1/v_i$, with variance of *B* given by $V = \left[\sum_i w_i\right]^{-1}$

- A test for association over all studies is given by $X^2 = B^2/V$, having an approximate chi-squared distribution with one degree of freedom.

Fixed-effects meta-analysis

- If estimates of the effect size are not available, or are not on the same "scale", we can make use of Stouffer's method for fixed-effects meta-analysis.
- Let Z_i denote the directed Z-score (aligned to a fixed baseline allele) for the ith study, and n_i denote the reported sample size.
- Combined test statistic is given by:

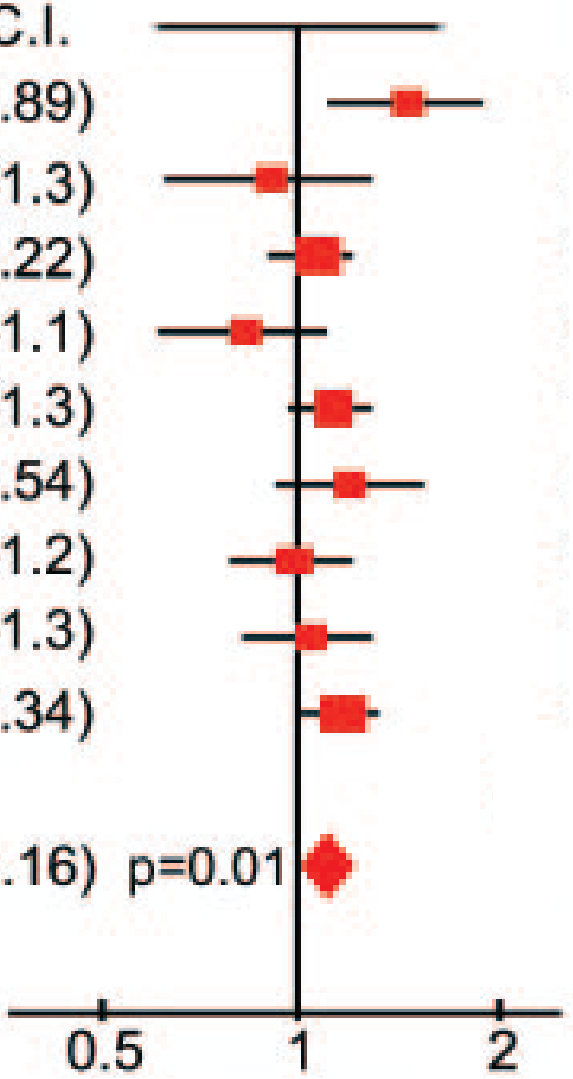
$$Z = \frac{\sum_i Z_i \sqrt{n_i}}{\sqrt{\sum_i n_i}} \sim N(0,1)$$

Assessing heterogeneity

- Fixed effects meta-analysis assumes the same odds ratio (allelic effect) over all studies. [Note this is indeed true under the null].
- We can test for heterogeneity between effects using the statistic $Q = \sum_i w_i (B - \beta_i)^2$, which has an approximate chi-squared distribution with *N*-1 degrees of freedom.
- An alternative statistic, I^2=[*Q*-(*N*-1)]/*Q*, quantifies the extent of heterogeneity from a collection of allelic effect sizes.
  - It is important to investigate the source of potential heterogeneity between the effects from the different studies.

Example: sporadic amyotrophic lateral sclerosis

Study	Cases/Controls	OR	95% C.I.
3	185/437	1.45	(1.12-1.89)
5	143/143	0.90	(0.62-1.3)
9	822/782	1.05	(0.91-1.22)
6	219/200	0.82	(0.61-1.1)
8	805/909	1.12	(0.97-1.3)
10	276/271	1.19	(0.92-1.54)
12	437/426	0.97	(0.79-1.2)
23	229/709	1.03	(0.82-1.3)
7	1035/7607	1.16	(1.01-1.34)
		OR (fixed)	
		1.09	(1.02-1.16) p=0.01



Q=12.12 (p=0.15) and I^2 = 35%

Random-effects meta-analysis

- Random-effects meta-analysis: assumes a *distribution* of true allelic effects instead of a single underlying true effect size. [Note this is almost surely true under the alternative].
- Random-effects variance component given by

$$\tau^2 = \max\left(0, \frac{Q-(N-1)}{\sum_i w_i - \left(\sum_i w_i^2 \Big/ \sum_i w_i\right)}\right)$$

- Weight assigned to *i*th study then given by

$$w_i^* = (\tau^2 + v_i)^{-1}.$$

A comment on random effects…

- Random-effects meta-analysis is often (a bit controversially?) utilised when a SNP demonstrates significant evidence of heterogeneity in allelic effects between studies.
- Important to investigate the *source* of heterogeneity between studies: variability may be due to phenotype definition, population background, interaction with exposure to environmental risk factor etc.

Allele/strand alignment

- Study 1: OR of 1.1 for allele A relative to allele G (aligned to + strand).
- Study 2: OR of 1.3 for allele C relative to allele T (aligned to – strand).
- Effect in study 2 is actually in the *opposite direction* to study 1, since A is not complementary to C.
- It is straightforward to overcome this issue for non-AT or non-GC SNPs: otherwise rely on correct strand information or matching of allele frequencies (possibly with reference to HapMap or 1000 Genomes data).

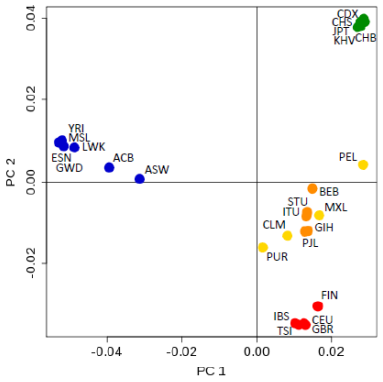
Software

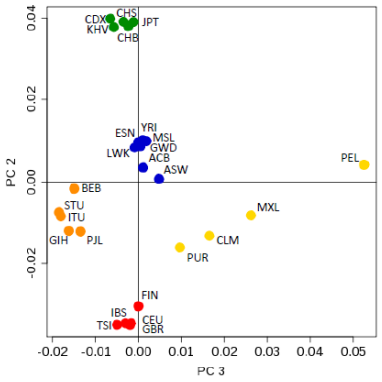
- Fixed- and random-effects meta-analysis can be performed for individual SNPs in standard statistical software packages such as R.
- Specialised software for genome-wide association meta-analysis that can handle large numbers of SNPs and studies, and can incorporate checks for strand alignment:
  - METAL: http://genome.sph.umich.edu/wiki/METAL_Program
  - GWAMA: http://www.geenivaramu.ee/en/tools/gwama
  - METASOFT: http://genetics.cs.ucla.edu/meta/
  - META: https://jmarchini.org/software/#meta

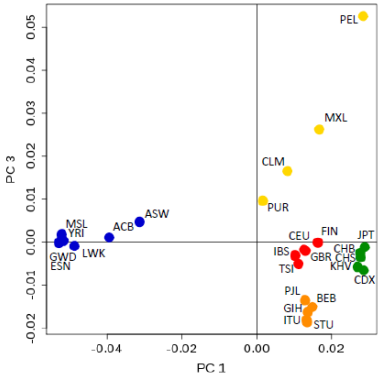
Trans-ethnic meta-analysis

- GWAS are increasingly undertaken across diverse populations.
- We expect increased heterogeneity in allelic effects between GWAS from diverse populations:
  - Differences in LD structure between ancestry groups: useful for fine-mapping causal variants.
  - Differences in exposure to an environmental risk factor that interacts with causal SNP.
- Random effects meta-analysis does not formally model the source of heterogeneity:
  - We expect “similar” populations to have less heterogeneity…..

Trans-ethnic meta-regression





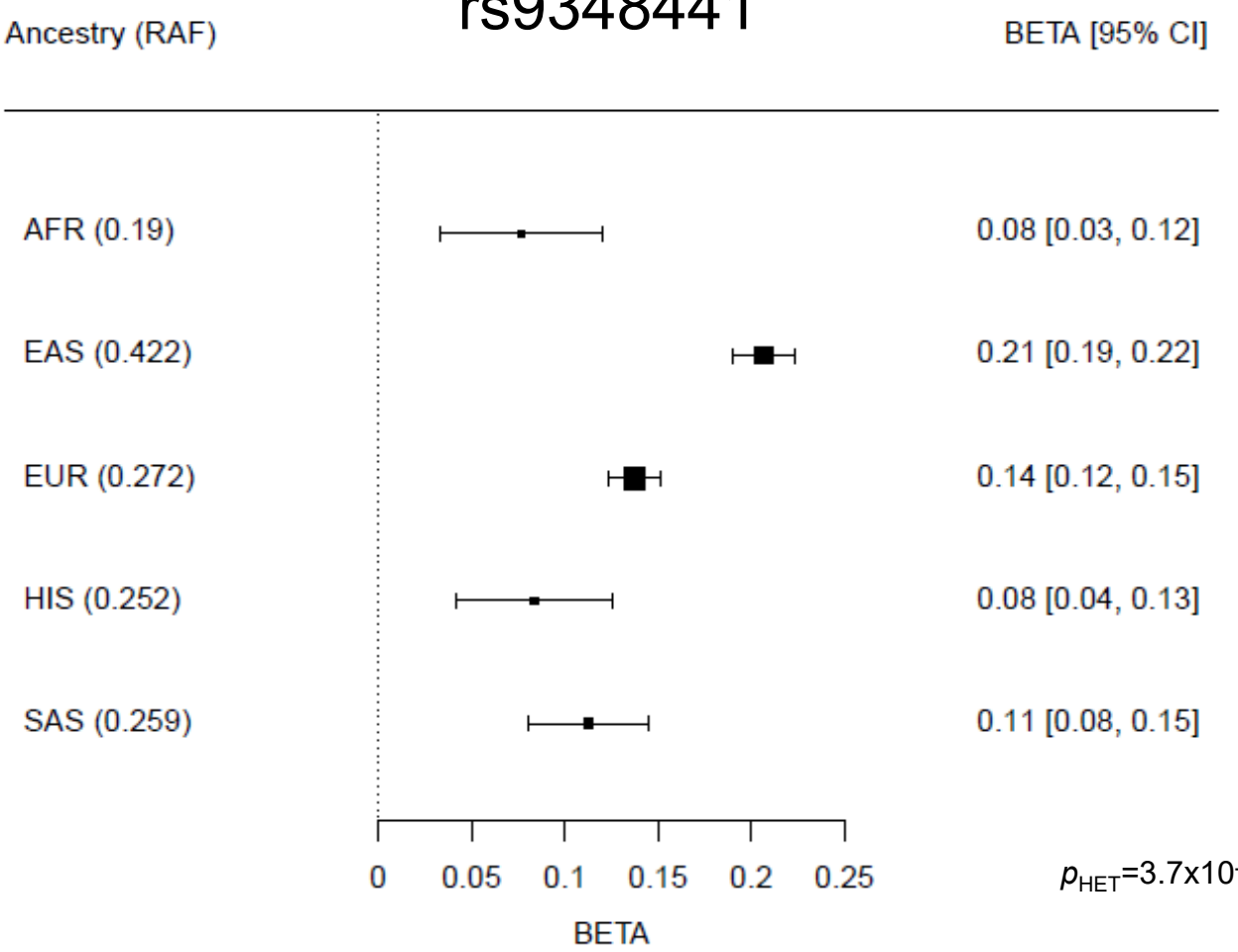


AFR
AMR
EAS
EUR
SAS

- Meta-regression can be used to model heterogeneity due to a "confounder" that differs between GWAS as a covariate.
- Principal components analysis can be used to derive "axes of genetic variation" that explain genetic differences between populations.
- MR-MEGA uses these axes to model heterogeneity in allelic effects between GWAS: the idea is those that are genetically dissimilar are more likely to have heterogeneous effects.
- https://genomics.ut.ee/en/tools/mr-mega

Type 2 diabetes

CDKAL1
rs9348441



$p_{HET}=3.7x10^{-14}$

Summary

- Meta-analysis of GWAS summary statistics allows increase in power to detect association without direct exchange of genotype (and other relevant phenotype) data.
- Fixed-effects meta-analysis assumes homogenous allelic effect across studies
  - Important to assess evidence for heterogeneity.
- Software designed for GWAS meta-analysis has features that allow for strand alignment and SNP filtering.
- Particularly important to model heterogeneity in trans-ethnic meta-analysis.