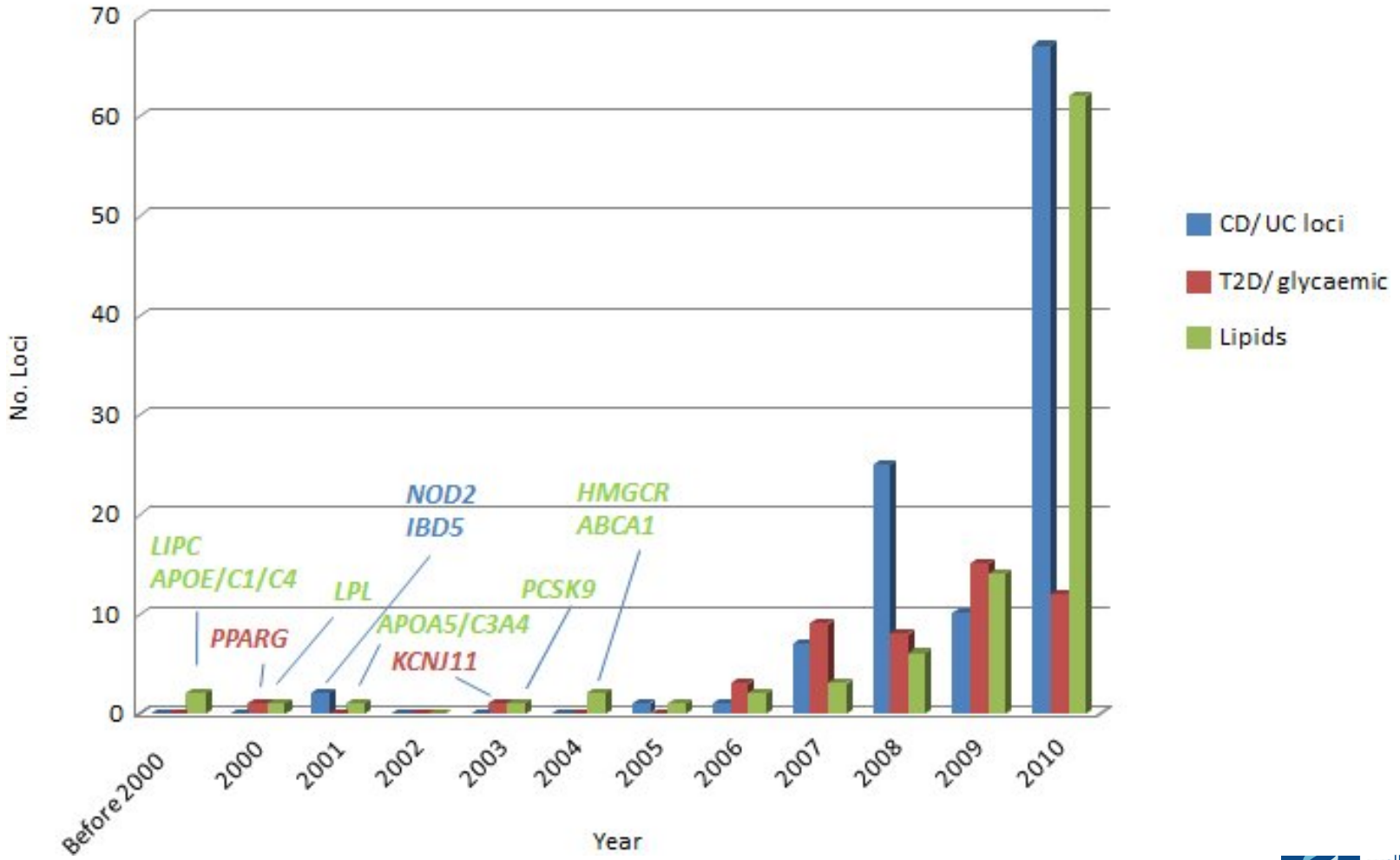
Meta-analysis strategies in genome-wide association studies

Ioanna Tachmazidou

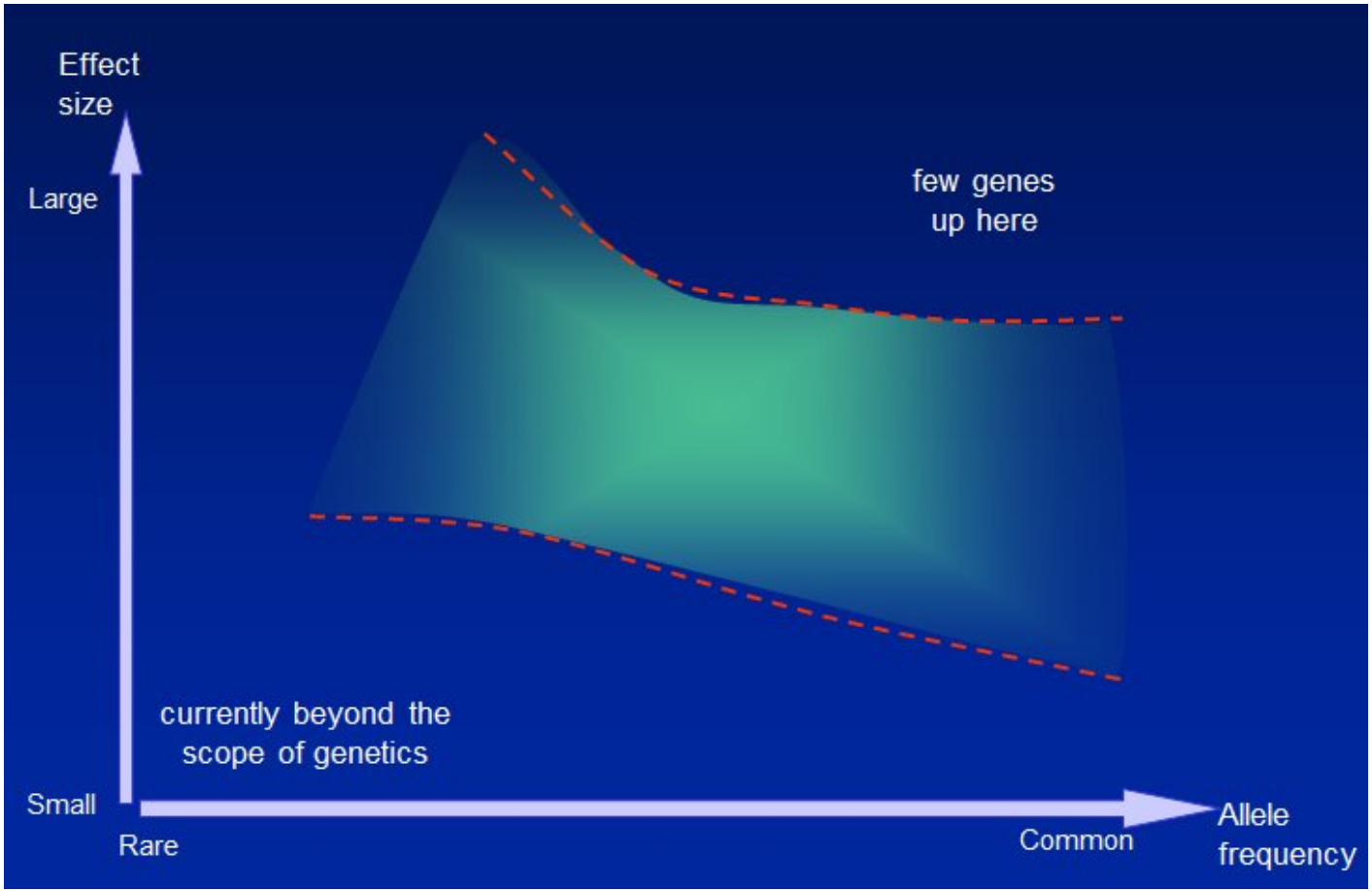
The Wellcome Trust Sanger Institute

Volos Summer School 2017

Genetic landscape pre- and post-GWAS



Atlas of complex disease susceptibility



Motivation for meta-analysis of GWAS

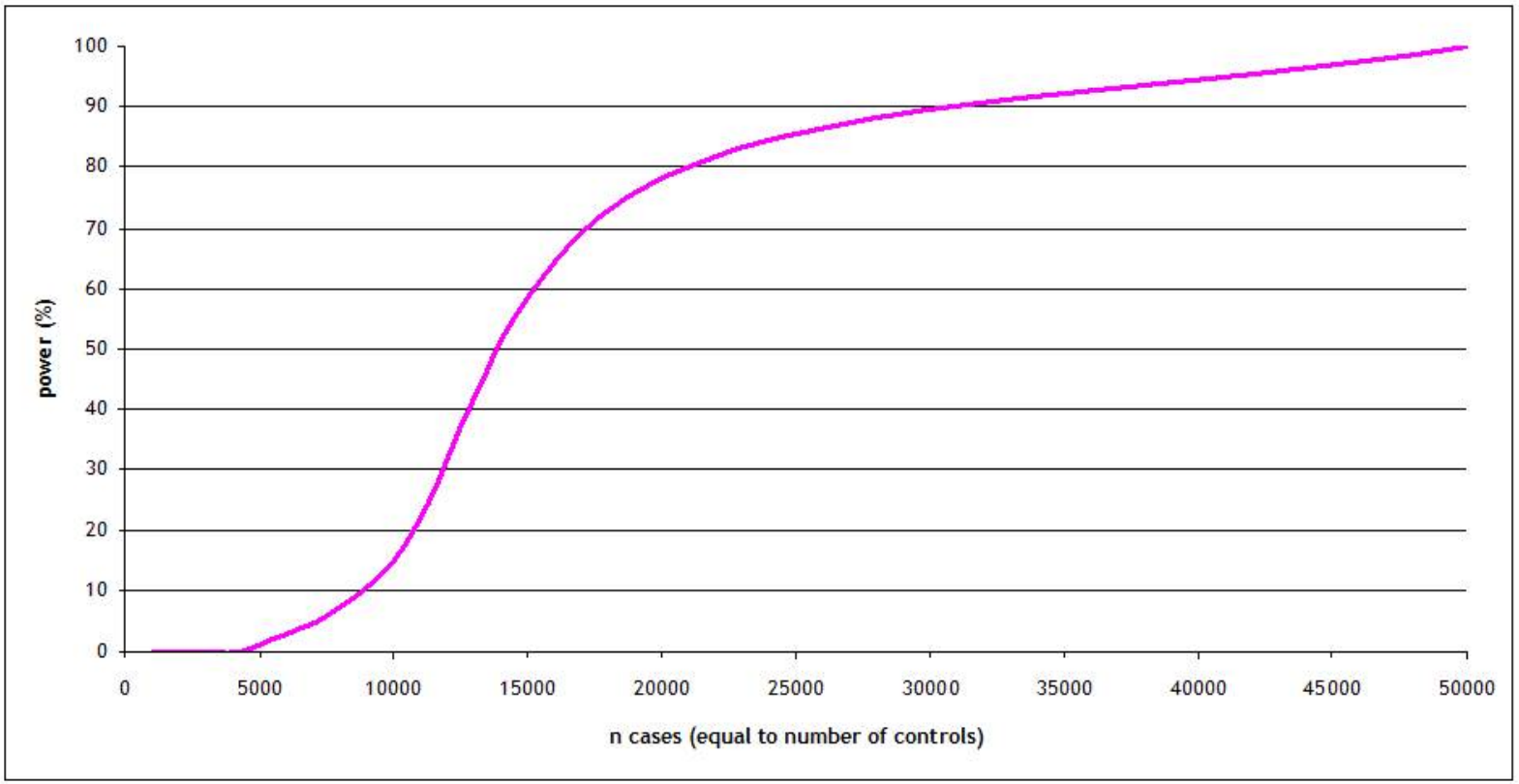
Data from diverse studies examining the same trait can be synthesized within a meta-analysis framework

Increase in power through:

- Increased sample size
- Imputation of untyped variants

Motivation for meta-analysis of GWAS

Power to detect association ($p = 5 \times 10^{-8}$) at a variant with risk allele frequency 0.30 and allelic OR 1.10



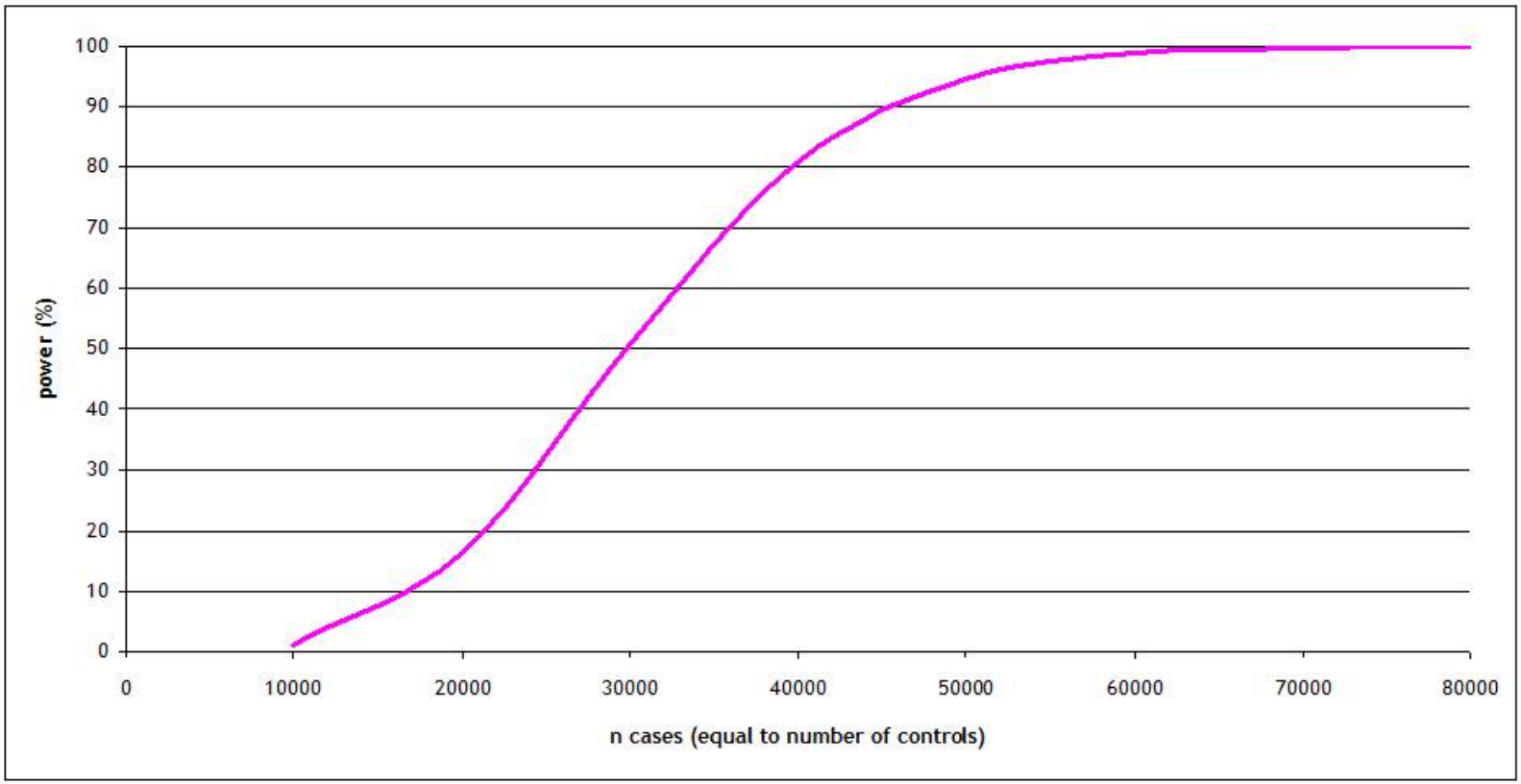
Motivation for meta-analysis of GWAS

- Recent shift of the complex trait genetics field towards low frequency (between 1-5%) and rare (less than 1%) variation
- The issue of sample size and power is more pronounced in the study of rare variation, because large effect sizes are unlikely to exist for polygenic disorders

wellcome trust sanger institute

Motivation for meta-analysis of GWAS

Power to detect association ($p = 5 \times 10^{-8}$) at a variant with risk allele frequency 0.005 and allelic OR 1.50



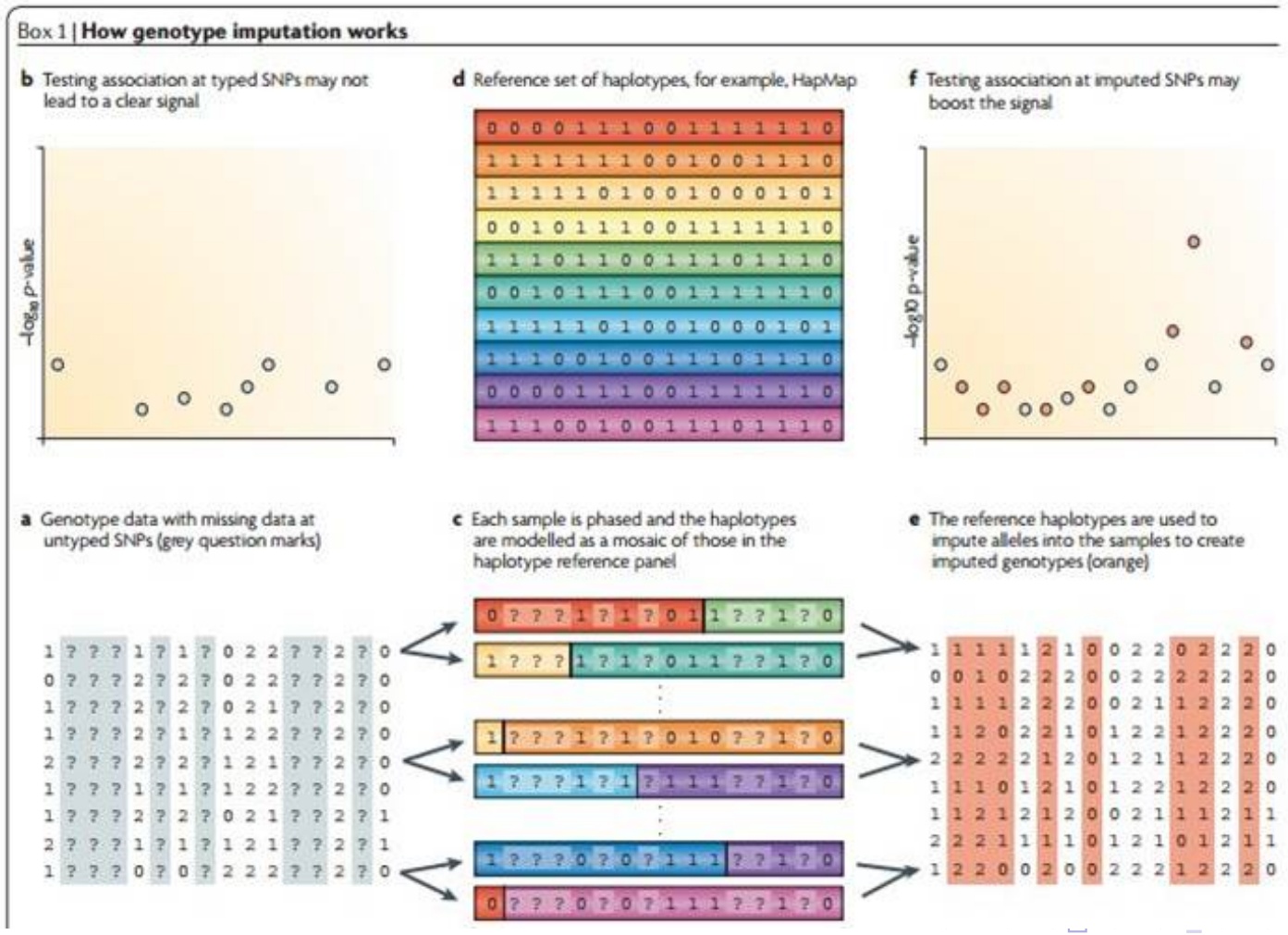
Principles of meta-analysis

- Synthesis of different datasets to obtain a summary based on evidence from the combined data
- Carried out to help the discovery of susceptibility variants of moderate/small effect size that would have otherwise escaped detection due to low power
- They can be carried out sequentially, and can be updated when new GWAS for the same trait emerge
- Facilitated by imputation, which enables the combination of data across different genotyping platforms

wellcome trust sanger institute

Imputation

Reference datasets, such as www.1000genomes.org, www.uk10k.org and www.haplotype-reference-consortium.org, are used to impute genotypes at untyped positions in the target dataset



Forming consortia

The first step in many GWAS meta-analyses involves setting up consortia to study specific traits of interest

DIAbetes Genetics
Replication And Meta-analysis

wellcome trust sanger institute

Big effort

ARTICLES

nature genetics

Defining the role of common variation in the genomic and biological architecture of adult human height

Andrew R Wood[1,323], Tonu Esko[2–5,323], Jian Yang[6,7,323], Sailaja Vedantam[3,4,323], Tune H Pers[3–5,8,323], Stefan Gustafsson[9,10,323], Audrey Y Chu[11], Karol Estrada[4,12,13], Jian'an Luan[14], Zoltán Kutalik[15–17], Najaf Amin[18], Martin L Buchkovich[19], Damien C Croteau-Chonka[19,20], Felix R Day[14], Yanan Duan[21], Tove Fall[9,10,22], Rudolf Fehrmann[23], Teresa Ferreira[24], Anne U Jackson[25], Juha Karjalainen[23], Ken Sin Lo[26], Adam E Locke[25], Reedik Mägi[2,24], Evelin Mihailov[2,27], Eleonora Porcu[28], Joshua C Randall[24,29], André Scherag[30,31], Anna A E Vinkhuyzen[6], Harm-Jan Westra[23], Thomas W Winkler[32], Tsegaselassie Workalemahu[33], Jing Hua Zhao[14], Devin Absher[34], Eva Albrecht[35], Denise Anderson[36], Jeffrey Baron[37], Marian Beekman[38,39], Ayse Demirkan[18,40], Georg B Ehret[41,42], Bjarke Feenstra[43], Mary F Feitosa[44], Krista Fischer[2], Ross M Fraser[45], Anuj Goel[24,46], Jian Gong[47], Anne E Justice[48], Stavroula Kanoni[49], Marcus E Kleber[50,51], Kati Kristiansson[52], Unhee Lim[53], Vaneet Lotay[54], Julian C Lui[37], Massimo Mangino[55], Irene Mateo Leach[56], Carolina Medina-Gomez[12,57,58], Michael A Nalls[59], Dale R Nyholt[60], Cameron D Palmer[3,4], Dorota Pasko[1], Sonali Pechlivanis[30], Inga Prokopenko[24,61,62], Janina S Ried[35], Stephan Ripke[13,63], Dmitry Shungin[64–66], Alena Stancáková[67], Rona J Strawbridge[68], Yun Ju Sung[69], Toshiko Tanaka[70], Alexander Teumer[71], Stella Trompet[72,73], Sander W van der Laan[74], Jessica van Setten[75], Jana V Van Vliet-Ostaptchouk[76], Zhaoming Wang[77–80], Loïc Yengo[81–83], Weihua Zhang[84,85], Uzma Afzal[84,85], Johan Ärnlöv[9,10,86], Gillian M Arscott[87], Stefania Bandinelli[88], Amy Barrett[61], Claire Bellis[89], Amanda J Bennett[61], Christian Berne[90], Matthias Blüher[91,92], Jennifer L Bolton[45], Yvonne Böttcher[91], Heather A Boyd[43], Marcel Bruinenberg[93], Brendan M Buckley[94], Steven Buyske[95,96], Ida H Caspersen[97], Peter S Chines[98], Robert Clarke[99], Simone Claudi-Boehm[100], Matthew Cooper[36], E Warwick Daw[44], Pim A De Jong[101], Joris Deelen[38,39], Graciela Delgado[50], Josh C Denny[102], Rosalie Dhonukshe-Rutten[103], Maria Dimitriou[104], Alex S F Doney[105], Marcus Dörr[77,106], Niina Eklund[52,107], Elodie Eury[81–83], Lasse Folkersen[68], Melissa E Garcia[108], Frank Geller[43], Vilmantas Giedraitis[109], Alan S Go[110], Harald Grallert[35,111,112], Tanja B Grammer[50], Jürgen Gräßler[113], Henrik Grönberg[22], Lisette C P G M de Groot[103], Christopher J Groves[61], Jeffrey Haessler[47], Per Hall[22], Toomas Haller[2], Goran Hallmans[114], Anke Hannemann[78], Catharina A Hartman[115], Maija Hassinen[116],

wellcome trust sanger institute

Meta-analysis of GWAS

- Requires a robust predefined protocol, specifying genetic model examined, strategy for covariate adjustments etc
- The majority of GWAS meta-analyses combine data retrospectively, making harmonization of study design difficult
- Requires summary statistics at each variant
- Information on:
  - Analysis method and covariates used
  - Size of the study
  - Independence of samples
  - Approaches taken to adjust for any population stratification
  - Strand and build of the human genome, on which allele coding has been based

wellcome trust sanger institute

Standardizing QC and analysis method

- Phenotype definition
- Genotype level QC (call rate, HWE, MAF)
- Imputation (reference, analysis)
- Association testing
- Correcting for population structure prior to the meta-analysis
- Centralized v distributed meta-analysis

Typical data sharing table format

STUDY TITLE	
General information	
Name of study	
Name of analyst	
Email of analyst	
Study design	population-based, family-based –please give details
Sample information	
Number of cases (females)	
Number of controls (males)	
Ethnic composition	
Possible relatedness issues	are individuals related (how?)
Possible structure issues	mixed population?
Genotyping and imputation information	
Genotyping platform	
Summary of key QC metrics	
# SNPs passed QC	
Imputation method	
Imputation settings	
Reference data used for imputation	including build
Analytical information	
Association analysis method for imputed genotypes	accounting for uncertainty using SNPTEST or other (which?) program, using only genotypes with P(call)>X (which threshold?) as hard calls, using best guess genotypes
Calculated GC lambda (typed SNPs)	
Calculated GC lambda (imputed SNPs)	
Covariates included	PCA, GC, none
Genetic model	

Typical data sharing table format

Column header	Description
SNP	SNP rs number (if unknown, e.g. with some Affymetrix SNPs, report Affy SNP ID)
build	e.g. “36”, human genome build used
strand	e.g. “+”, human genome strand used
chromosome	chromosome on which SNP resides
position	position of SNP on chromosome in base pairs, based on human genome build used
imputed	“1” for imputed, “0” for directly-typed SNP passing QC
major_allele	e.g. “G”, major allele at that SNP, based on control frequency
minor_allele	e.g. “A”, minor allele at that SNP, based on control frequency
MAF_controls	e.g. “0.246”, minor allele frequency in controls -provide 3 digits to the right of the decimal
OR_allele	e.g. “A”, allele to which the OR has been estimated
call_rate	e.g. “0.985”, call rate for this SNP across cases and controls -provide 3 digits to the right of the decimal
exact_HWE_cases	exact HWE p value in cases
exact_HWE_controls	exact HWE p value in controls
OR	e.g. “1.097”, allelic odds ratio –provide 3 digits to the right of the decimal
lower_95%CI	e.g. “0.874”, lower 95% confidence interval of the OR –provide 3 digits to the right of the decimal
upper_95%CI	e.g. “1.267”, upper 95% confidence interval of the OR –provide 3 digits to the right of the decimal
additive_p_uncorr	additive model p value, uncorrected for genomic control
additive_p_corr	additive model p value, corrected for genomic control
impute_acc	e.g. “0.98”, metric for imputation accuracy (i.e. value for r^2hat or proper_info measures, depending on imputation programme used; if some other measure used, please specify)

Overview of meta-analysis methods

- Fixed effects
- Random effects
- Bayesian
- Based on estimate of effect size (e.g. OR-based) or on p-value
  - Must have independent set of effect sizes
  - Larger studies should carry more weight

Fixed effects meta-analysis

- Inverse variance based

  β_i - effect size estimate for study i

  se_i - standard error for study i

  - $w_i = 1/se_i^2$
  - $SE = \sqrt{1/\sum_i w_i}$
  - $\beta = \sum_i(\beta_i \times w_i)/\sum_i w_i$
  - $Z = \beta/SE \Rightarrow P = 2 \times (1 - \Phi(|Z|))$

Fixed effects meta-analysis

- Inverse variance based

  β_i - effect size estimate for study i

  se_i - standard error for study i

  - $w_i = 1/se_i^2$
  - $SE = \sqrt{1/\sum_i w_i}$
  - $\beta = \sum_i(\beta_i \times w_i)/\sum_i w_i$
  - $Z = \beta/SE \Rightarrow P = 2 \times (1 - \Phi(|Z|))$

- Sample size based

  N_i - sample size for study i

  P_i - p-value for study i

  Δ_i - direction of effect for study i

  - $Z_i = \Phi^{-1}(P_i/2) \times \text{sign}(\Delta_i)$
  - $w_i = \sqrt{N_i}$
  - $Z = \frac{\sum_i(Z_i \times w_i)}{\sqrt{\sum_i w_i^2}} \Rightarrow P = 2 \times (1 - \Phi(|Z|))$

Correcting for population structure

- Within study variation
  - Correct test statistics $\chi_i^2 = (\beta_i / se_i)^2$ by the genomic control inflation factor $\lambda = \text{median}(\chi_i^2)/0.456$.
  - Calculate λ separately for directly genotyped and imputed SNPs, λ_{Di} and $\lambda_{D^\star i}$ respectively for study i.
  - Adjust weights to $w_i^{\text{adj}} = \lambda_{Ki} \times w_i$, where K is replaced by D or $D^\star$ as appropriate.
- Between studies variation
  - $X^2 = Z^2 = \beta^2/(\lambda \times SE^2)$, where λ is the genomic control inflation factor over all meta-analyzed association test statistics.

Assessment of heterogeneity

The first step in carrying out a meta-analysis involves assessing heterogeneity across the combined studies

Statistics:

- Cochran's Q: is there heterogeneity?
  Q_j for SNP j depends on the number of studies for which an allelic effect is reported

- I^2: how much heterogeneity is there?
  I^2_j is more robust to variability in the number of studies included in the meta-analysis for SNP j

Potential reasons for heterogeneity

- Variable LD patterns across studies: the identified marker is not the causal polymorphism, but has a different LD pattern with the causal polymorphism across different studies.
- Gene-environment interactions with different environmental exposures across populations.
- Genuine genetic heterogeneity in effect sizes across different ethnic backgrounds and population-specific effects.
- Study-specific bias: genotyping errors, population stratification, phenotype missclassification.
- Winner's curse: The originally identified effect size is likely to be overestimated in comparison to its true value.

Interpreting heterogeneity

- Heterogeneity may represent genuine differences in genetic effects across different populations and different biological setting (truly informative heterogeneity).
- Informative heterogeneity may reveal interesting facts about biology, e.g. the mechanism through which the variant is acting on disease risk.
- Recognizing the potential for heterogeneity can rescue associations from being discarded as replication failures.

Interpreting heterogeneity

- Heterogeneity may represent genuine differences in genetic effects across different populations and different biological setting (truly informative heterogeneity).
- Informative heterogeneity may reveal interesting facts about biology, e.g. the mechanism through which the variant is acting on disease risk.
- Recognizing the potential for heterogeneity can rescue associations from being discarded as replication failures.
- An example of informative heterogeneity is the relationship between the obesity associated (*FTO*) gene and T2D. Overall the evidence suggests that *FTO* affects weight, which in turn increases risk of T2D. As a consequence of being on the same pathway, *FTO* showed an association with T2D in population-based studies but failed to replicate in studies that controlled for weight by only recruiting lean subjects.

wellcome trust sanger institute

Random effects meta-analysis

A variance component τ^2 is used to inflate the variance of the estimated allelic effect in each study

$$w_i = 1/se_i^2 \Rightarrow w_i^\star = 1/(\tau^2 + se_i^2)$$

Specialized software for GWAS meta-analysis

GWAMA
Genome-Wide Association Meta-Analysis
www.well.ox.ac.uk/gwama/contact.shtml

METAL
Meta-Analysis Helper
www.sph.umich.edu/csg/abecasis/metal

META
www.stats.ox.ac.uk/ jsliu/meta.html

MetABEL
R package part of GenABEL, an R library for GWAS analysis
www.genabel.org/packages/MetABEL

METASOFT
Han and Eskin's Random Effects model, Binary Effects model
http://genetics.cs.ucla.edu/meta

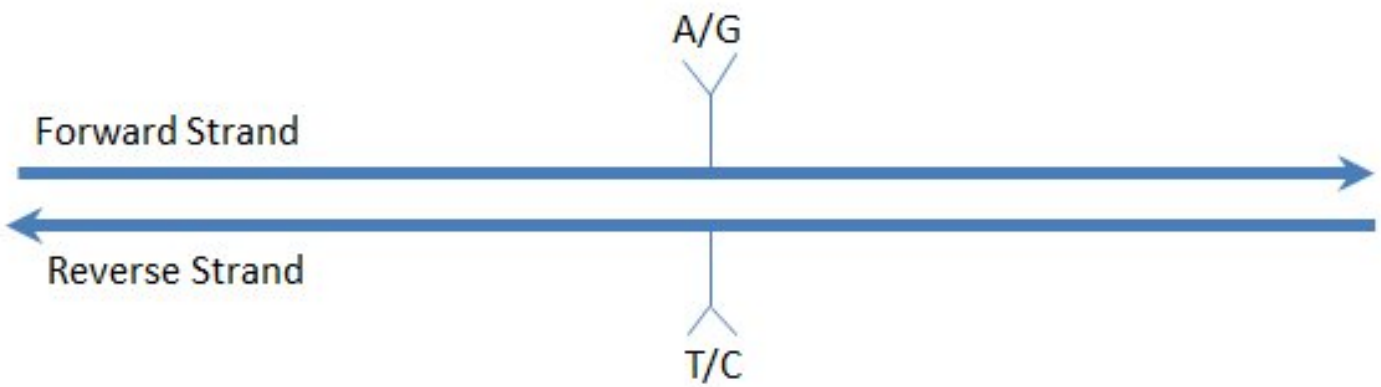
Comparison of software

Magi and Morris, BMC Bioinformatics 2010

Software package	METAL	MetABEL	META	GWAMA
Pre-processing of GWA files	No	ABEL	SNPTEST	SNPTEST,PLINK
Strand flipping	Yes	Yes	Yes	Yes
Fixed effect analysis	Yes	Yes	Yes	Yes
Random effect analysis	No	No	Yes	Yes
Heterogeneity statistics	Q	No	Q, I^2	Q, I^2
Genomic control	Yes	Yes	Yes	Yes
Graphical visualization	No	Forest plot	No	QQ and Manhattan plots

Remapping genome build and strand flipping

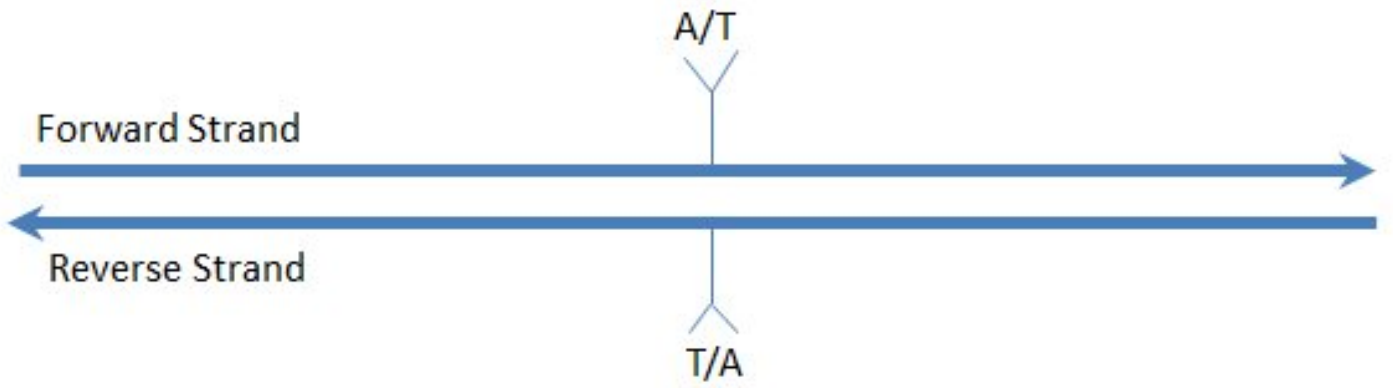
- A key issue to be dealt with prior to any meta-analysis is ensuring all the data sets are aligned to the same strand on the same genome build. This is usually the forward or positive (+) strand.
- For most SNPs the strand can be identified by the alleles.



- SNP is A/G on the forward strand and T/C on the reverse, thereby uniquely identifying the strand.

Alleles Are Not The Full Story

- For some SNPs the strand cannot be determined using the alleles.



- In this case the alleles on either strand are A/T.
- This is the same for G/C SNPs as well.

Using Allele Frequency to Determine Strand

- Assume SNP is A/T, the Minor Allele is A with a frequency (MAF) of 30%.
- A second study with the SNP listed as Minor Allele T with a frequency of 32% is likely on the opposite strand.
- This is not conclusive as SNPs vary in frequency between populations.
- A/T or G/C SNPs with a frequency near 50% are particularly difficult to determine using frequency.
- Files containing strand and position information for common genotype chips can be downloaded from http://www.well.ox.ac.uk/ wrayner/strand/

wellcome trust sanger institute

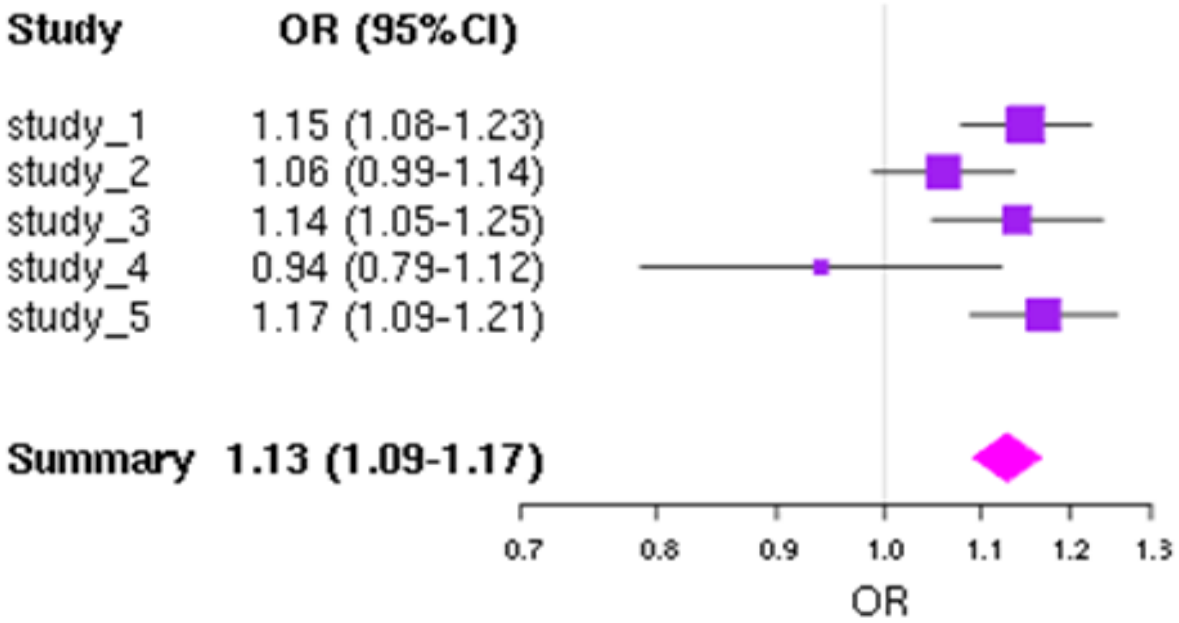
Study alignment and error trapping

Example of alignment of allelic effects and error trapping for a single SNP in a meta-analysis of five studies of a dichotomous phenotype. Magi and Morris, BMC Bioinformatics 2010

Study	Reported strand	Effect allele[1]	Other allele	RAF	Odds ratio (95% confidence interval)	Aligned allelic effect (standard error)	Comment
1	+	A	G	0.12	1.12 (1.07-1.16)	0.11 (0.02)	Allele A taken as reference effect allele.
2	+	G	A	0.85	0.92 (0.87-0.98)	0.08 (0.03)	Effect aligned to allele A.
3	-	T	C	0.12	1.06 (1.02-1.10)	0.06 (0.02)	Effect aligned to allele A on + strand.
4	+	T	C	0.13	1.07 (0.99-1.16)	0.07 (0.04)	Effect aligned to allele A on + strand. Strand error reported to log file.
5	+	A	G	0.87	0.95 (0.90-1.01)	-0.05 (0.03)	Large discrepancy in EAF reported to log file.

[1] Effects are aligned to the reference allele in the first study. Errors in the reported strand are recorded in the log file together with warnings regarding potential discrepancies in reported data between studies, for example the aligned reference allele frequency (RAF).

Visualizing meta-analysis results



Interpreting meta-analysis results

- In the absence of between-study heterogeneity, fixed and random effects calculations yield identical point estimates and confidence intervals.
- With increasing between-study heterogeneity, the random effects summary estimates have larger variance (wider confidence intervals) and usually less prominent statistical significance.
- Most meta-analysts would typically run both models, but prefer placing emphasis on random effects.
- Statistically significant associations in fixed or random effects calculations need replication.

wellcome trust sanger institute

Interpreting meta-analysis results

- Effect sizes should be known with precision, if this information were to be used successfully for predictive risk modelling.
  - The analytical methods used, e.g. genetic model specification, may affect the magnitude of the effect size.
  - Significant associations are likely to have observed effect sizes that are inflated compared with the true effect size.

Interpreting meta-analysis results

- Effect sizes should be known with precision, if this information were to be used successfully for predictive risk modelling.
  - The analytical methods used, e.g. genetic model specification, may affect the magnitude of the effect size.
  - Significant associations are likely to have observed effect sizes that are inflated compared with the true effect size.
- It is difficult to know whether a newly discovered genetic variant is the causal variant or simply linked to it.
  - Even large-scale GWA meta-analyses require extensive fine-mapping and targeted resequencing experiments before the truly causal variants can be identified.

wellcome trust sanger institute

Interpreting meta-analysis results

- Effect sizes should be known with precision, if this information were to be used successfully for predictive risk modelling.
  - The analytical methods used, e.g. genetic model specification, may affect the magnitude of the effect size.
  - Significant associations are likely to have observed effect sizes that are inflated compared with the true effect size.
- It is difficult to know whether a newly discovered genetic variant is the causal variant or simply linked to it.
  - Even large-scale GWA meta-analyses require extensive fine-mapping and targeted resequencing experiments before the truly causal variants can be identified.
- Discovery of a genetic locus has important implications on its own, regardless of the observed effect size, e.g. it may highlight some interesting biological pathway and may give some insights into developing new therapeutics.

Replication of meta-analysis findings

- Prioritize interesting signals (whether they reach genome-wide significance or not) for follow-up and replication.
- Follow-up sample sets should be adequately powered to detect the association. The replication stage could be a large meta-analysis itself.
- Issues of set-up, information aggregation, estimation of heterogeneity and summary effects in a replication effort are similar to those described for discovery meta-analysis.
- Meta-analyze the discovery with the replication data to capture the totality of the evidence.
- Review the literature and bioinformatic databases to identify candidate variants both for the particular trait under study and for associated traits.
- Replication of previously published hits confirms previous publications and gives validity to the meta-analysis.

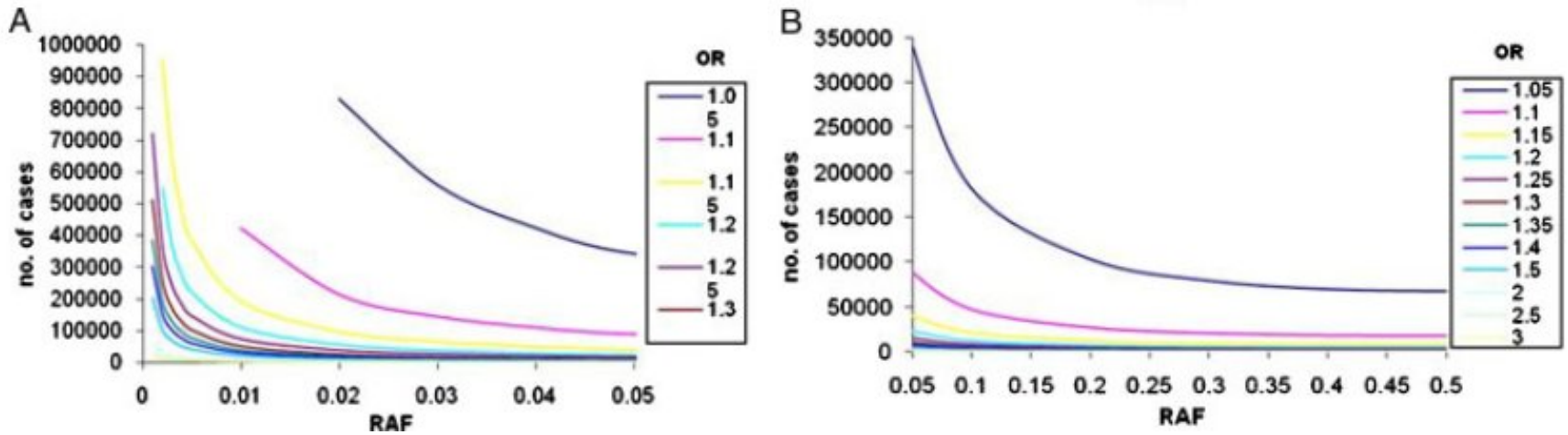
wellcome trust sanger institute

Stages of a meta-analysis

- Sensitivity analyses
  - Does the effect extend over a chromosomal region or is it confined to one variant?
  - Do the results depend critically on a single study? Why?
  - Examine cluster plots of the confirmed signals.
  - Is the effect stronger under a recessive or dominant genetic model?
  - Is there large heterogeneity at a locus?
- Secondary analyses
  - Assess the importance of some variants adjusted for others, in order to see if the two sets act independently.
  - Adjust for phenotypes that lie on potential causal pathways.
  - Is the signal driven or stronger in males or females?

Power limits of current genome-wide meta-analysis

Sample sizes required to reach 80% power at $\alpha = 5 \times 10^{-8}$ for low (left) and common allele frequencies (right).

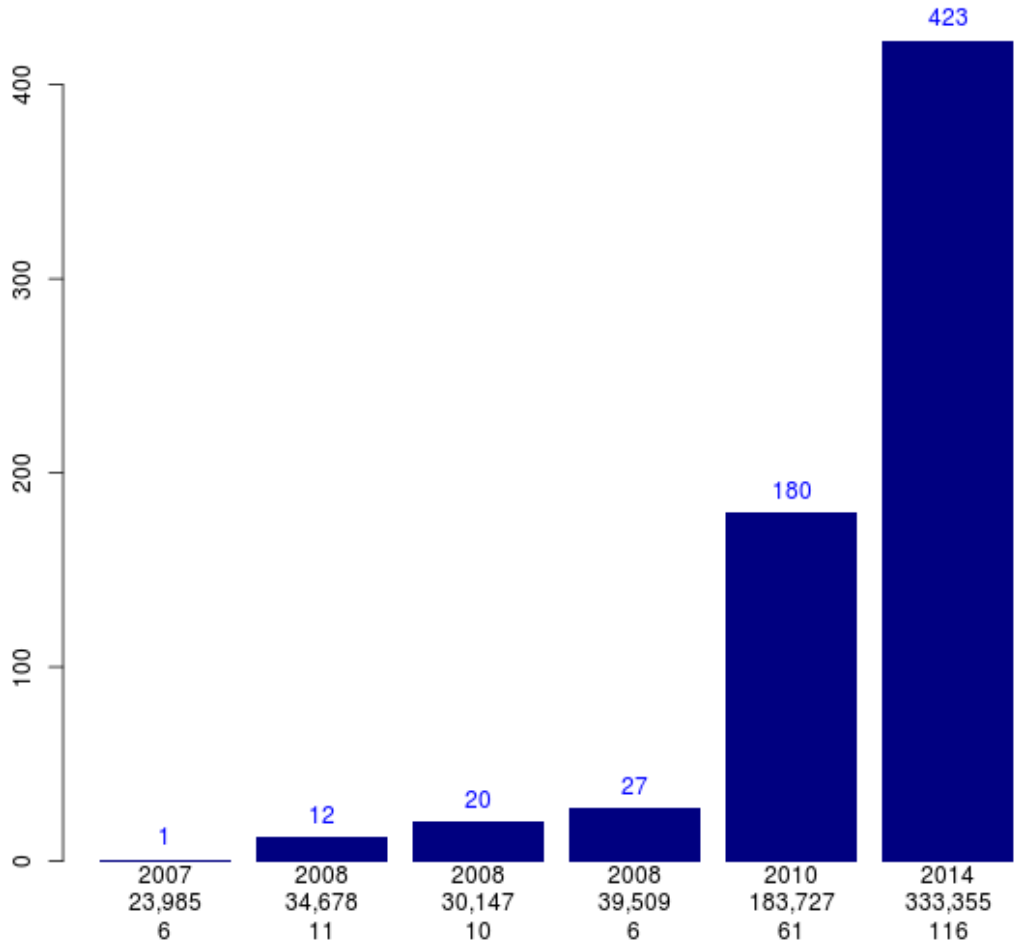


Chapman et al, GenEpi 2011

- Large consortia with adequate power to detect association with low frequency alleles (MAF down to 0.01) and effect sizes of at least 1.10.
- The AF and effect size of genetic determinants could be diverse (common variants of small effect size, low-frequency or rare variants of modest effect size).

Height discoveries

Per year as a function of sample size and number of studies



Summary

- Genetic effects of common variants associated with complex diseases are mostly moderate/small and require very large sample sizes to identify with certainty.
- Meta-analysis of genome-wide data can improve the power for detecting and validating such associations.
- Meta-analysis of data from studies using different platforms is enhanced by the use of imputed genotype data.
- Careful collection and quality-checking of information is essential to avoid errors.
- A wide array of methods may be used, including fixed effects, random effects, and Bayesian meta-analysis and they have particular advantages and disadvantages.
- Application of the meta-analysis methodology in genome-wide data has been successful in identifying more disease-related genes for some conditions.

wellcome trust sanger institute

Appendix

Random effects meta-analysis

The random effects variance component at SNP j is

$$\tau^2 = \max\left(0, \frac{Q_j - (N_j - 1)}{\sum_i w_{ij} - (\sum_i w_{ij}^2 / \sum_i w_{ij})}\right)$$

The combined allelic effect across all studies is $B_j^\star = \frac{\sum_{i=1}^N \beta_{ij} w_{ij}^\star}{\sum_{i=1}^N w_{ij}^\star}$, where $w_{ij}^\star = 1/(\tau^2 + se_{ij}^2)$

The random effects test statistic $X_j^2 = B_j^{\star 2} / V_j^\star \sim \chi_1^2$, where $V_j^\star = 1/\sum_{i=1}^N w_{ij}^\star$

The random effects variance component τ^2 is used to inflate the variance of the estimated allelic effect in each study

Random effects model that increases power under heterogeneity

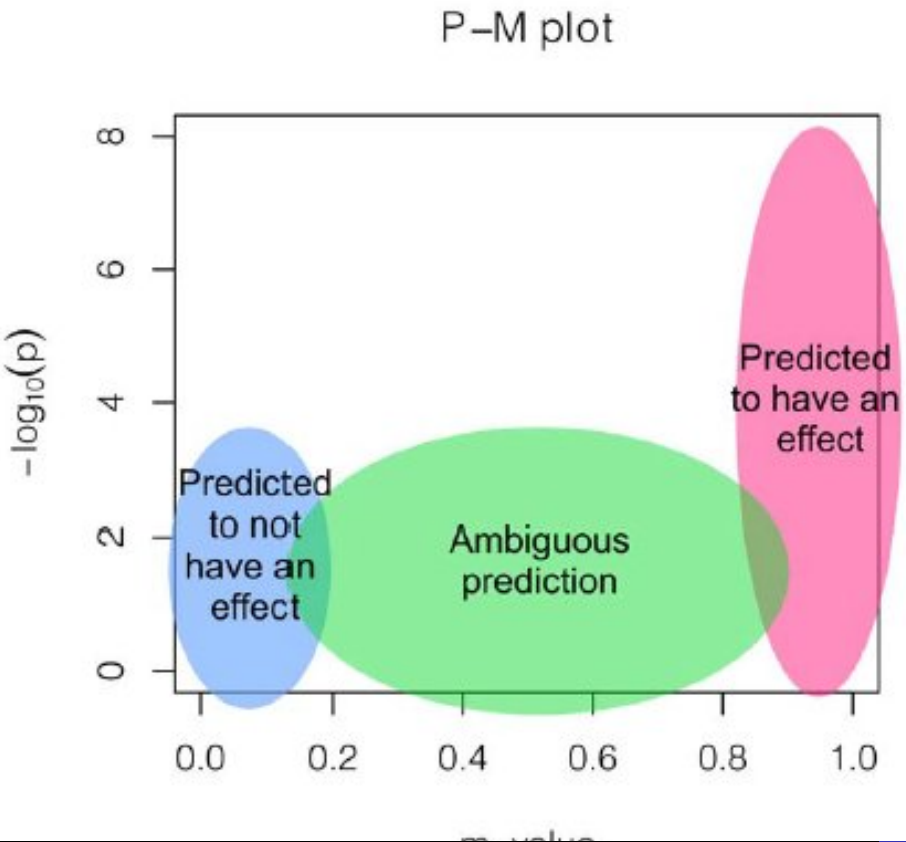
Han and Eskin, AJHG 2011

- The traditional RE implicitly assumes heterogeneity under the null hypothesis, which makes it conservative.
- Alternative RE model assumes no heterogeneity under the null hypothesis.
  - The effect size estimate and its confidence interval is the same.
  - A wide confidence interval might not always correspond to a statistically nonsignificant result.
- Likelihood ratio test statistic $S^{\star}_{RE} = S_{FE} + S_{HET}$, where $S_{FE} = Z^2_{FE}$ and S_{HET} is the test statistic testing for heterogeneity.
- Its statistical significance is estimated using tabulated values.

Bayesian meta-analysis: Binary effects model

Han and Eskin, PLoS Gen 2012

- Based on the posterior probability that the effect exists in each study (m-value).
- Estimated using cross-study information via MCMC.
- Segregates the studies predicted to have an effect, the studies predicted to not have an effect, and the underpowered ones.



If m-value > 0.9, the study is predicted to have an effect. If m-value < 0.1, the study is predicted to not have an effect.

Binary effects test statistic

$S_{BE} = \frac{\sum m_i \sqrt{W_i} Z_i}{\sqrt{\sum m_i^2 W_i}}$, where

$Z_i = \beta_i / se_i$ and $\sqrt{W_i} = \sqrt{N_i}$.