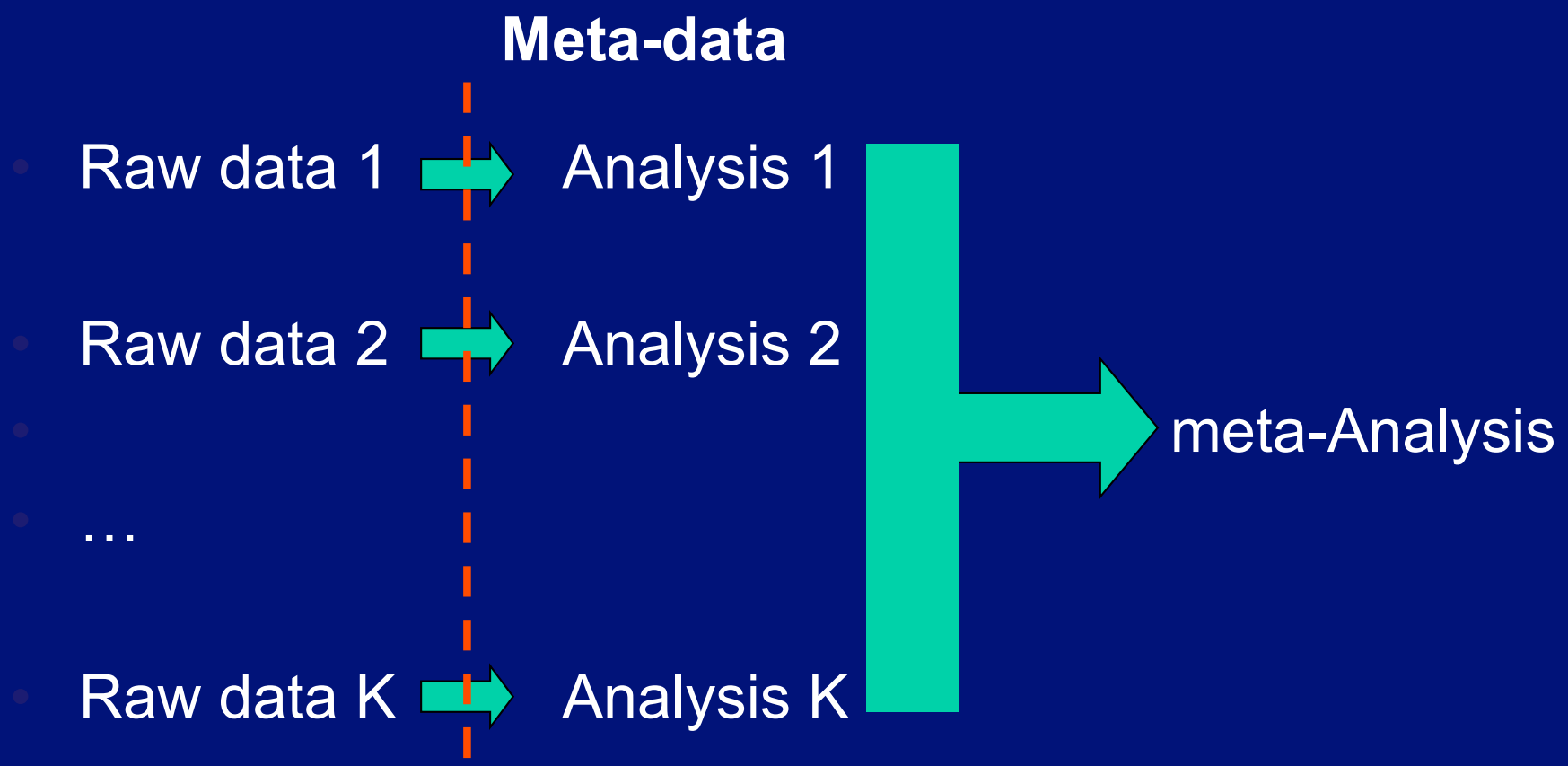
Meta-analysis of GWA Studies

Yurii Aulchenko

Outline

- Introduction: why meta-GWA?
- Methodology
  - Meta-analysis methods
  - Joint vs. meta-analysis
  - Random vs. fixed effects
  - Specific of analysis of individual study
- Technology: what to report for meta-GWA?

Outline of meta-GWAS



- **Avoid bias: all results reported (no selection on P-values, betas, etc.)**

Meta-data

Study	SNP	n	β	s.e.
1	rs355456	2640	0.11	0.032
2	rs355456	2370	0.08	0.041
3	rs355456	1310	-0.01	0.030
1	rs765865	2644	0.01	0.044
2	rs765865	2311	-0.03	0.037
3	rs765865	1312	0.02	0.055
1	rs485698	2583	0.001	0.029
2	rs485698	879	-0.12	0.033

Inverse variance meta-analysis

- Available from each of N studies
  - β_i (i=, …, N): effect estimates
  - s_i (i=, …, N) standard errors of the estimates

- Compute weights as $$w_i = \frac{1}{s_i^2}$$
- Pooled estimate of the effect is $$\beta = \frac{\sum_{i=1}^{N} w_i \beta_i}{\sum_{i=1}^{N} w_i}$$
- Pooled estimate of the standard error $$s^2 = \frac{1}{\sum_{i=1}^{N} w_i}$$
- Pooled Z-test value $$Z = \frac{\beta}{s} = \frac{\sum_{i=1}^{N} w_i \beta_i}{\sqrt{\sum_{i=1}^{N} w_i}}$$

Z-test based meta-analysis

- We do not quite believe that the effect estimates are consistent across studies because of differences in e.g. study design
- Use only “significance and sign” as characterized by study specific value of the *Z*-test (Z_i)
- Compute a study weight as the square root of the number of subjects used $w_i = \sqrt{n_i}$
- Pooled *Z*-score is $$Z = \frac{\sum_{i=0}^{N} w_i Z_i}{\sqrt{\sum_{i=0}^{N} w_i^2}}$$

Genomic Control with inverse variance

- K studies reporting reporting results for M SNPs. For particular study *k*, SNP *m*
  - effect estimate (β_{km}) and
  - its standard error (s_{km}) is reported
- Compute $T_{km}^2 = (\beta_{km} / s_{km})^2$
- For each study *k* estimate GC λ_k :
  - $\lambda_k = \text{Median}(T_{k1}^2, T_{k2}^2, \ldots T_{kM}^2) / 0.455$
- For each study *k* marker *m*, adjust standard error by λ_k :
  - $s'_{km}{}^2 = \lambda_k * s_{km}^2$
- Perform meta-analysis using corrected standard errors

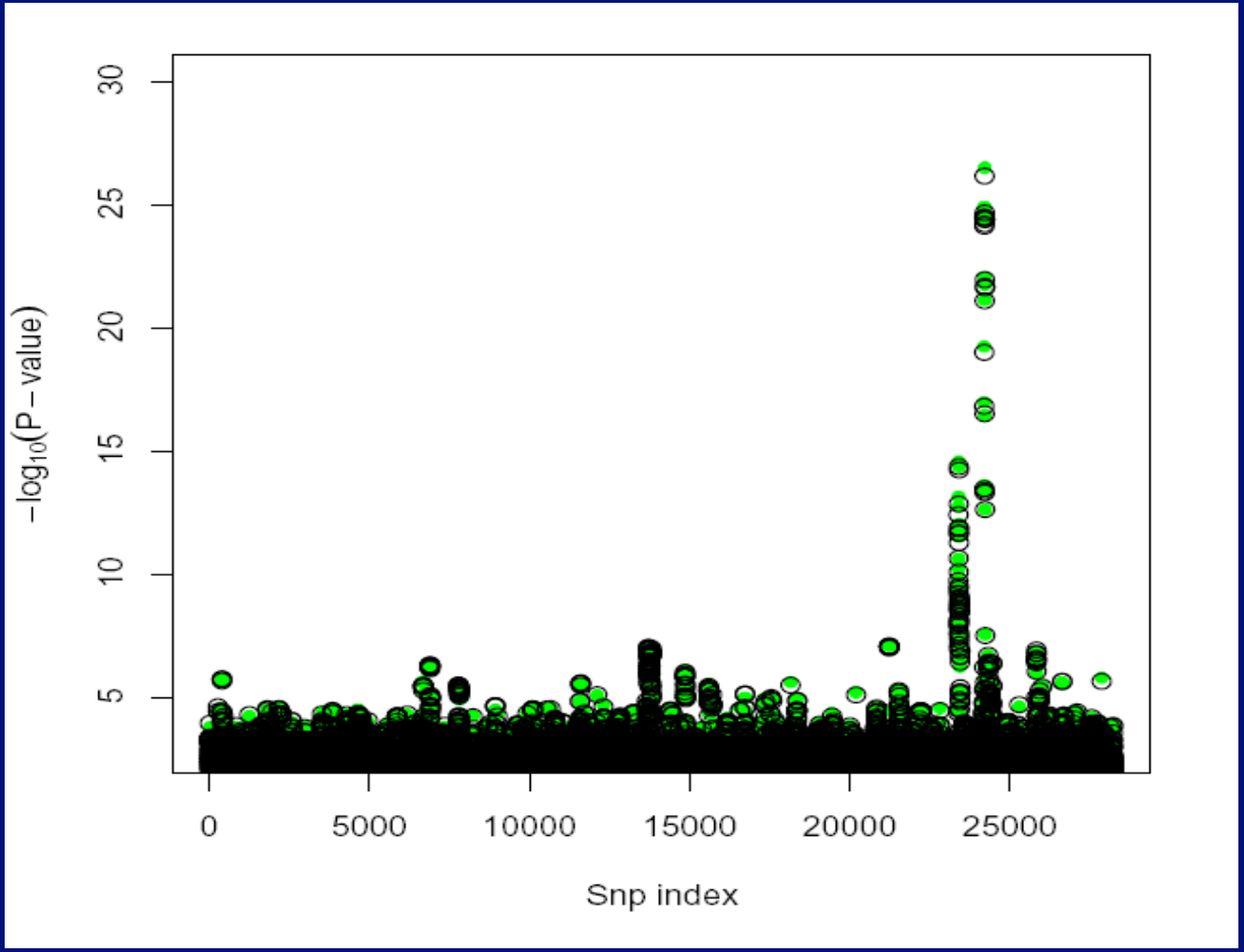
GC with Z-test meta-analysis

- K studies reporting reporting results for M SNPs. For particular study *k*, SNP *m*
  - *Z*-statistics value (Z_{km}) and
  - Number of subjects (n_{km}) is reported
- For each study *k* estimate GC λ_k :
  - $\lambda_k = \text{Median}(Z_{k1}^2, Z_{k2}^2, \ldots Z_{kM}^2) / 0.455$
- For each study *k* marker *m* re-compute Z scores
  - $Z'_{km} = Z_{km} / \text{Sqrt}(\lambda_k)$
- Perform meta-analysis using Z-score method

Outline

- Introduction
- Methodology
  - Meta-analysis methods
  - **Joint vs. meta-analysis**
  - Random vs. fixed effects
  - Specific of analysis of individual study
- Technology: what to report for meta-GWA?

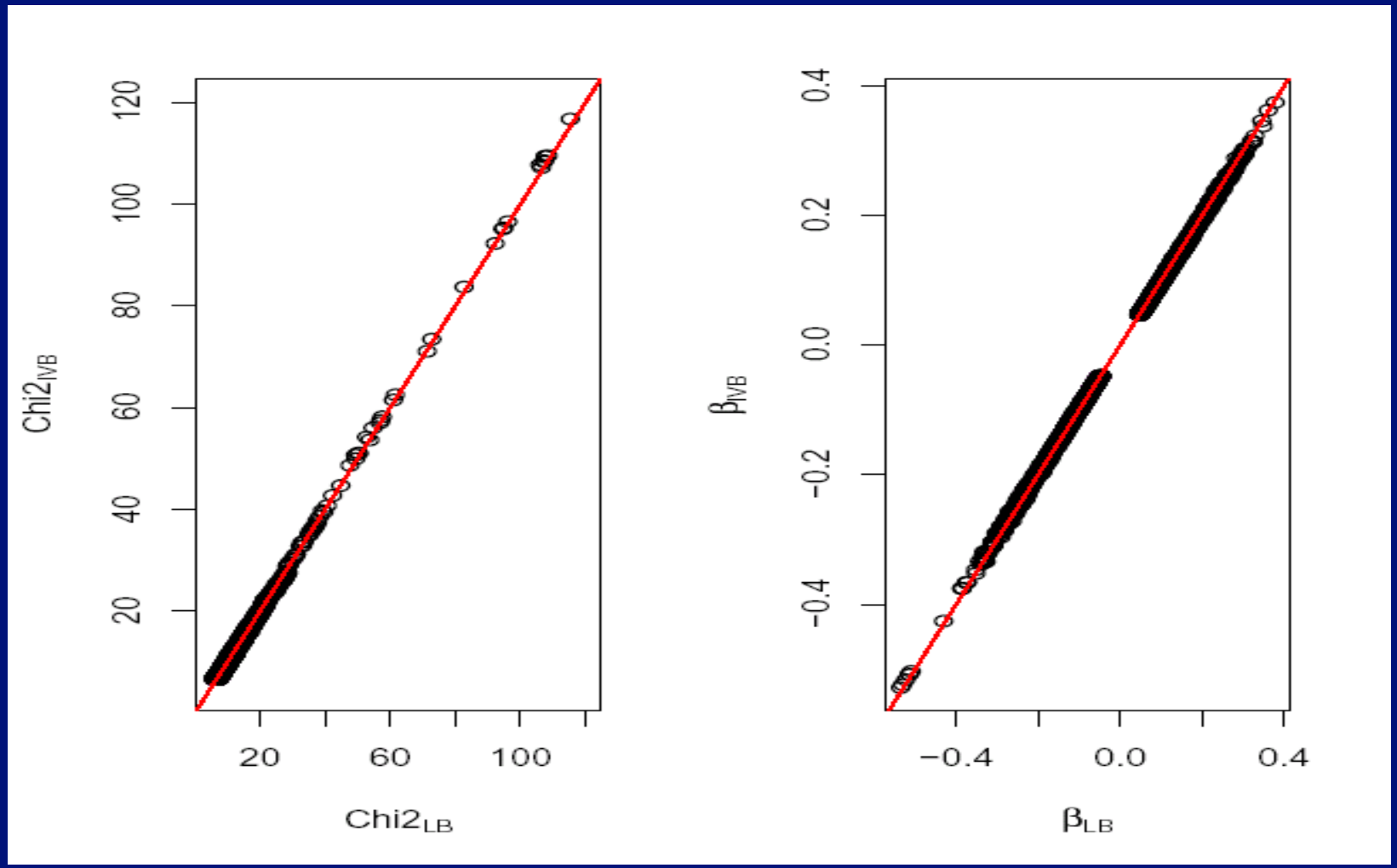
Joint less powerful than Meta?



- Green – *meta-analysis*
- Black – *joint analysis*

Yurii Aulchenko

Joint vs Meta: chi2's and beta's



Chi2: slope=0.999+/-2E-04

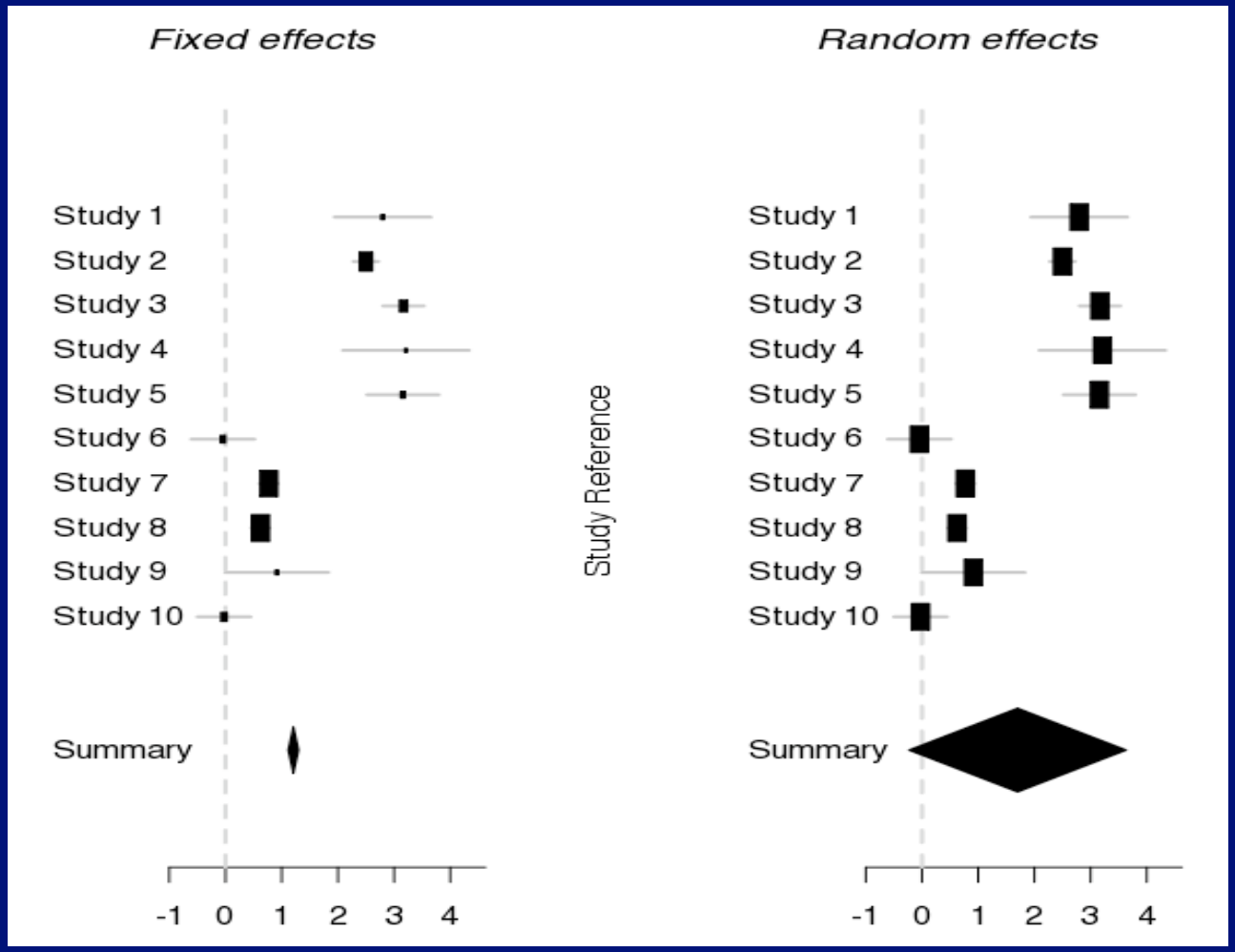
beta: slope=1.016+/-1E-04

Yurii Aulchenko

Outline

- Introduction
- Methodology
  - Meta-analysis methods
  - Joint vs. meta-analysis
  - **Random vs. fixed effects**
  - Specific of analysis of individual study
- Technology: what to report for meta-GWA?

Fixed vs Random



Yurii Aulchenko

Standard meta-analysis tests

Consider *k* studies with corresponding SNP effects β_i , i = 1, . . . , k

Fixed effect model null hypothesis: β1 = β2 = . . . = βk = 0

- Alternative: β1 = β2 = . . . = βk = β ≠ 0
- Random effect model assumes that β1, . . . , βk arises from a N (μ, σ2)
- Null hypothesis: μ = 0
- **Alternative: μ > 0 (you are not interested in that!)**
- Actually, for gene-discovery you are interested in alternative β ≠ 0 in one or more populations, and you do not care if these are heterogeneous!

Outline

- Introduction
- Methodology
  - Meta-analysis methods
  - Joint vs. meta-analysis
  - Random vs. fixed effects
  - **Specific of analysis of individual study**
- Technology: what to report for meta-GWA?

Analysis of individual study

- Meta-data: extract information the best way you can
- What is minimally needed for meta-analysis?
  - Number of people measured for the trait and the SNP genotype and Z-test values

    AND/OR
  - Unbiased effect estimates and standard errors
- <Slight> inflation of the test statistic can be corrected using Genomic Control in meta-GWAS

Best analysis providing the required characteristics!

QC for meta-GWAS

- Unit of meta-data: SNP characteristics
- Only exclude data points for which QC characteristics **can not** be reported in meta-data (and effectively used in metaGWA analysis)
- This usually translates to:
- (a) Identify and exclude “bad” samples
  - Use SNP and individual-level filters to identify “bad” samples
  - Exclude “bad” samples, but keep all SNPs
- (b) Perform GWA, report SNP-level QC characteristics (call rate, P-value HWE, AF, etc.)

Trait's distribution

- Significance derived based on effect estimate and standard error (e.g. Z-test) is correct
  - when number of measurements is very large

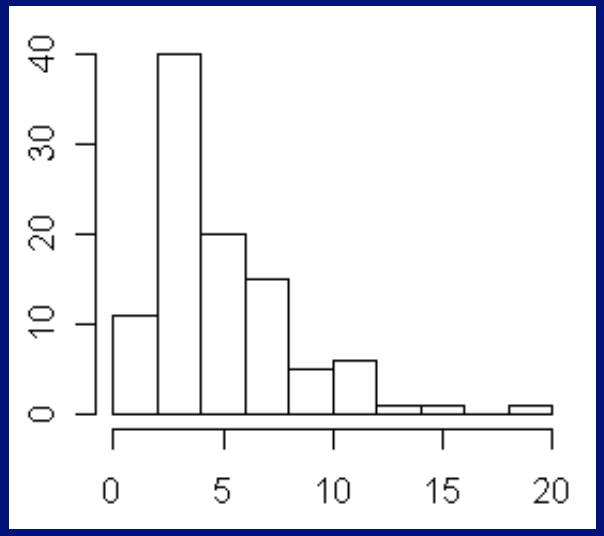
    and/or

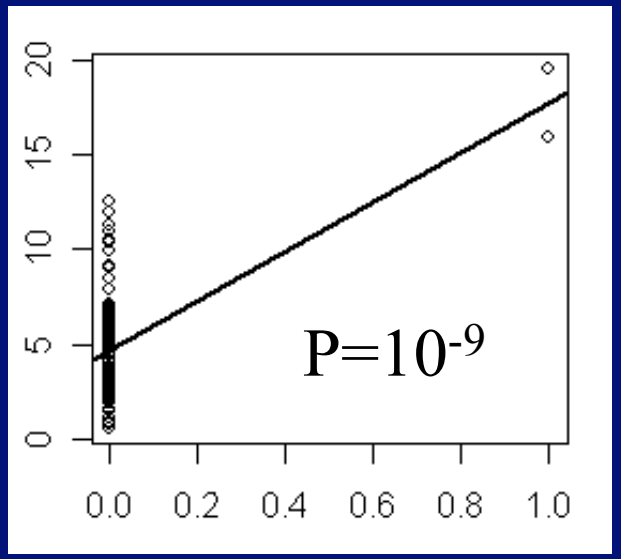
  - trait's residuals are distributed normally

Outliers generate false positives in individual GWAS

- (a) Presence of outliers
- (b) Small number of people
- (c) Rare polymorphisms

- => False-positive association





Yurii Aulchenko

Solution for individual study

- Trait's transformation:
  - Log-transformation: y' = log y
  - Square root transformation: y' = sqrt(y)
  - Box-Cox transformation $$y^{(\lambda)} = \begin{cases} (y^\lambda - 1)/\lambda, & \text{if } \lambda \neq 0 \\ \log y, & \text{if } \lambda = 0 \end{cases}$$
  - Rank-transformation to Normal
    - Ranks projected to Normal
    - Guarantees perfect fit to Normal in absence of ties

- Empirical procedures: they do not rely on normality assumption (but can not use in meta unless some new methods are developed)

Meta-analysis: large numbers are good!

- The larger are the numbers, the more non-normality you can afford
- If the number of cohorts and total number of subjects studied in meta-analysis is really large, say
  - Each study > 1,000 subjects
  - In total, > 20,000 subjects
  - In total, >10 cohorts
- Then there is little problem in (moderate) non-normality of the trait distribution
- False positives due to combination of rare allele and non-normality can be easily detected: you will see a huge effect coming from a single study
- … thus checking heterogeneity may be a good idea
- … at least for your “top” hits

Outline

- Introduction
- Methodology
  - Meta-analysis methods
  - Joint vs. meta-analysis
  - Random vs. fixed effects
  - Specific of analysis of individual study
- **Technology: what to report for meta-GWA?**

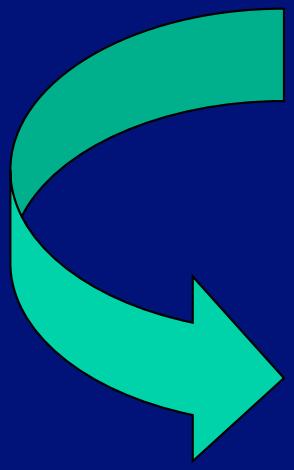
Meta-data: what to report?

- For meta-analysis one needs
  - Effect estimate
  - Estimate's standard error
  - Number of people measured for both trait and SNP
- Suggested format 1:

Study	n	β	s.e.	P
1	2644	0.11	0.032	0.0005
2	2311	0.08	0.030	0.0003
3	2375	-0.12	0.028	0.0001
Meta	7330	0.01	0.013	0.45

Reference & Effective alleles

Study	Ref.	Eff.	n	β	s.e.
1	A	G	2644	0.11	0.032
2	A	G	2311	0.08	0.030
3	**G**	**A**	**2375**	**-0.12**	**0.028**



Study	Ref.	Eff.	n	β	s.e.	P
1	A	G	2644	0.11	0.032	0.0005
2	A	G	2311	0.08	0.030	0.0003
3	A	G	2375	**+0.12**	0.028	0.0001
Meta	A	G	7330	0.10	0.013	10^{-9}

Yurii Aulchenko

Suggested format 2

- Effect estimates (sign of Z) should be reported for the same allele (A/T/G/C) across all studies
- …or individual study results should provide enough information about reference and effective allele
- E.g. report coding A_1A_2 where A_1 is always reference
-

Study	Ref.	Eff.	n	β	s.e.	P
1	A	G	2644	0.11	0.032	0.0005
2	A	G	2311	0.08	0.030	0.0003
3	A	G	2375	0.12	0.028	0.0001
Meta	A	G	7330	0.10	0.013	10^{-9}

No association again?

Study	Ref.	Eff.	n	β	s.e.	P
1	A	T	2644	0.11	0.032	0.0005
2	A	T	2311	0.08	0.030	0.0003
3	A	T	2375	-0.12	0.028	0.0001
Meta	A	T	7330	0.01	0.013	0.45

Specifics of A/T and G/C SNPs

Study	Ref.	Eff.	Strand	n	β	s.e.	
1	A	T	+	2644	0.11	0.032	
2	A	T	+	2311	0.08	0.030	
3	**A**	**T**	**-**	2375	**-0.12**	0.028	
1	A	T	+	2644	0.11	0.032	
2	A	T	+	2311	0.08	0.030	
3	**T**	**A**	+	2375	**-0.12**	0.028	
1	A	T	+	2644	0.11	0.032	
2	A	T	+	2311	0.08	0.030	
3	A	T	+	2375	**0.12**	0.028	
Meta	A	T	+	7330	0.10	0.013	10^{-9}

Yurii Aulchenko

Minimal suggested format

- From analysis:
  - SNP name
  - Reference allele
  - Effective allele
  - Strand
  - Genomic build
  - Number of people with trait & genotype
  - Effect estimate
  - Standard error of the effect estimate
- From QC:
  - Call rate
  - P-value HWE
  - Effective allele frequency

Software

- MetABEL
  - by Yurii Aulchenko & Maksim Struchalin
  - Inverse variance method
  - http://mga.bionet.nsc.ru/~yurii/ABEL/
- METAL
  - by Goncalo Abecasis
  - Z-score method
  - Inverse variance method
  - http://www.sph.umich.edu/csg/abecasis/Metal/index.html
- R library “rmeta”
  - by Thomas Lamley
  - General wide-scope meta-analysis library
  - Implements multiple methods and great forest-plot graphics
  - Not quite suited for meta-GWAS

Conclusions

- Meta-analysis of GWAS is a powerful tool to detect common loci, even of small effect
- Meta is almost as powerful as joint analysis
- Use fixed effects models for meta-GWA; new tests are coming
- Large numbers are good
- Bio-informatics matters: mind the build, strand, and coding