GWAS Meta-Analysis

Sarah Medland

The 2023 International Statistical Genetic Workshop


International
Women's
Day

Honouring a woman…

Marjolein Kriek

3 language

Article Talk Read Edit View h

From Wikipedia, the free encyclopedia

Marjolein Kriek (born 22 November 1973)[1] is a Dutch clinical geneticist at the Leiden University Medical Center. In 2008, at age 34, she became the first woman and probably the first European to have her total DNA genome sequenced.[2][3][4][5]

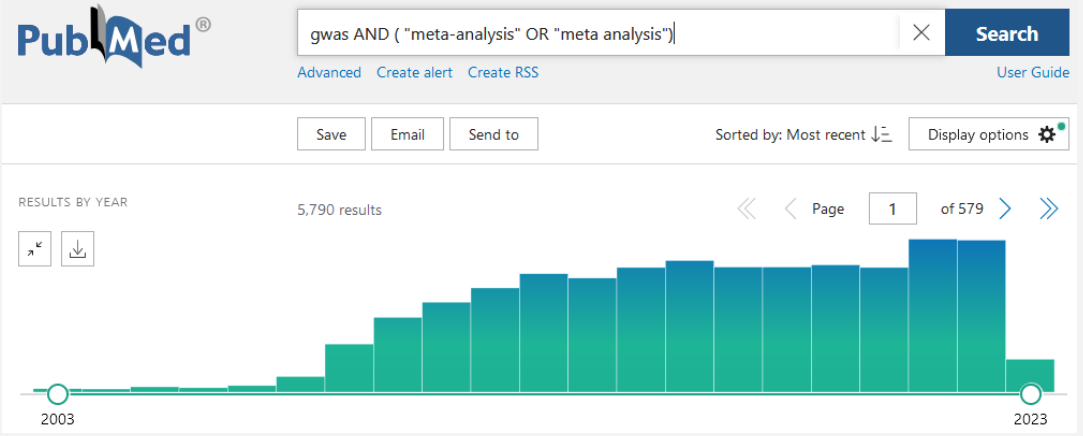
Sequencing project [edit]

Leiden University announced the completion of the nine-month-long[4] sequencing of Kriek's genome on 26 May 2008,[3][5] though the results of the study were published later. The study was initiated by Gert-Jan van Ommen of the LUMC team and director of the Center for Medical Systems Biology (CMSB), to gain insight in X-chromosome variability. The data set contained significant redundancy, as each base pair was sampled an average of seven to eight times.[3][5] The sequencing has been performed in close collaboration with a spin-off company, 'ServiceXS'. [6]

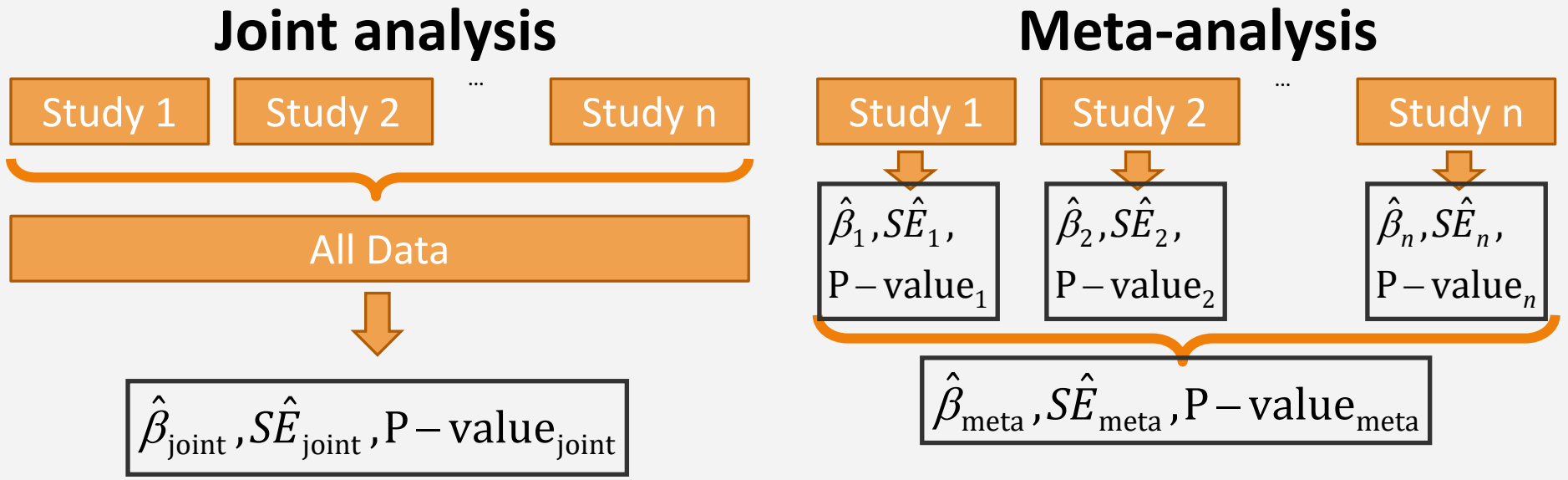
At the time, Kriek was one of five or six people to have their entire genome sequenced, the others being James D. Watson, Craig Venter, two Yoruba-African men[2][3] and possibly Dan Stoicescu of Switzerland.[7]

Combining data across studies

- Aims:
  - Estimate the overall, or combined effect
  - Explore differences between cohorts – heterogeneity
  - Improve power
  - Replicate effects



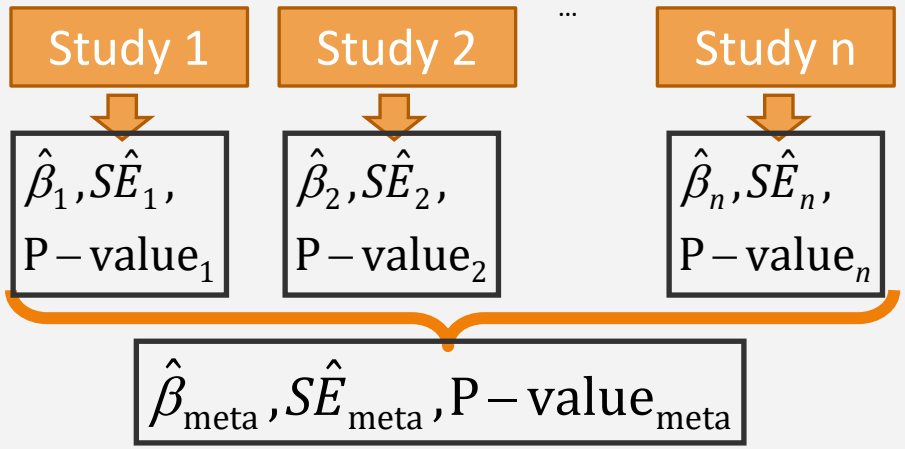
Joint vs Meta-analysis



For common variants, joint and meta-analysis have similar power (Lin & Zeng, *Genet. Epidemiol.*, 2010)

How we use meta-analysis in GWAS

- Commissioned analyses rather than MA of previously published findings
  - Analysis protocol
    - Imputation reference
    - Phenotypic definition
    - Covariates to be included
    - Population stratification
    - Analyses to be run
    - Output format



Types of meta-analysis – Fixed Effect

- Each SNP has a “true effect size” on the trait.
- This effect is shared by all cohorts.
- The observed effect sizes in the different cohorts will be distributed around the “true effect size” with a variance that depends on the precision of the different cohorts

Types of meta-analysis – Fixed Effect

- We weight each cohorts effect size by it's precision
- Error in our estimate is due to random error within studies
- The combined effect = the meta-analytic estimate

Types of meta-analysis – Random Effect

- The true effect for a SNP varies between cohorts.
- The studies included in the meta-analysis are assumed to be a random sample reflecting the distribution of true effects

Types of meta-analysis – Random Effect

- Error in our estimate is due to random error within AND between studies
- Weights reflect these two sources of error and are less dependent on sample size
- The mean effect in this distribution = the meta-analytic estimate

Types of meta-analysis – Random Effect

- Traditionally, null hypothesis for RE model is that the mean of the effects is 0
  - lower power than FE
- Correct null hypothesis for GWAS is that all effects are 0
  - Modification proposed by Han & Eskin 2011 tests the null hypothesis of exactly 0 effect in every study – RE2
  - Implemented in Meta (https://mathgen.stats.ox.ac.uk/genetics_software/meta/meta.html)

Types of meta-analysis – Others

- Bayesian partition model
  - Cohorts are grouped into clusters. Effect is assumed be the same within but different between clusters.
  - MANTRA (Meta-ANalysis of Transethnic Association studies) - Andrew Morris - Genet Epidemiol. 2011; 35(8): 809–822. PMCID: PMC3460225
  - SMetABF - Sun J et al. - PLOS Computational Biology 2022; 18(3): e1009948.

Types of meta-analysis – Others

- Multivariate GWAS packages
  - MTAG, GenomicSEM (later today)
- Continuous & Binary meta-analysis
  - Demontis, Walters et al Nat Genet. 2019; 51(1): 63–75. PMCID: PMC6481311 (Sup information)

Setting up a meta-analysis

- Start with an analysis plan
  - Decide on QC of input and analytic approach
  - Define your primary analyses
  - Define any secondary analyses – sensitivity, populations
  - Map out intended follow-ups
  - Define replication
  - Preregistration or public posting is strongly encouraged
  - Allow a lot more time than you think you will need

Software for running meta-analysis

- Important considerations
  - Types of analyses the program can run
  - QC requirements
  - Strand flipping
    - Allele frequency tracking
  - What you want to do your output
    - Genomic control vs LDscore regression
    - Beta & SE vs Z

Software for running meta-analysis

- Software
  - METAL
  - ? GWAMA (Magi & Morris)
  - ? Meta(Han & Eskin)
  - R & Stata packages

QC of cohort level data

- Variant naming
- Allele Frequency - MAF .5 or 1%
- Imputation accuracy (r^2) - Typically .6 (.8 if hard calls were analysed)
- Plots - Manhattan, QQ, P-Z
- MAF compared to reference - Strand
- Lambda calculation - Checking for confounding
- Packages are available to help with this i.e. EasyQC

Choosing an analytic approach

- Metal
  - Fixed effect analysis – Inverse variance weighted
    - Requires: beta, SE, alleles
    - Outputs: beta, SE, p, N, heterogeneity, MAF

	Inverse variance based
Inputs	β_i- effect size estimate for study *i*
	se_i - standard error for study *i*
Intermediate Statistics	$w_i = 1/SE_i^2$
	$se = \sqrt{1/\sum_i w_i}$
	$\beta = \sum_i \beta_i w_i / \sum_i w_i$
Overall *Z*-Score	$Z = \beta / SE$
Overall *P*-value	

Choosing an analytic approach

- Metal
  - Fixed effect analysis – Sample size weighted
    - Requires: direction, P, N , alleles
    - Outputs: Z, p, N, heterogeneity, MAF
    - Sample overlap correction

	Sample size based
Inputs	N_i - sample size for study i
	P_i–P-value for study i
	Δ_i - direction of effect for study i
Intermediate Statistics	$Z_i = \Phi^{-1}(P_i/2) * \text{sign}(\Delta_i)$
	$w_i = \sqrt{N_i}$
Overall Z-Score	$Z = \frac{\sum_i Z_i w_i}{\sqrt{\sum_i w_i^2}}$
Overall P-value	$P=2\Phi(\lvert -Z \rvert)$

Choosing an analytic approach

- Random Metal
  - Requires: beta, SE, p, N, alleles
  - Outputs: Both FE and RE results beta, SE, p, heterogeneity, tau^2

Why would you chose N weighted or RE?

- Inverse variance FE meta-analysis are sensitive to deviations in scaling between studies
- N weighted FE meta-analyses are less sensitive to this
- RE meta-analyses are more appropriate for situations where the effect size differs between cohorts

Practical

cp /faculty/sarah/2023/Meta

The link to the Qualtrics is in the file called *Link_to_workbook.txt*

Questions?