Meta-analysis

in genetic association studies

Marleen H.M. de Moor

Department of Biological Psychology

VU University Amsterdam

mhm.de.moor@psy.vu.nl

vrije Universiteit amsterdam

Outline of this session

- When and why use meta-analysis in genetic studies
- Theoretical background of meta-analysis

- Introduction of the example dataset (Dorret Boomsma)
- Some background information on genotyping (Nick Martin)
- Practical 1: Run PLINK, compute needed parameters
- Practical 2: Compute meta-analysis results by hand
- Practical 3: Run meta-analysis using METAL

- Some example studies

Genetic association studies

- Meta-analysis of candidate gene studies
  → Adhoc analysis of published results
  → Replication

- Meta-analysis of GWA studies
  → Replication of most significant hits from discovery sample
  → International consortia

Why not combine samples for GWAS?

- Privacy
- Population stratification

Goal of meta-analysis

- Quantitative synthesis of results from different samples/studies
- Larger N -> More power!
- Done by pooling:
  - Genetic effect of a SNP on a phenotype
  - P-value of the association test

Types of meta-analysis

- Pooling effect estimates

  - What is ‘true’ effect in population?

  ➡ Inverse variance weighted method

  ➡ Fixed vs. Random models

- Pooling p-values

  - Is association significant?

  ➡ Pooled z-score method

Pooling effect estimates

Phenotype	Analysis	Effect estimate
Case-control	Chi-square test	$OR = e^{\beta}$ $\beta = \ln(OR)$
Case-control	Logistic regression	$OR = e^{\beta}$ $\beta = \ln(OR)$
Quantitative trait	Linear regression	β

Inverse variance weighted method

Fixed models

Assumptions:

- There is one underlying ‘true’ effect
- All deviations of sample effects from the ‘true’ effect are due to chance

Prerequisites:

- Same scale must be used across samples!
- Same reference allele on same strand!

Inverse variance weighted method

Computing pooled effect:

$$\text{Pooled effect} = \frac{\text{Sum (weights * effect estimates)}}{\text{Sum (weights)}}$$

$$\beta_{pooled} = \frac{\sum_{i=1}^{N} (w_i * \beta_i)}{\sum_{i=1}^{N} (w_i)}$$

$$w_i = \frac{1}{\text{var}(\beta_i)}$$

$$\text{var}(\beta_i) = se\ (\beta_i)^2$$

$$se\ (\beta_i) = \frac{SD_i}{\sqrt{n_i}}$$

i=1...N samples

Inverse variance weighted method

Computing pooled standard error:

$$\text{Pooled standard error} = \text{Square root}\left(\frac{1}{\text{Sum (weights)}}\right)$$

$$se_{pooled} = \sqrt{\frac{1}{\sum_{i=1}^{N}(w_i)}}$$

$$w_i = \frac{1}{\text{var}(\beta_i)}$$

$$\text{var}(\beta_i) = se\ (\beta_i)^2$$

Inverse variance weighted method

Computing 95% confidence interval:

Pooled effect +/- 1.96 * pooled standard error

Inverse variance weighted method

Computing test statistic:

$$\chi^2_{df=1} = \frac{\beta_{pooled}^{\ 2}}{se_{pooled}^{\ 2}} = \frac{(\sum_{i=1}^{N} w_i * \beta_i)^2}{\sum_{i=1}^{N} w_i}$$

$$z = \frac{\beta_{pooled}}{se_{pooled}} = \frac{\sum_{i=1}^{N} w_i * \beta_i}{\sqrt{\sum_{i=1}^{N} w_i}}$$

Look up or compute the associated p-value

Inverse variance weighted method

Computing test statistic:

$$\chi^2_{df=1} = \frac{\beta_{pooled}^{\ 2}}{se_{pooled}^{\ 2}} = \frac{(\sum_{i=1}^{N} w_i * \beta_i)^2}{\sum_{i=1}^{N} w_i}$$

$$z = \frac{\beta_{pooled}}{se_{pooled}} = \frac{\sum_{i=1}^{N} w_i * \beta_i}{\sqrt{\sum_{i=1}^{N} w_i}}$$

P=0.05 → χ^2=3.84
Z=1.96

P=0.001 → χ^2=10.83
Z=3.29

Do assumptions of fixed model hold?

Test of homogeneity

- Cochran's Q statistic

$$Q = \sum_{i=1}^{N} w_i (\beta_i - \beta_{pooled})^2$$

χ^2-distributed with df=k-1

k=Number of samples

α=0.10

➡ With small sample size, low power!!

Quantify heterogeneity

I^2 statistic

$$I^2 = \frac{Q - (k-1)}{Q} * 100$$

Range 0-100%

I^2 > 50%: Large heterogeneity

I^2 > 75%: Very large heterogeneity

Causes of heterogeneity

Possible causes related to bias in samples:

- Differential selection of cases and controls
- Poor genotyping
- Poor genotype data cleaning
- Different SNP platforms
  (imputed vs. observed SNPs)
- Poor/differential phenotyping
- Population stratification

Causes of heterogeneity

Possible causes related to genuine differences across samples:

- Different LD structure across populations
  (truly associated SNP vs. tested SNP)
- Different correlations of phenotypes across populations
  (truly associated phenotype vs. tested phenotype)

Solution to heterogeneity

Random effects model

Assumptions:

- Assume that there is one underlying *distribution* of effects
- Normal distribution of effects

Random effects models

Used if:

- Large differences across samples
  (expected or observed)
- Same scale is used across samples

But:

- Number of samples should be sufficiently large

Random effects models

Estimate between study variance

(DerSimonian Laird estimator)

$$\tau^2 = \frac{Q - (k-1)}{\sum_{i-1}^{N} w_i - \frac{\sum w_i^2}{\sum w_i}}$$

- τ^2 is incorporated in the weights
- Random effects model are more conservative (larger se)

Z-score pooling method

Good to use if:

- Large differences across samples
- Number of samples is small
- Same scale is NOT used across samples

Z-score pooling method

Computing pooled z-score:

$$\text{Pooled z-score} = \frac{\text{Sum (weights * z-scores)}}{\text{Sum (weights)}}$$

Individual z-scores computed by:

- Converting individual p-values into z-scores
- Taking the sign of the effects into account

Z-score pooling method

Computing pooled z-score:

$$\text{Pooled z-score} = \frac{\text{Sum (weights * z-scores)}}{\text{Sum (weights)}}$$

$$z_{pooled} = \frac{\sum_{i=1}^{N}(w_i * z_i)}{\sqrt{\sum_{i=1}^{N}(w_i^2)}}$$

$$w_i = \sqrt{n_i}$$

Z-score pooling method

When combining samples using different platforms

- Incorporate uncertainty information in weights

MACH: r-squared
IMPUTE: proper_info

Introduction of GWA study for MDD

- Data from 7 replication samples used for the practicals

- Dorret Boomsma: introduction of the study
- Nick Martin: background of genotyping

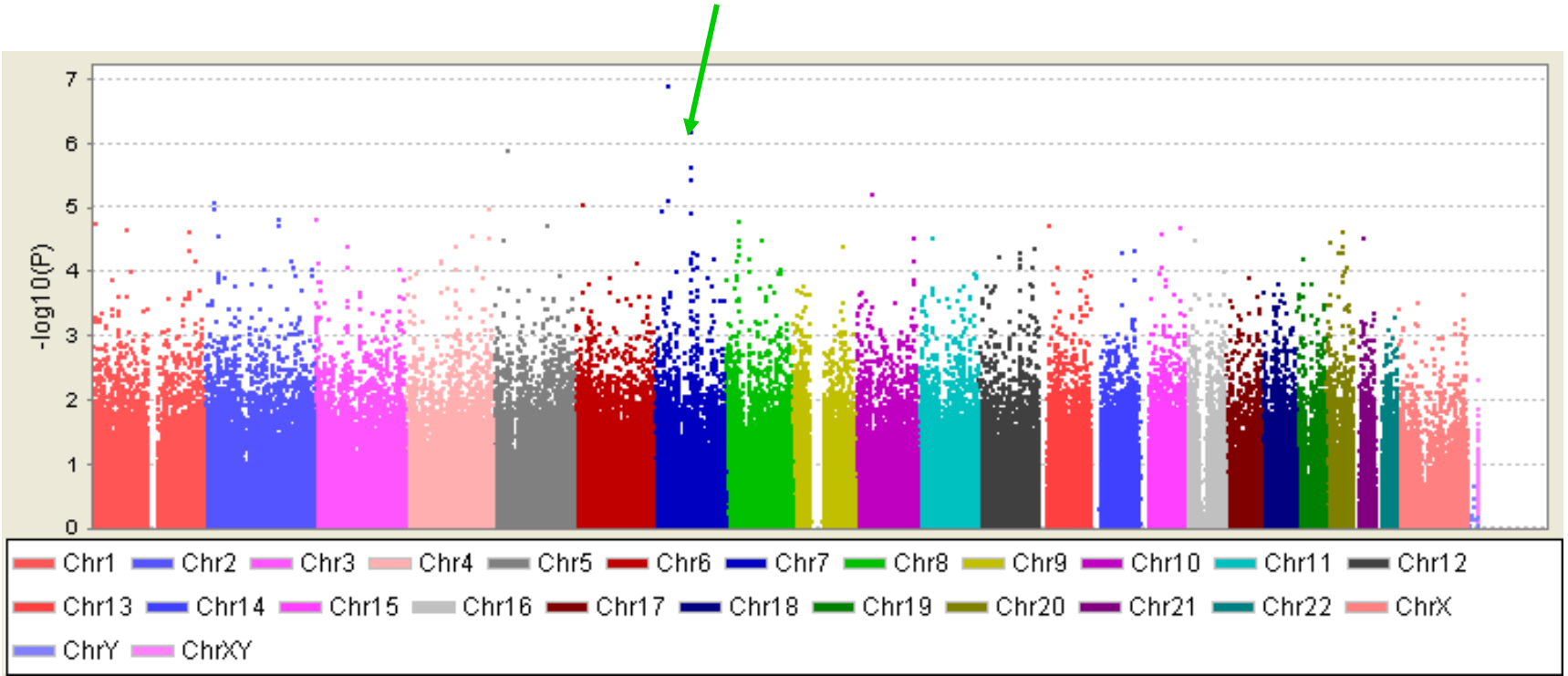
GWA of MDD

Dorret Boomsma
Nick Martin

Boulder, 2009

GWA-GAIN: MDD in discovery sample

Eleven of the top 200 signals localized to a 167 kb region overlapping the gene piccolo (*PCLO*, whose protein product localizes to the cytomatrix of the presynaptic active zone and plays an important role in monoaminergic neurotransmission in the brain)



Molecular Psychiatry (2008), 1–17

www.nature.com/mp

IMMEDIATE COMMUNICATION

Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo

PF Sullivan[1], EJC de Geus[2], G Willemsen[2], MR James[3], JH Smit[4], T Zandbelt[4], V Arolt[5], BT Baune[6], D Blackwood[7], S Cichon[8], WL Coventry[9], K Domschke[5], A Farmer[10], M Fava[11], SD Gordon[3], Q He[1], A Heath[12], P Heutink[4], F Holsboer[13], WJ Hoogendijk[4], JJ Hottenga[2], Y Hu[1], M Kohli[13], D Lin[1], S Lucae[13], DJ MacIntyre[14], W Maier[8], KA McGhee[7], P McGuffin[10], G Montgomery[3], WJ Muir[7], W Nolen[15], MM Nöthen[8], RH Perlis[11], K Pirlo[10], D Posthuma[2], M Rietschel[16], P Rizzu[4], A Schosser[10], AB Smit[2], JW Smoller[11], J-Y Tzeng[17], R van Dyck[4], M Verhage[2], FG Zitman[18], NG Martin[3], NR Wray[3], DI Boomsma[2,19] and BWJH Penninx[4,19]

[1]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA; [2]VU University Amsterdam; [3]Queensland Institute for Medical Research; [4]VU University Medical Center Amsterdam; [5]University of Münster; [6]James Cook University Queensland; [7]University of Edinburgh; [8]University of Bonn; [9]University of New England; [10]Institute of Psychiatry; [11]Harvard Medical School; [12]Washington University, St. Louis; [13]Max-Planck Institute of Psychiatry; [14]Royal Edinburgh Hospital; [15]University Medical Center Groningen; [16]University of Heidelberg; [17]North Carolina State University and [18]Leiden University Medical Center

Major depressive disorder (MDD) is a common complex trait with enormous public health significance. As part of the Genetic Association Information Network initiative of the US Foundation for the National Institutes of Health, we conducted a genome-wide association study of 435 291 single nucleotide polymorphisms (SNPs) genotyped in 1738 MDD cases and 1802 controls selected to be at low liability for MDD. Of the top 200, 11 signals localized to a 167 kb region overlapping the gene piccolo (*PCLO*, whose protein product localizes to the

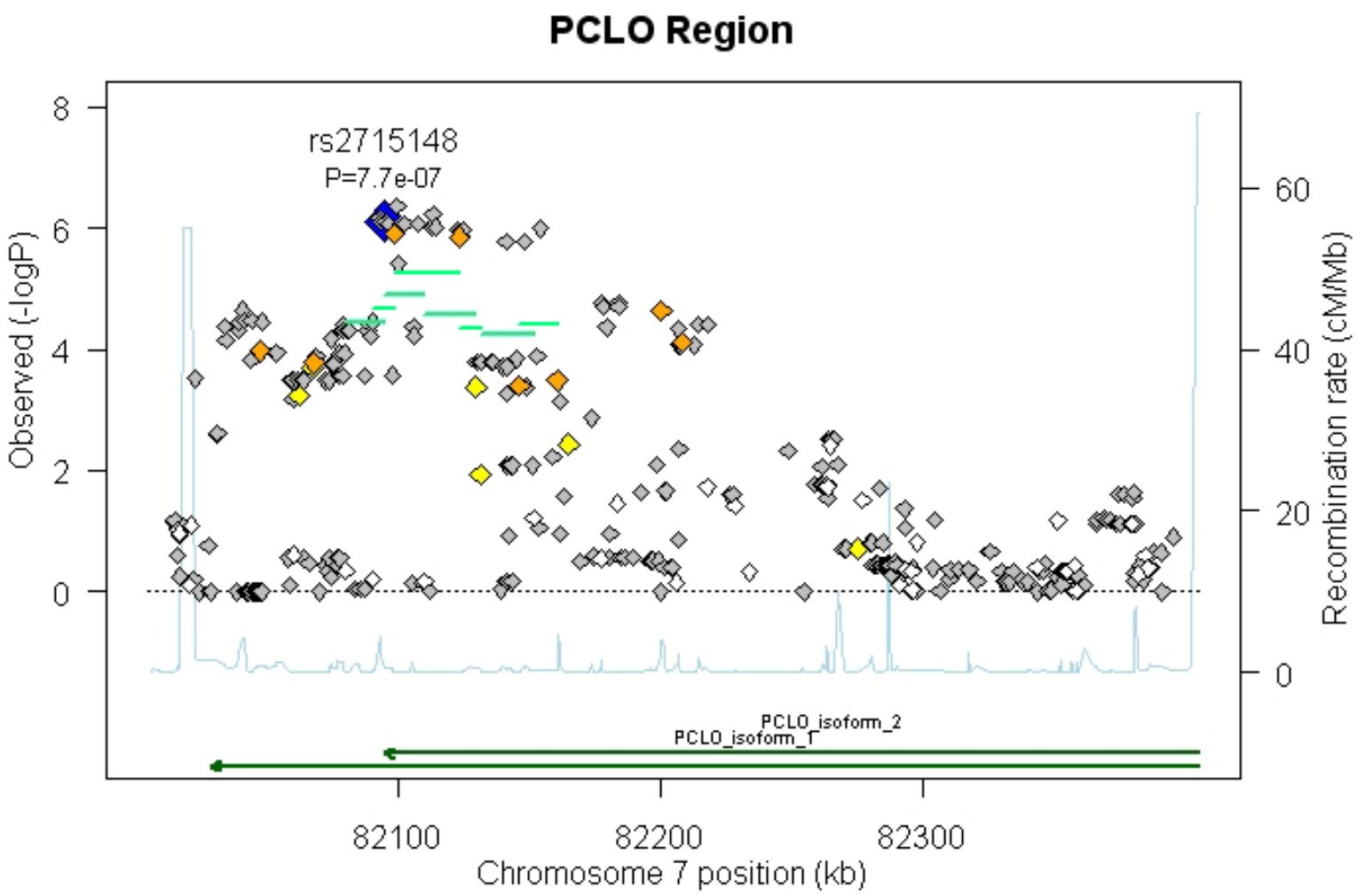
Discovery Sample: GAIN: Background

- Case-control study of MDD
- Dutch subjects from the Netherlands Twin Register (NTR).
- Participants in the Netherlands Study of Depression and Anxiety (NESDA).
- 1,862 participants with a diagnosis of MDD
- 1,857 controls at low liability for MDD
- Perlegen 4 chip platform of 600K SNPs that were selected as tags for individuals of European ancestry.
- 75% SNPs pass QC.

Result

- 11 of top 200 p-vals in piccolo, *PCLO*
- Couldn't make it go away – not due to any bias we detected
  - Not population stratification
  - Not due to very high LD
  - Not due to genotyping artifact
  - Not due to "funny" controls
  - Not resolvable with haplotypes or imputation
  - No known CNV

Ch	Allele	Gene	SLEP ‡	Rank	OR (CI)	P-empirical
2	A/G			10	1.26 (1.14-1.39)	0.000014
2	A/G			7	1.26 (1.14-1.40)	0.000011
2	T/C	*ALK*	CNV, mutated in colon CA	12	1.34 (1.17-1.54)	0.000020
2	A/G		Near CNV	13	1.31 (1.16-1.49)	0.000027
6	C/T		SCZ linkage meta-analsyis	6	0.79 (0.71-0.88)	0.000007
7	C/T		MDD linkage	1	1.27 (1.16-1.40)	0.000002
7	G/C		MDD linkage	14	1.22 (1.12-1.35)	0.000038
7	A/C	*PCLO*	BIP GWAS	2	0.79 (0.72-0.87)	0.000003
7	C/A	*PCLO*	BIP GWAS	3	1.26 (1.15-1.39)	0.000002
7	G/T	*PCLO*	BIP GWAS	4	1.25 (1.14-1.38)	0.000003
7	A/T	*PCLO*		8	0.81 (0.74-0.89)	0.000007
8	T/C		BIP GWAS	15	0.76 (0.66-0.86)	0.000036
10	A/G			5	0.80 (0.73-0.88)	0.000013
12	T/C	*TMEM16F*	MDD linkage	11	0.78 (0.69-0.87)	0.000023
15	C/T	*SHC4*		9	0.72 (0.62-0.84)	0.000009



Expressed in brain, localizes to neuronal presynaptic active zone

Role in monoamine neurotransmission, a venerable hypothesis of MDD (Schildkraut 1965) ; rs2522833 (#3): nsSNP (ala-4814-ser, MAF 0.45), near *PCLO* C2A calcium binding domain, predicted major effect on PCLO

PCLO Replication

- Independent samples: N=11,972 (6,079 cases, 5,893 controls)
- 6 samples, similar to NESDA cases & NTR controls
  - Cases: EUR adults, similar inclusion & exclusion criteria. Most clinical, one population.
  - Controls: EUR adults, MDD removed, pop-based
- **Genotyping: 30 SNPs in 5 & 2 SNPs in 1 sample. Initial sample re-genotyped (agreement 0.9987). Genotyping done at QIMR for majority of the samples**
- Power.
  - Assume: log-additive, MDD risk 0.15, MAF=0.45 (similar to rs2522833), GRR 1.14 ("shrunk" from observed GRR of 1.26 for rs2522833), & two-tailed type 1 error rate of 0.00167
  - 97.2% for 2 SNPs in all samples (N=11,972)
  - 90.4% for the remaining SNPs (N=9,278)

NARSAD
The World's Leading Charity Dedicated to Mental Health Research

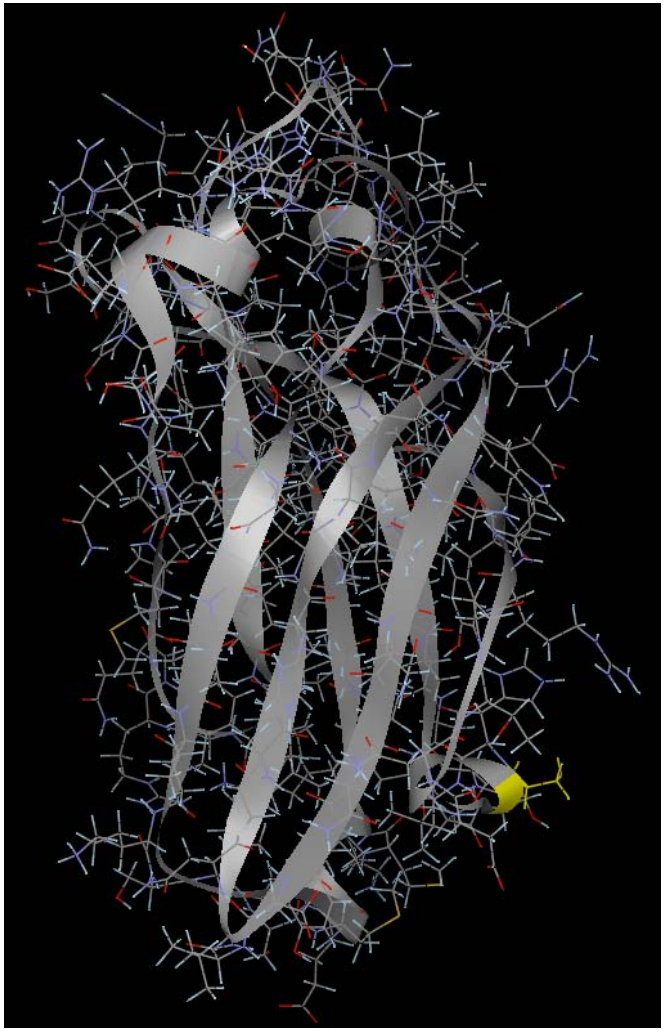
Replication samples for PCLO

Feature	NESDA-NTR	Bonn-Mannheim	Münster	MPIP	DeCC	STAR*D	QIMR	UEDIN
Type	Original	Replicate	Replicate	Replicate	Replicate	Replicate	Replicate	Replicate
Location	NL	Germany	Germany	Germany	UK	US	Australia	UK
Design	Cohort	Cross-sectional	Cross-sectional	Cross-sectional	Cross-sectional	Clinical trial	Cohort	Cross-sectional
SNPs	30	30	30	30	3	30	30	30

Comparisons								
N Case	1866	777	452	940	1403	1187	966	354
N Controls	1792	1260	0	967	1291	864	1039	472

We will analyze the replication data from 3 out of 30 SNPs in the PCLO region in independent samples (6,079 MDD independent cases and 5,893 controls). Data have been permuted (maintaining case – control status)

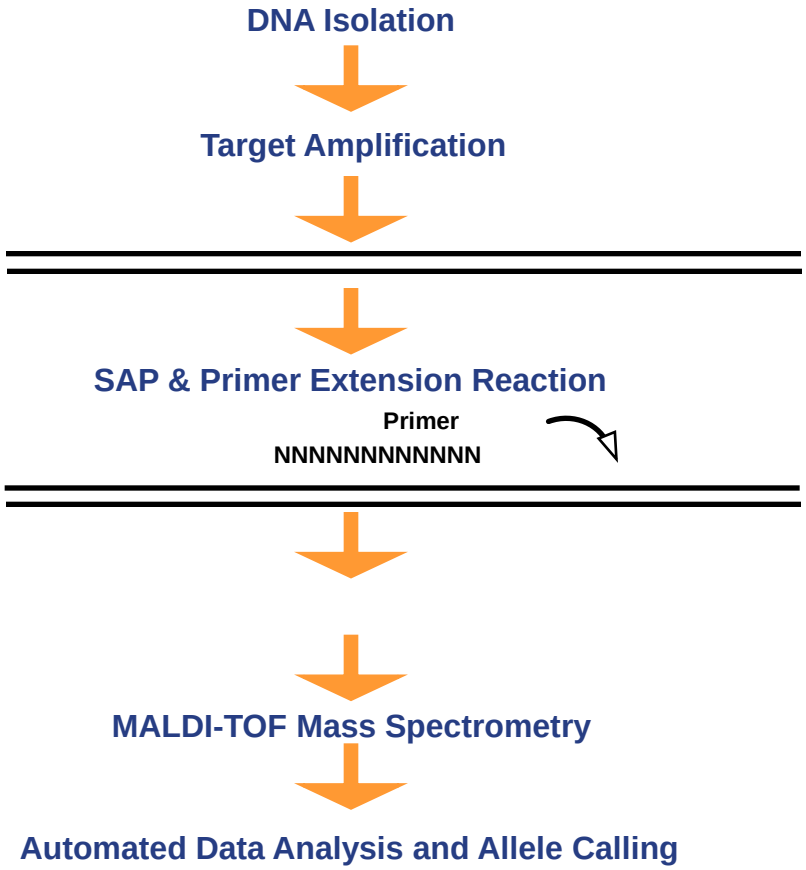
Replication



No SNP exceeded the replication significance threshold.

However, there was heterogeneity in the replication samples, and analysis of the original sample with the sample of greatest similarity (Australia; population based) yielded **p=6.4x10-8** for non-synonymous SNP rs2522833 that gives rise to a serine to alanine substitution near a C2 calcium-binding-domain of the PCLO protein.

Overview of Genotyping Process

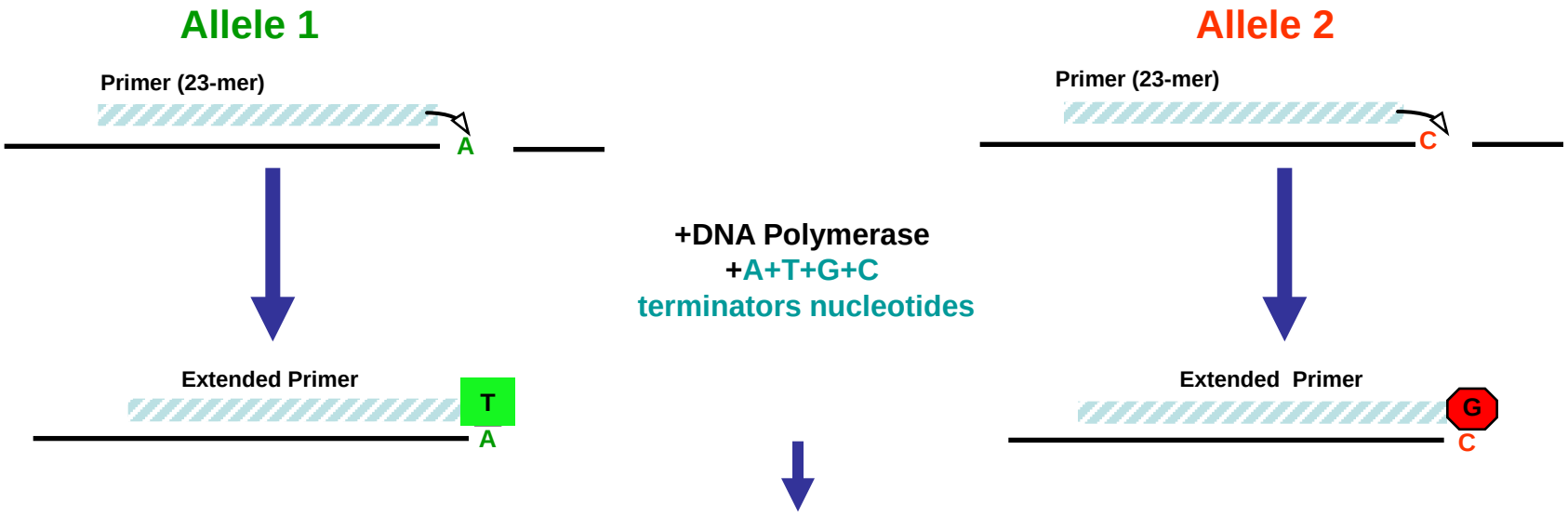


Queensland Institute of Medical Research

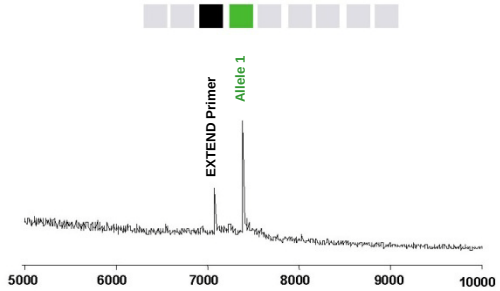
James Aug.2004

SEQUENOM INDUSTRIAL GENOMICS™

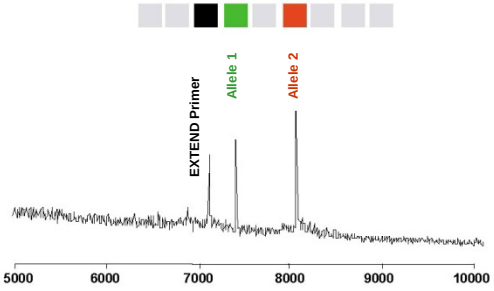
Primer Extension SNP Assay
SBE-ASSAY



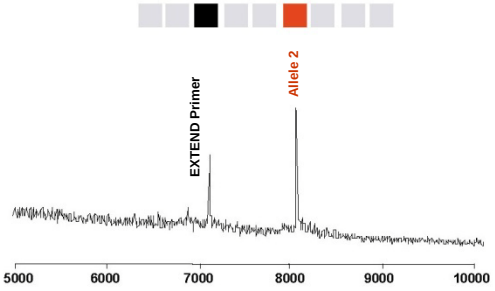
Separation of extension products by Mass



A-Homozygote



AC-Heterozygote

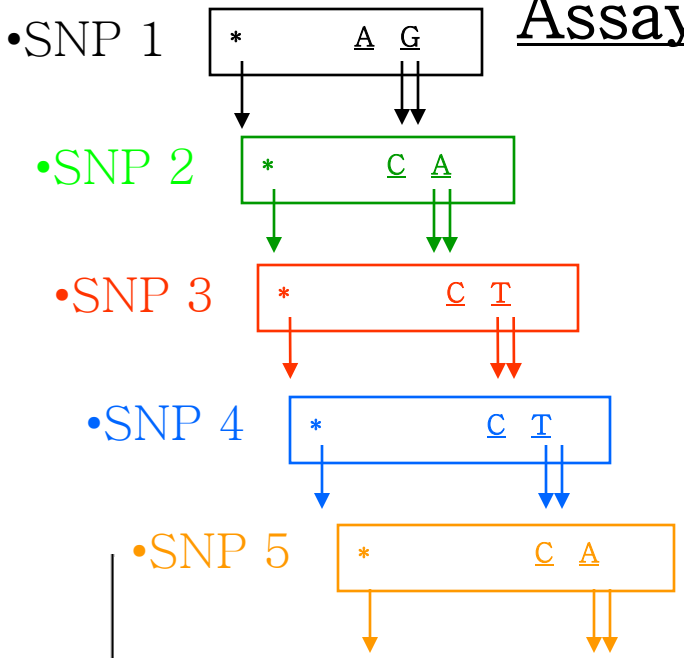


C-Homozygote

Queensland Institute of Medical Research

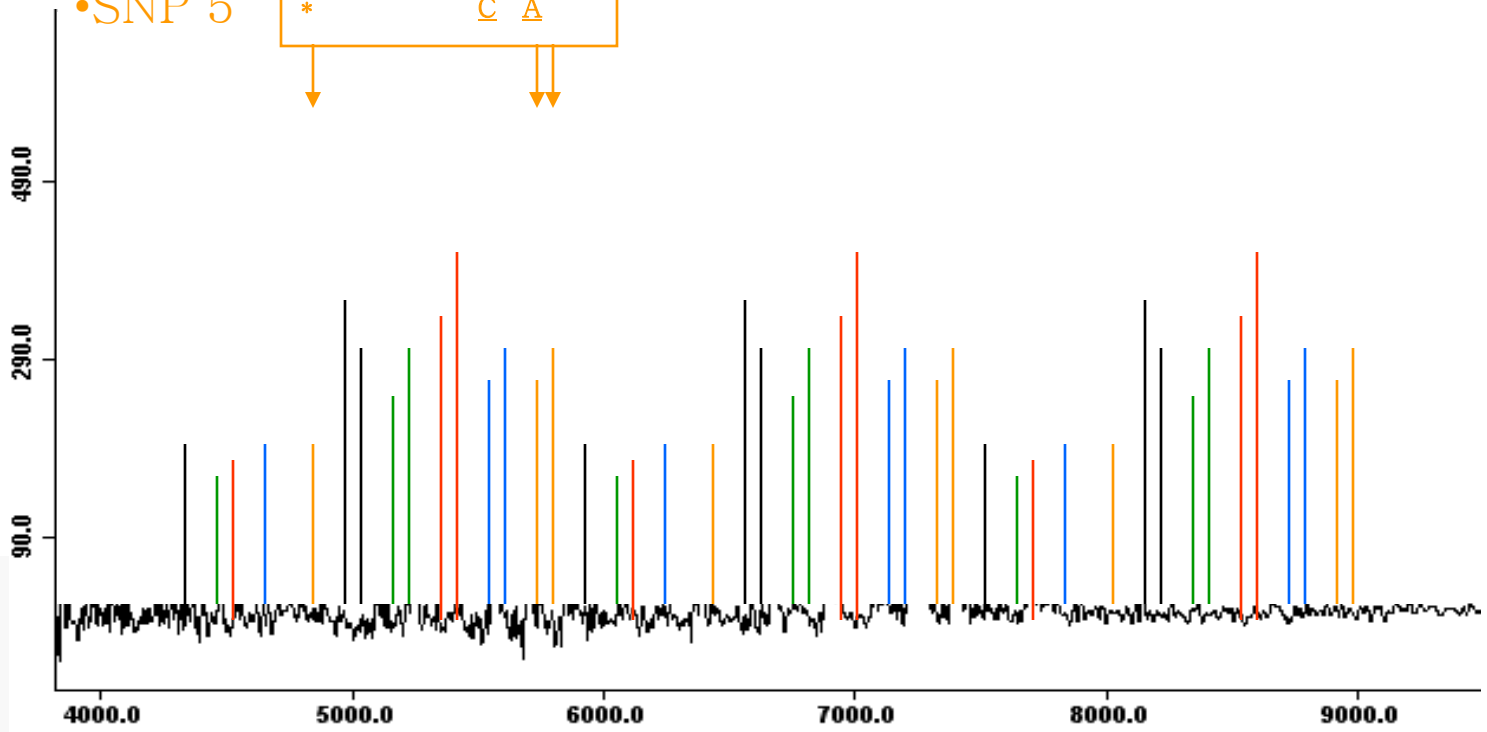
James Aug.2004

Multiplexing SBE Assay



Group 2

Group 3



Queensland Institute of Medical Research

QIMR's Sequenom MassARRAY Installation (CCRC-E floor)

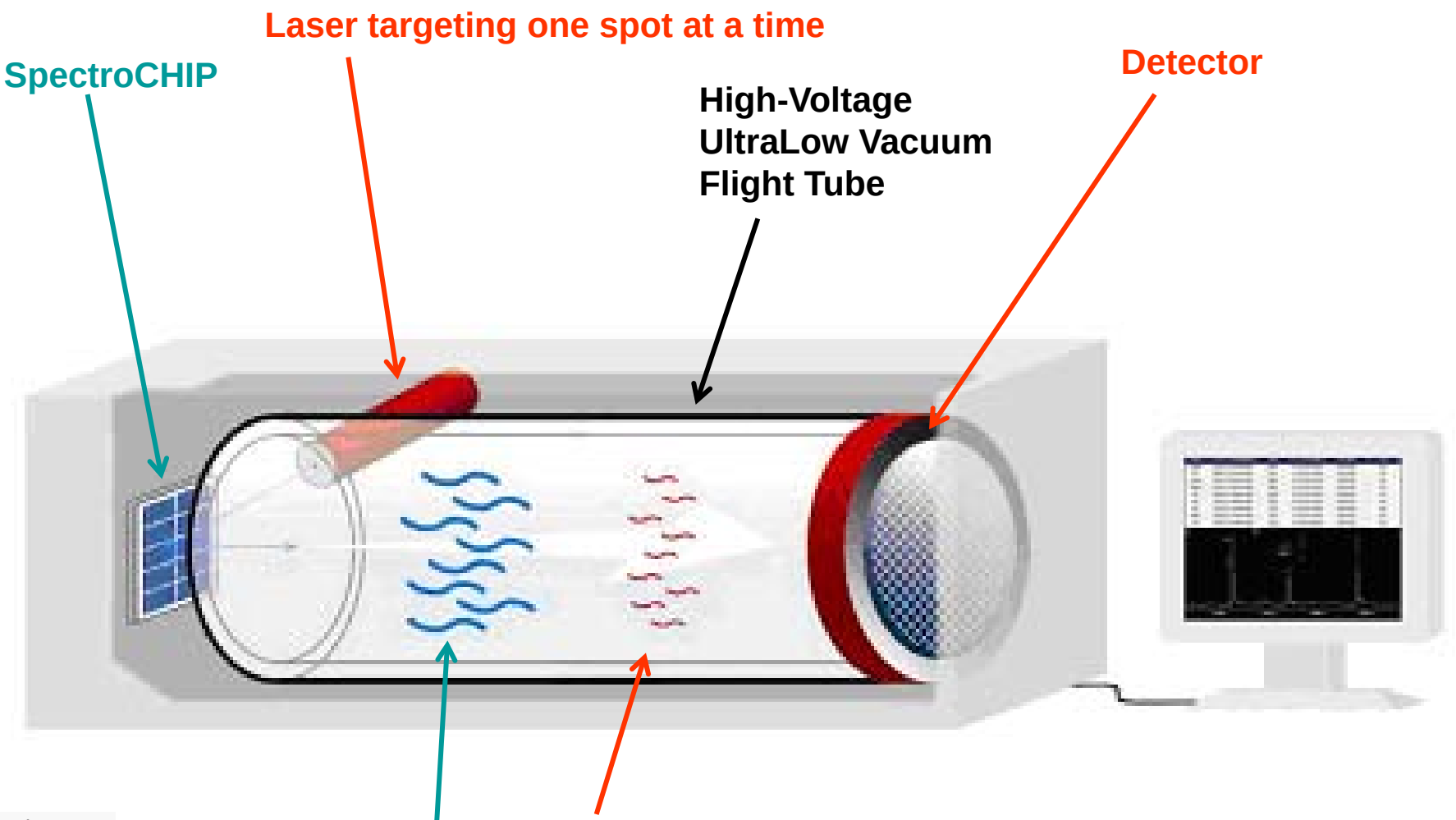
Queensland Institute of Medical Research

SEQUENOM INDUSTRIAL GENOMICS™

SpectroCHIP™

384 spots of dried DNA + matrix

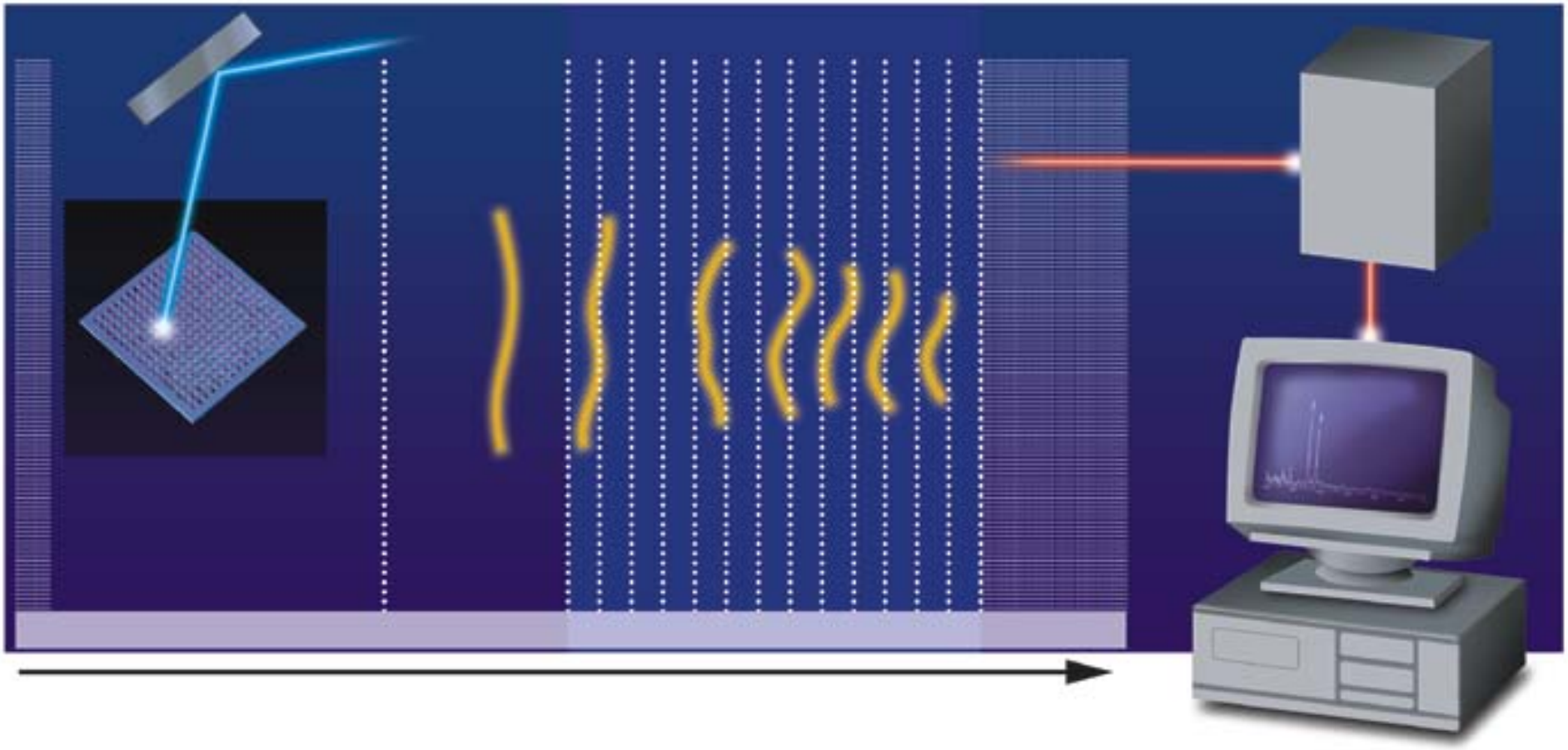
MALDI-TOF Mass Spectrometry

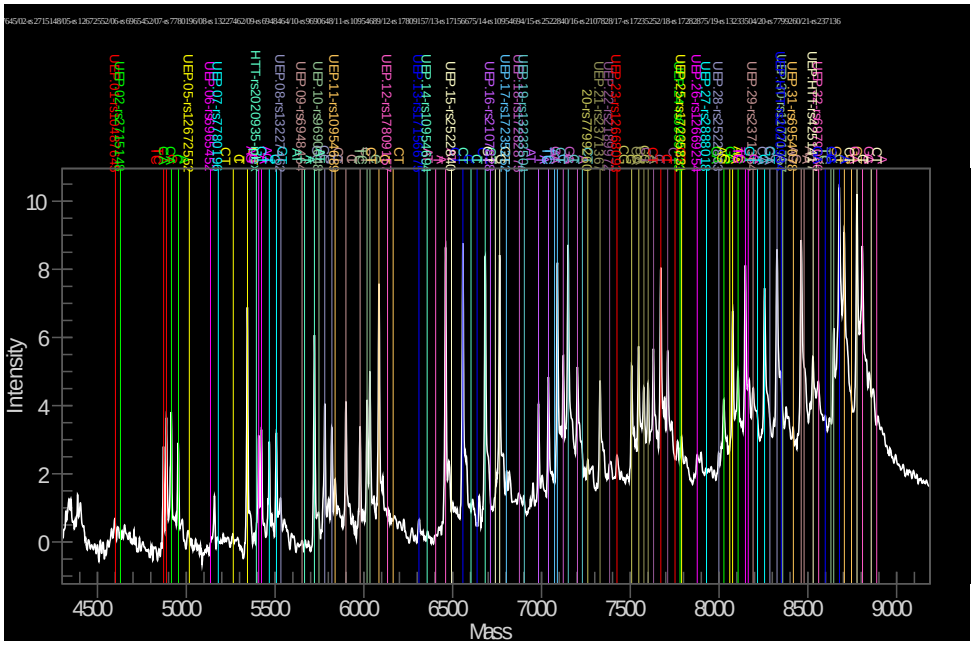


Queensland Institute of Medical Research

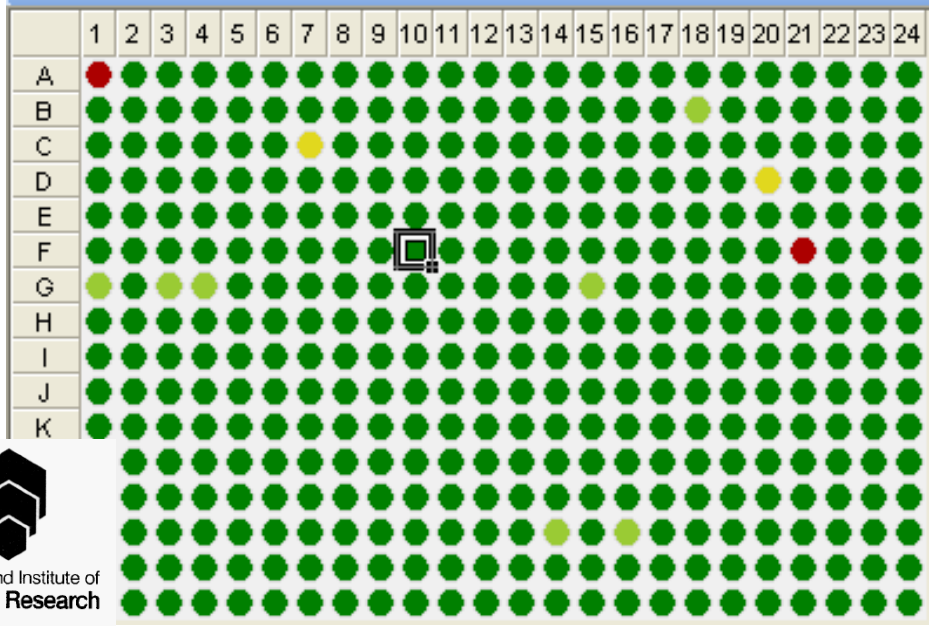
SEQUENOM INDUSTRIAL GENOMICS™

MALDI-TOF Mass Spectrometry





Traffic Light



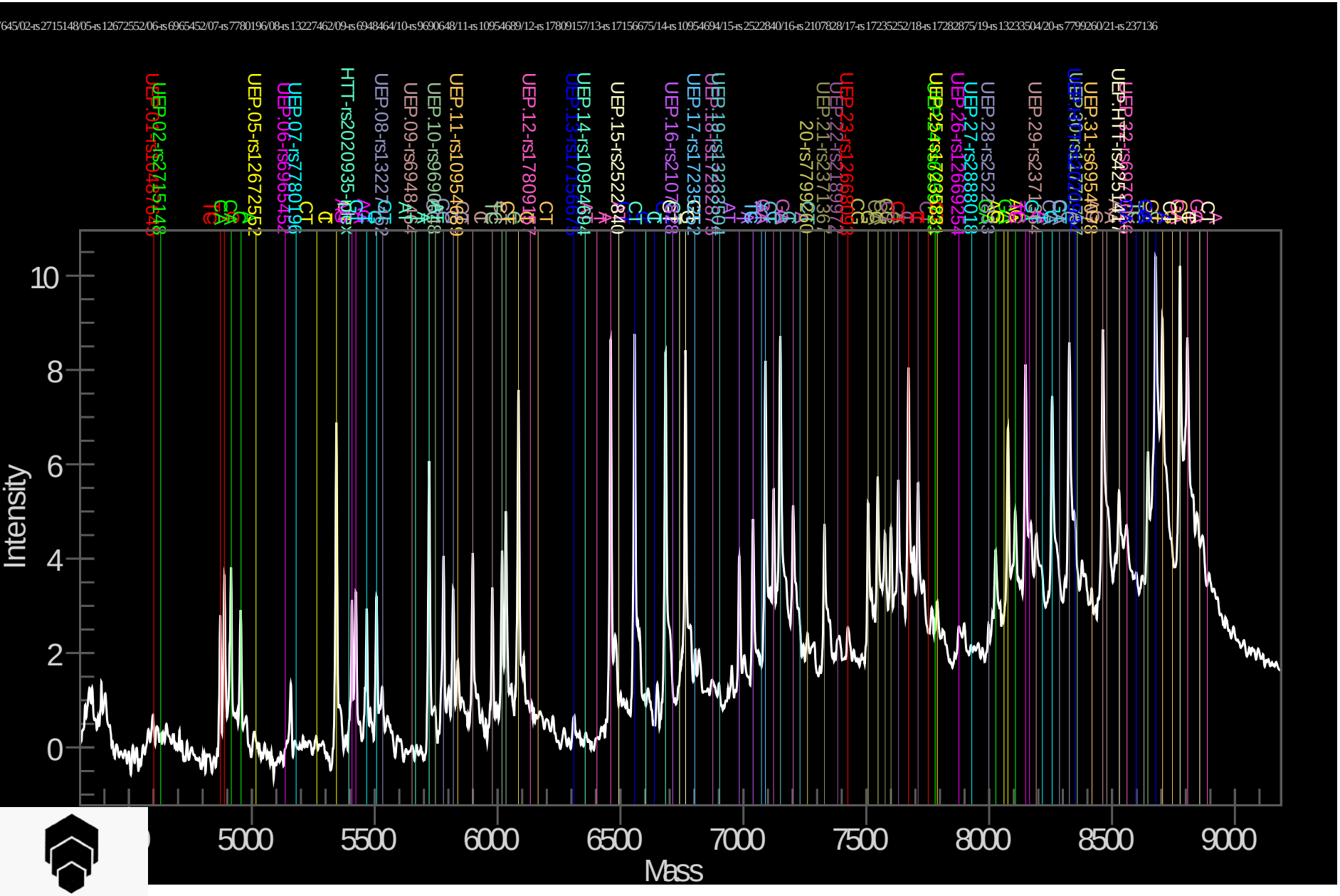
Well Data

ASSAY_ID	SA...	C...	DESCRIPTION
01-rs10487645	A 196	T	B.Moderate
02-rs2715148	A 196	CA	A.Conservative
05-rs12672552	A 196	T	A.Conservative
06-rs6965452	A 196	G	A.Conservative
07-rs7780196	A 196	T	A.Conservative
08-rs13227462	A 196	G	A.Conservative
09-rs6948464	A 196	CT	B.Moderate
10-rs9690648	A 196	T	A.Conservative
11-rs10954689	A 196	C	A.Conservative
12-rs17809157	A 196	T	A.Conservative
13-rs17156675	A 196	CT	B.Moderate
14-rs10954694	A 196	C	A.Conservative
15-rs2522840	A 196	GT	A.Conservative
16-rs2107828	A 196	T	A.Conservative
17-rs17235252	A 196	C	A.Conservative
18-rs17282875	A 196	A	A.Conservative
19-rs13233504	A 196	C	A.Conservative
20-rs7799260	A 196	G	A.Conservative
21-rs2371367	A 196	C	B.Moderate
22-rs2189972	A 196	C	A.Conservative
23-rs12668093	A 196	C	A.Conservative
24-rs6959723	A 196	A	A.Conservative
25-rs17235831	A 196	G	A.Conservative
26-rs12669254	A 196	T	B.Moderate
27-rs2888018	A 196	T	A.Conservative
28-rs2522833	A 196	CA	A.Conservative
29-rs2371364	A 196	G	A.Conservative
30-rs11771757	A 196	C	C.Aggressive
31-rs6954078	A 196	GT	C.Aggressive
32-rs6979066	A 196	GA	B.Moderate

Queensland Institute of Medical Research

James Aug.2008

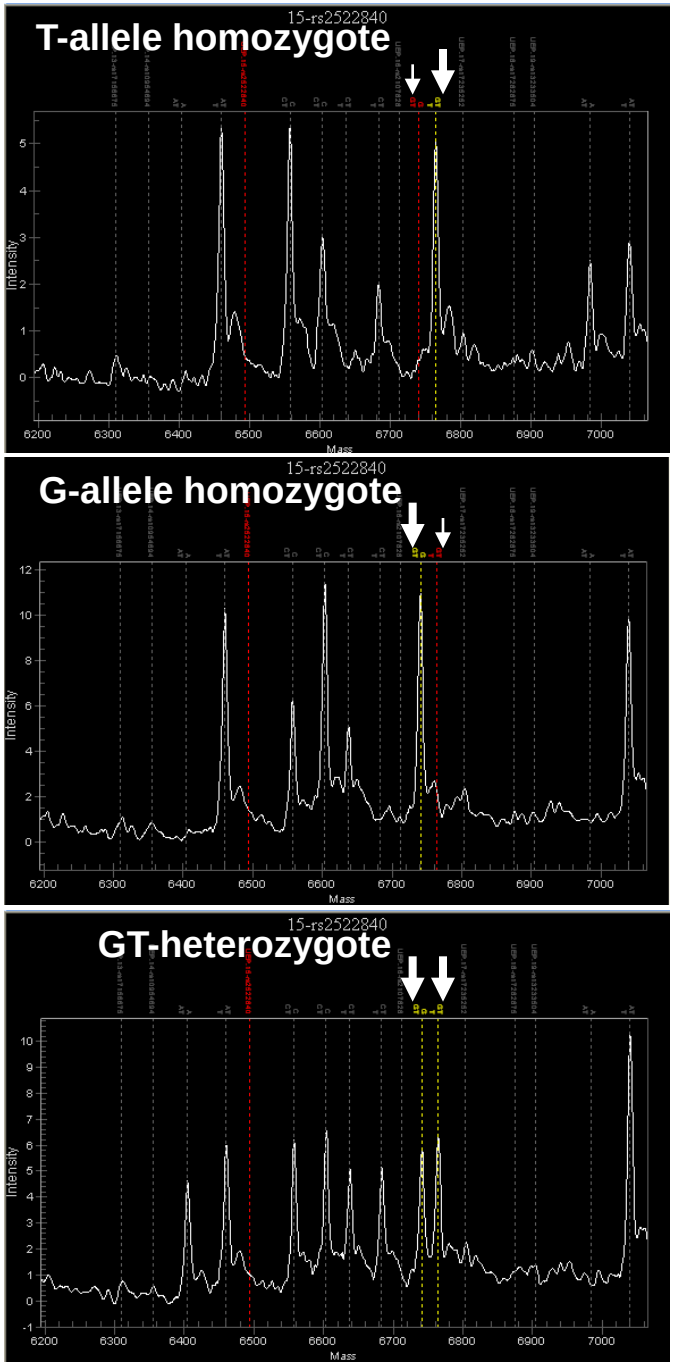
Mass Spectrum of 30-plex SNPs



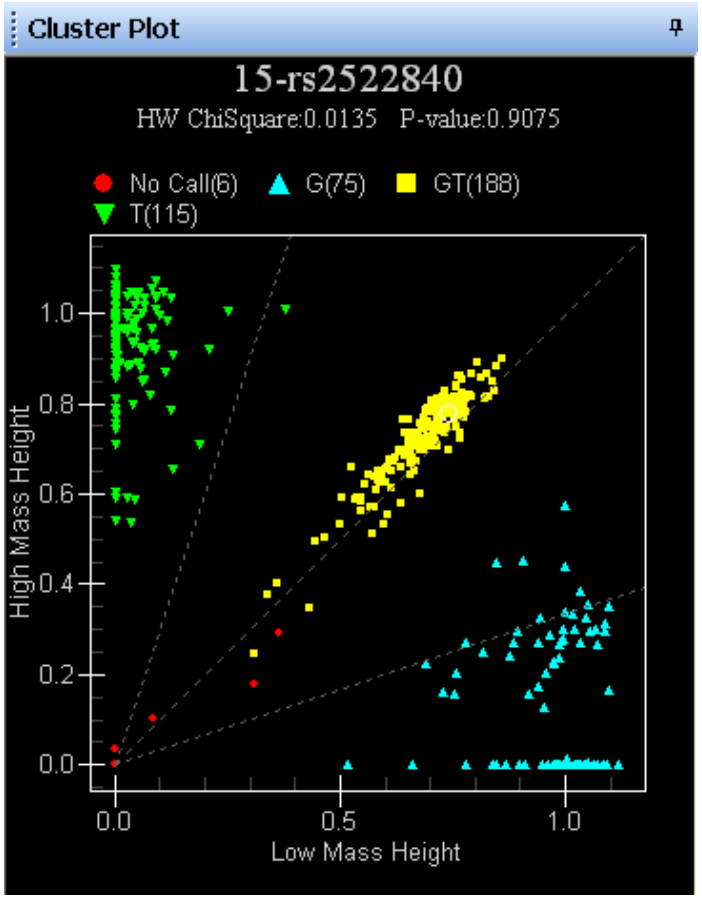
Queensland Institute of Medical Research

James Aug.2008

Spectra of rs2522840



rs2522840: cluster plot of data for 384 samples



Queensland Institute of Medical Research

Results of Genotyping Project:
(QIMR replication of GAIN *pclo*)

- **30-SNP multiplex**
- **2008 case/control samples**

- **Av. 98.9% genotype success**
- **120,326 alleles called**
- **77 missing genotypes**

	Alleles Cases	Controls	Sum	Alleles % genotyped
rs7780196	1938	2076	4014	0.99950
rs17282875	1934	2076	4010	0.99851
rs10954689	1936	2078	4014	0.99950
rs12672552	1938	2078	4016	1.00000
rs6948464	1938	2078	4016	1.00000
rs13227462	1938	2076	4014	0.99950
rs17156675	1938	2078	4016	1.00000
rs6979066	1924	2048	3972	0.98904
rs6965452	1936	2072	4008	0.99801
rs11771757	1924	2068	3992	0.99402
rs12668093	1938	2078	4016	1.00000
rs6954078	1916	2070	3986	0.99253
rs2715148	1936	2076	4012	0.99900
rs2522833	1936	2074	4010	0.99851
rs2522840	1936	2078	4014	0.99950
rs13233504	1934	2078	4012	0.99900
rs2888018	1938	2078	4016	1.00000
rs2371364	1936	2078	4014	0.99950
rs2371367	1936	2078	4014	0.99950
rs2189972	1936	2078	4014	0.99950
rs17235252	1936	2078	4014	0.99950
rs17809157	1936	2076	4012	0.99900
rs2107828	1938	2078	4016	1.00000
rs10954694	1938	2078	4016	1.00000
rs10487645	1938	2078	4016	1.00000
rs9690648	1938	2078	4016	1.00000
rs17235831	1936	2076	4012	0.99900
rs6959723	1936	2076	4012	0.99900
rs7799260	1938	2078	4016	1.00000
rs12669254	1938	2078	4016	1.00000
			No. >99%	
			Minimum	**0.98904**
		Sum	120,326	
		Missing genotypes	77	

Queensland Institute of Medical Research

Practical 1

- Run PLINK on the Dutch dataset, SNP rs2715148

Files:	Type of test:	Options:
Dutch_sample.ped	1 df χ^2-test	--allow-no-sex
Mapfile.map		--ci 0.95
		--snp rs#

- Use the output to compute the β, se and 95% CI
- Compare your CI with that given in the output

- If you have time, repeat for another SNP or dataset

Formulas needed

$\beta = \ln(OR)$

$$\chi^2 = \frac{\beta^2}{se^2}$$

95% CI (β) = β +/- 1.96 * se

95% CI (OR) = $OR = e^{\beta}$

Answer Practical 1

SNP rs2715148, Dutch dataset

Output: $\chi^2=26.36$ OR=0.786

$$\chi^2 = \frac{\beta^2}{se^2}$$

$\beta=\ln(OR)=-0.2408$

$se=\sqrt{[\beta^2/\chi^2]}= \sqrt{[(-.2408)^2/26.36]}=0.0469$

Lower 95% CI (β) = $\beta-1.96*se=-0.2408-1.96*0.0469=-0.3327$

Upper 95% CI (β) = $\beta+1.96*se=-0.2408+1.96*0.0469=-0.1489$

Lower 95% CI (OR) = $e^{-0.3327} = 0.7169$

Upper 95% CI (OR) = $e^{-0.1489} = 0.8617$

Practical 2

Needed information for meta-analysis:

Inverse variance weighted method:

- Beta
- Standard error

Z-score pooling method:

- P-value
- Direction of effect (Beta or +/- column)
- Effective sample size per SNP

Practical 2

- Compute meta-analysis results by hand

Inverse variance weighted method

Files:

Americanresults.txt

Australianresults.txt

British1results.txt

British2results.txt

Dutchresults.txt

German1results.txt

German2results.txt

Strand / reference allele flips

Scheme that METAL uses:

Example 1; Strand flips required

	ALLELES	EFFECT	ALLELES Analyzed	EFFECT Analyzed
Input file 1	T/G	+	a/c	+
Input file 2	T/G	+	a/c	+
Input file 3	A/C	+	a/c	+
Output			a/c	+

Example 2; Reference allele flips required

	ALLELES	EFFECT	ALLELES Analyzed	EFFECT Analyzed
Input file 1	C/A	-	a/c	+
Input file 2	C/A	-	a/c	+
Input file 3	A/C	+	a/c	+
Output			a/c	+

Example 2; Strand flips, numeric flips, and reference allele flips required

	ALLELES	EFFECT	ALLELES Analyzed	EFFECT Analyzed
Input file 1	G/T	-	a/c	+
Input file 2	2/1	-	a/c	+
Input file 3	A/C	+	a/c	+
Output			a/c	+

Taken from: http://www.sph.umich.edu/csg/abecasis/Metal/index.html

Practical 2

- Compute meta-analysis results by hand

Inverse variance weighted method

Files:

Americanresults.txt
Australianresults.txt
British1results.txt
British2results.txt
Dutchresults.txt
German1results.txt
German2results.txt

Meta-analysis_Inverse_Variance_Weighted_Method.xls

Practical 2

- Compute meta-analysis results by hand

- Use the information in the files to compute the pooled β, pooled se, chi-square test statistic, p-value
- Start with SNP rs2715148
- If you have time, repeat for rs2522833 and rs2371367

Formulas needed

$$\beta_{pooled} = \frac{\sum_{i=1}^{N} (w_i * \beta_i)}{\sum_{i=1}^{N} (w_i)}$$

$$se_{pooled} = \sqrt{\frac{1}{\sum_{i=1}^{N} (w_i)}}$$

$$\chi^2_{df=1} = \frac{\beta_{pooled}^{\ 2}}{se_{pooled}^{\ 2}} = \frac{(\sum_{i=1}^{N} w_i * \beta_i)^2}{\sum_{i=1}^{N} w_i}$$

Answer Practical 2

	SNP	**rs2715148**					
Study	***Reference allele***	***Non-reference allele***	***Beta***	***Standard error***	***Beta refall=A***	***Weight***	***Weight*Beta***
Dutch	A	C	-0.2408	0.0469	-0.2408	454.6089	-109.4691
American	A	C	0.1310	0.0634	0.1310	248.9457	32.6189
Australian	C	A	0.1124	0.0631	-0.1124	251.3897	-28.2651
British 1	C	A	-0.0183	0.0557	0.0183	322.5035	5.8908
British 2	C	A	-0.0155	0.0993	0.0155	101.3848	1.5735
German 1	C	A	0.0431	0.0648	-0.0431	238.0645	-10.2509
German 2	C	A	-0.0383	0.0568	0.0383	310.1141	11.8852
					Sum:	1927.0111	-96.0168
					Pooled effect:	-0.0498	
					Pooled se:	0.0228	
					Chi-square:	4.7842	
					Z-score:	-2.1873	
					P-value:	0.0287	

METAL website

http://www.sph.umich.edu/csg/abecasis/Metal/index.html
Goncalo Abecasis

- Pdf with instructions
- More options found when running METAL

Running METAL

- Copy metal.exe and all results files to your working directory
- Open Cygwin, go to your working directory
- Type ./metal (this loads the program)

Running METAL

Z-score pooling method:

```
MARKER SNP
ALLELE A1 A2
PVALUE P
EFFECT BETA
WEIGHTLABEL NMISS
PROCESS Dutchresults.txt
PROCESS Americanresults.txt
PROCESS Australianresults.txt
PROCESS British1results.txt
PROCESS British2results.txt
PROCESS German1results.txt
PROCESS German2results.txt
OUTFILE metaanalysis_zscore .txt
ANALYZE
```

Inverse variance weighted method:

```
MARKER SNP
ALLELE A1 A2
PVALUE P
EFFECT BETA
WEIGHTLABEL NMISS
STDERRLABEL SE
SCHEME STDERR
PROCESS Dutchresults.txt
PROCESS Americanresults.txt
PROCESS Australianresults.txt
PROCESS British1results.txt
PROCESS British2results.txt
PROCESS German1results.txt
PROCESS German2results.txt
OUTFILE metaanalysis_inverse .txt
ANALYZE
```

Practical 3

- Run meta-analysis using METAL

Files:

Americanresults.txt
Australianresults.txt
British1results.txt
British2results.txt
Dutchresults.txt
German1results.txt
German2results.txt

- Compare the output to the results from the Excel file

Answer Practical 3

MarkerName	Allele1	Allele2	Effect	StdErr	P-value	Direction
rs2715148	a	c	-0.0498	0.0228	0.02872	-+-++-+
rs2522833	a	c	-0.0594	0.0229	0.009397	-+-++--
rs2371367	a	c	-0.0496	0.0266	0.06199	---?-+-

Further reading

- http://www.cochrane-net.org/openlearning/
- Kavvoura & Ioannidis (2008). Methods for meta-analysis in genetic association: a review of their potential and pitfalls. *Human Genetics. 123*:1-14.
- De Bakker et al. (2008). Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Human Molecular Genetics. 17*:R122-R128.

Meta-analysis in candidate gene studies

Example study: Munafo et al. (2008) Biological Psychiatry

Serotonin Transporter (5-HTTLPR) Genotype and Amygdala Activation: A Meta-Analysis

Marcus R. Munafò, Sarah M. Brown, and Ahmad R. Hariri

Background: We evaluated the magnitude of the reported associations between amygdala activation and the serotonin transporter gene linked polymorphic region (5-HTTLPR) and the likely effect size of this relationship.

Methods: We used meta-analytic techniques to combine data from existing published and unpublished studies. We also tested for possible publication bias and explored possible moderating influences on any association, such as sample ancestry.

Results: Our results provide support for the association of the 5-HTTLPR polymorphism and amygdala activation and suggest that this locus may account for up to 10% of phenotypic variance. Although we did not observe evidence for potential publication bias in our main analysis, this was due in part to efforts to obtain unpublished data pertinent to this meta-analysis, and when three unpublished data sets were excluded we did observe evidence of such bias. We also observed evidence that the first published study may provide an overestimate of the true effect size, which is consistent with findings from genetic association studies of other phenotypes.

Conclusions: Although our analysis provides support for the association of the 5-HTTLPR polymorphism and amygdala activation, it also suggests that most studies to date are nevertheless lacking in statistical power. Increasing the sample sizes of future imaging genetics studies will allow a more accurate characterization of any true effect size and afford adequate power to examine the impact of multiple polymorphisms that likely work in concert to affect gene function and, in turn, bias neural processes mediating dispositional traits such as temperament and personality.

Key Words: 5-HTTLPR, amygdala, fMRI, meta-analysis, serotonin transporter gene

Individual differences in trait negative affect are important predictors of vulnerability for a spectrum of health-related disorders including depression, anxiety, and cardiovascular

aptic 5-HT receptor stimulation. In 1996, Lesch and colleagues (5,6) identified a relatively common functional promoter polymorphism in the human 5-HTT gene (*SLC6A4*). The so-called 5-HTT gene linked polymorphic region, or 5-HTTLPR, is typically defined by two variable nucleotide tandem repeat elements, a short (S) allele comprising 14 copies of a 20-23 base pair repeat

Meta-analysis in candidate gene studies

Example study: Munafo et al.

Forest plot

Winners curse

Studyname	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z Value	p Value
Hariri 2002	3.748	.627	.394	2.518	4.977	5.973	.000
Heinz 2004	.688	.411	.169	- .118	1.495	1.673	.094
Bertolino 2005	.193	.395	.156	- .581	.968	0.490	.624
Canli 2005	.542	.341	.116	- .126	1.210	1.590	.112
Hariri 2005	.603	.233	.054	.145	1.060	2.583	.010
Brown 2006	.643	.284	.081	.086	1.200	2.261	.024
Canli 2006	1.459	.503	.253	.474	2.444	2.903	.004
Dannlowski 2006	.886	.425	.181	.034	1.699	2.039	.041
Domschke 2006	1.014	.495	.245	.043	1.985	2.046	.041
Dannlowski 2007	.403	.328	.108	- .240	1.045	1.228	.219
Rao 2007	.892	.411	.169	.086	1.698	2.168	.030
Steinunpub	1.204	.664	.440	- .096	2.505	1.815	.070
Surguladze unpub	.333	.436	.190	- .522	1.188	0.764	.445
Taylor unpub	-1.882	.499	.249	-2.861	- .903	-3.769	.000
	.626	.103	.011	.424	.828	6.087	.000

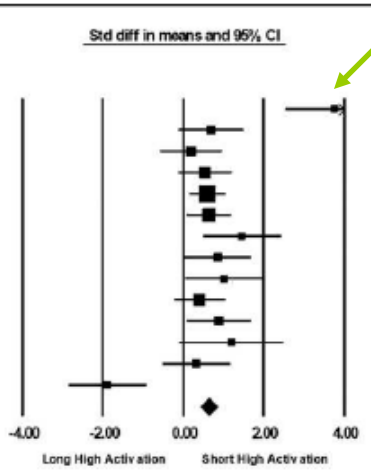


Figure 1. Meta-analysis of association studies of 5-HTTLPR genotype and amygdala activation. Meta-analysis indicates significant association between 5-HTTLPR genotype and amygdala activation ($p < .001$). Bars represent individual study 95% confidence intervals, with a central block proportional to study size. The summary diamond bar represents the pooled effect size estimate and 95% confidence interval (CI).

Meta-analysis in GWA studies

Example study: Frayling et al. (2007) Science

A Common Variant in the *FTO* Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity

Timothy M. Frayling,[1,2]* Nicholas J. Timpson,[3,4]* Michael N. Weedon,[1,2]* Eleftheria Zeggini,[3,5]* Rachel M. Freathy,[1,2] Cecilia M. Lindgren,[3,5] John R. B. Perry,[1,2] Katherine S. Elliott,[3] Hana Lango,[1,2] Nigel W. Rayner,[3,5] Beverley Shields,[2] Lorna W. Harries,[2] Jeffrey C. Barrett,[3] Sian Ellard,[2,6] Christopher J. Groves,[5] Bridget Knight,[2] Ann-Marie Patch,[2,6] Andrew R. Ness,[7] Shah Ebrahim,[8] Debbie A. Lawlor,[9] Susan M. Ring,[9] Yoav Ben-Shlomo,[9] Marjo-Riitta Jarvelin,[10,11] Ulla Sovio,[10,11] Amanda J. Bennett,[5] David Melzer,[1,12] Luigi Ferrucci,[13] Ruth J. F. Loos,[14] Inês Barroso,[15] Nicholas J. Wareham,[14] Fredrik Karpe,[5] Katharine R. Owen,[5] Lon R. Cardon,[3] Mark Walker,[16] Graham A. Hitman,[17] Colin N. A. Palmer,[18] Alex S. F. Doney,[19] Andrew D. Morris,[19] George Davey Smith,[4] The Wellcome Trust Case Control Consortium,† Andrew T. Hattersley,[1,2]‡§ Mark I. McCarthy[3,5]‡

Obesity is a serious international health problem that increases the risk of several common diseases. The genetic factors predisposing to obesity are poorly understood. A genome-wide search for type 2 diabetes–susceptibility genes identified a common variant in the *FTO* (fat mass and obesity associated) gene that predisposes to diabetes through an effect on body mass index (BMI). An additive association of the variant with BMI was replicated in 13 cohorts with 38,759 participants. The 16% of adults who are homozygous for the risk allele weighed about 3 kilograms more and had 1.67-fold increased odds of obesity when compared with those not inheriting a risk allele. This association was observed from age 7 years upward and reflects a specific increase in fat mass.

Obesity is a major cause of morbidity and mortality, associated with an increased risk of type 2 diabetes mellitus, heart disease, metabolic syndrome, hypertension, stroke, despite considerable efforts, there are, as yet, no examples of common genetic variants for which there is widely replicated evidence of association with obesity in the general population.

Meta-analysis in GWA studies

Example study: Frayling et al.

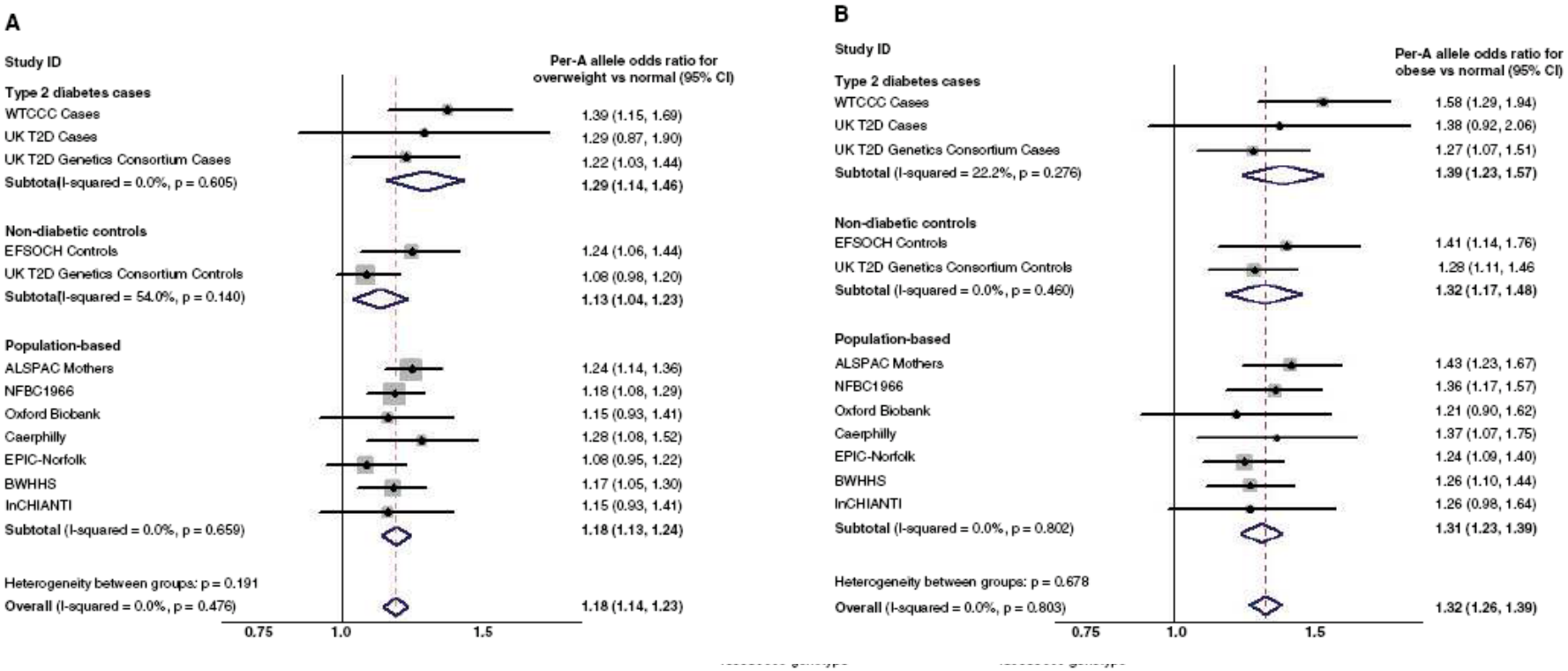


Fig. 2. (**A** and **B**) Meta-analysis plots for odds of (A) overweight and (B) obesity, compared with normal weight in adults for each copy of the A allele of rs9939609 carried. (**C** and **D**) Bar charts showing (C) DEXA-measured fat mass in 9-year-old children and (D) DEXA-measured lean mass in 9-year-old children, both from the ALSPAC study. Error bars represent 95% confidence intervals