Meta-analysis in genetics and genomics

Darlene Goldstein

École Polytechnique Fédérale de Lausanne

Joint work with Mauro Delorenzi, Thierry Sengstag and Ruth Luthi-Carter

Joint CSPS/IMS Meeting, Beijing

9 July 2005

Outline

- The problem – an illustration
- What is meta-analysis and why might it help
- Why meta-analysis might not be straightforward – study heterogeneity
- Combining information across linkage studies – 'simple', 'worst' case scenarios
- Combining information across microarray studies – *'simple'*, 'worst' case scenarios
- Concluding remarks

The problem: an example

- A study is carried out to examine mean IBD sharing between concordant obese (BMI > 35) and discordant sib pairs at 8 markers in the human OB gene region

The problem: an example

- A study is carried out to examine mean IBD sharing between concordant obese (BMI > 35) and discordant sib pairs at 8 markers in the human OB gene region
- A second study performs multipoint tests of linkage between obesity-related quantitative traits and 15 markers on sibship data from pedigrees ascertained on a type II diabetic proband

The problem: an example

- A study is carried out to examine mean IBD sharing between concordant obese (BMI > 35) and discordant sib pairs at 8 markers in the human OB gene region
- A second study performs multipoint tests of linkage between obesity-related quantitative traits and 15 markers on sibship data from pedigrees ascertained on a type II diabetic proband
- A third study investigates a single marker (different from those of Studies 1 and 2) in 400 sib pairs using Haseman-Elston regression

The problem: an example

- A study is carried out to examine mean IBD sharing between concordant obese (BMI > 35) and discordant sib pairs at 8 markers in the human OB gene region
- A second study performs multipoint tests of linkage between obesity-related quantitative traits and 15 markers on sibship data from pedigrees ascertained on a type II diabetic proband
- A third study investigates a single marker (different from those of Studies 1 and 2) in 400 sib pairs using Haseman-Elston regression
- ***Question:*** ***What is the evidence in favor of linkage between obesity and the OB gene?***

What is meta-analysis?

- Statistical methods for combining ***results*** of independent studies addressing related questions
- Several different methods, including
  - Comparative binary outcomes: combining *odds ratios*
  - Continuous outcomes: combining parameter estimates via *fixed effects* or *random effects models*
  - Any outcome type: combining *p-values* from hypothesis tests about the data
- Alternative to meta-analysis: combining (or pooling) ***data*** for the analysis
- Not always appropriate to pool – *Simpson's paradox*

Why might meta-analysis help?

- To obtain *increased power*
- Studies with small sample sizes are less likely to find effects even when they exist
- 'Integration-driven discovery' (*IDD*; Choi *et al.*, *Bioinformatics* 2003)
- Given the small (but increasing) size of many linkage and microarray experiments, meta-analysis might be considered a *natural approach* to the problem of integrating results

Combining information

To combine information across studies, we consider

- Combining (pooling) *raw data* and analyzing as a single data set (suggested in the linkage context by Lander and Kruglyak, *Nat. Genet.* 1995)
  - may not be possible – *e.g.* raw data may not be available
  - may not be sensible – *e.g.* too many differences between the studies
  - may not be feasible – *e.g.* limited computing power
- Combining *parameter estimates*
  - Fixed effects model (FE)
  - Random effects model (RE)
- Combining p*-values*

FE vs. RE meta-analysis (I)

- FE and RE are both ways to obtain a *single, combined parameter estimate* from a *set* of estimates obtained from different studies
- The combined estimates are *weighted averages*
- FE assumes there is *no heterogeneity between results* of the different studies
- If the *results are heterogeneous*, then there is assumed to be no single underlying value of HD effect but rather a *distribution* of values
- Differences among study results are considered to arise from *inter-study variation* of true effect size as well as chance variation

FE vs. RE meta-analysis (II)

- In FE meta-analysis, each individual study estimate receives *weight* inversely proportional to its variance
- RE meta-analysis assumes that individual studies may be estimating *different* treatment effects
- Study weights are adjusted to take into account the additional variability τ^2 *between* studies: $w_i^* = \frac{1}{(1/w_i)+\hat{\tau}^2}$
- When the additional variability between studies is 0, then the RE model reduces to the FE model
- If assume *normality* of the estimates, can get p-values (and q-values)

Fisher combined p-values

- Other methods for combining results focus on *p-values*
- *Usually preferable to combine data or parameter estimates*, but sometimes impossible – *e.g.*, only p-values available, no parameter estimates given
- Several methods for combining p-values, an old (1930s) and commonly used method is due to Fisher
- The *Fisher summary test statistic* $\mathrm{S} = -2\sum_{i=1}^{k} \log(p_i)$
  - The theoretical null distribution of S should be χ^2_{2k}
  - Rhodes *et al.* (*Cancer Res.* 2002) obtain a p-value for S by resampling

Problem: study heterogeneity

In general, studies may vary in

- scientific research goals
- population of interest
- design
- quality of implementation
- subject inclusion and exclusion criteria
- baseline status of subjects
- treatment dosage and timing
- management of study subjects
- outcome definition or measures
- statistical methods of analysis

Heterogeneity in genetics

In the *linkage context*, sources of heterogeneity include

- different family structures
- different ascertainment rules (*e.g.* concordant sib pairs, discordant sib pairs, families with at least one affected)
- genetic heterogeneity between study populations
- different genetic markers across studies
- different statistical tests of linkage
- phenotype definition

Heterogeneity in genomics

In the *microarray context*, sources of heterogeneity include

- differences in the technology used for the study
- heterogeneity of measured expression from the same probe occurring multiple times on the array
- multiple (different) probes for the same gene
- variability in probes used by different platforms
- different gene expression measures, even with the same technology

Linkage: simple case

- *Simulated data* from Genetic Analysis Workshop (GAW) 11
- *2 genetically distinct, clinically identical* diseases with *mild* and *severe* forms
  - *Single locus* with 3 alleles, interacting with environmental factor
  - *2 epistatically interacting loci*, associated with an allele at one of the marker loci
- Family linkage and disease data from *3 populations*, each with different genetic parameters
- Genotypes at *300 polymorphic markers*
- 4 studies (2 from same population with *different ascertainment schemes*)

Analyses: Simple case

- Described in detail in
  - Goldstein, Sain, Guerra and Etzel *Gen. Epi.* 1999
  - Guerra, Etzel, Goldstein and Sain *Gen. Epi.* 1999
- Mean IBD sharing (2-point sib-pair analysis)
  - Pooling raw data
  - FE meta-analysis
  - RE meta-analysis
  - Fisher combined p-value
- Multipoint NPL
  - Pooling raw data
  - Fisher combined p-value
- ***Assumption:*** all studies make p-values available if at least one study found a 'significant' result

Results: Simple case

- When possible, *better to combine raw data* or parameter estimates than to combine p-values
- Combining p-values led to *more false positives* and *missed more true linkages*
- If results heterogeneous, may not be possible to combine p-values (based on our assumption regarding p-value availability)
- Multipoint analysis had *fewer false positives* than 2-point

Linkage: worst case (I)

- Allison and Heo (*Genetics* 1998) combined p-values across obesity linkage studies (as in the example)
  - Not always straightforward to obtain the relevant p-values
  - In these cases, they used *ad hoc* methods to obtain a single p-value for each study
- Wise, Lanchbury and Lewis (*Ann. Hum. Genet.* 1999) proposed the Genome Search Meta-analysis method, *GSMA*
  - Also sometimes called Genome Scan Meta-analysis
  - Ranks genome intervals according to evidence for linkage, combines these ranks across studies and assesses significance

Linkage: worst case (II)

- Loesgen *et al.* (*Gen. Epi.* 2001) proposed weighting multipoint NPL scores
  - Applied to GAW 12 data
  - In a simulation study, Dempfle and Loesgen (*Ann. Hum. Genet.* 2004) found favorable power properties compared to Fisher combined p-values, GSMA
- Etzel and Guerra (*Am. J. Hum. Genet.* 2002) combined H-E regression slopes via WLS to estimate location of a QTL
  - Also provide SEs and test for linkage
  - Can also be applied to other statistics

Microarray: simple case

- Data from *two experiments on the R6/2 mouse*, a mouse model of human Huntington's disease (HD)
- Looked at *effects on gene expression of different drugs* on HD and normal (WT) mice
- Each experiment is a 2x2 factorial layout
  - drug/placebo treatment
  - HD/WT mouse
- Consider only the *control groups* for the two studies
  - Study I has *8 controls*, Study II has *6 controls*
  - In each study, *half of the mice HD, half WT*
  - Experiments carried out by the *same laboratory* a few months apart, using the *same protocols*
  - Affymetrix MOE 430A; 22,690 probe sets ('genes')

Analyses: simple case

- *Combined data analysis*: pretend all the data are from one study and analyze that single (combined) data set (14 mice)
- *Meta-analysis methods*: combine results from the two individual studies by 3 methods
  - Fixed effects model (FE)
  - Random effects model (RE)
  - Combining p-values (Fisher)
    - χ^2 assumption
    - Resampling p-value

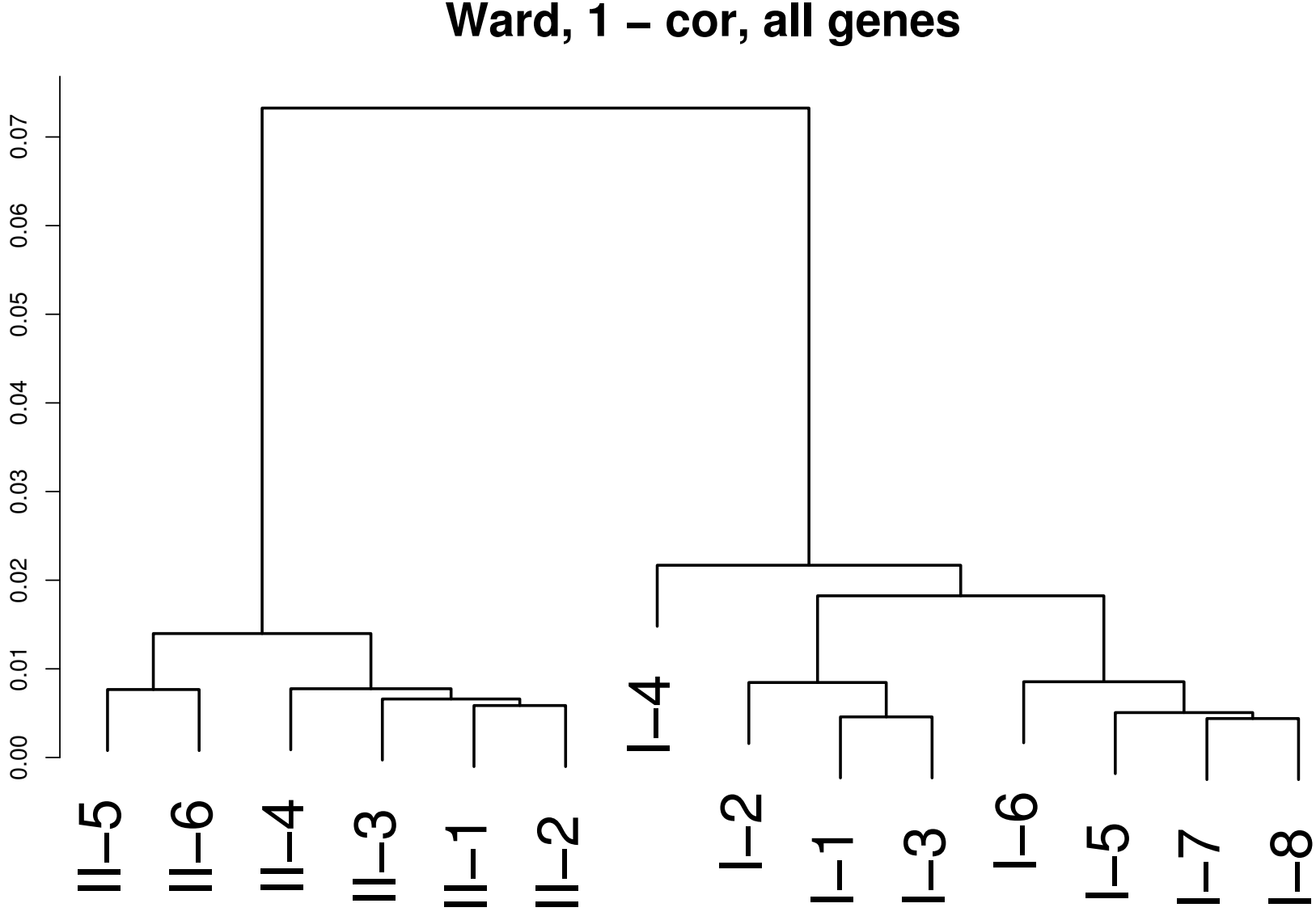
Combined data analysis

- Consider all 14 chips as a *single data set* from the same experiment
- RMA all 14 chips together (all passed quality check)
- ***Normalization***: would like to remove (all) *artifactual* differences between chips, leaving (only) *true biological differences*
- Identify DE genes using (single gene) *linear modeling* approach (`limma`, Smyth 2004)
  - Model A: $y = \beta_0 + \beta_{HD} I_{\{HD=1\}} + \epsilon$
  - Model B: $y = \beta_0 + \beta_{HD} I_{\{HD=1\}} + \beta_{batch} I_{\{batch=1\}} + \epsilon$
  - Model C: $y = \beta_0 + \beta_{HD} I_{\{HD=1\}} + \beta_{batch} I_{\{batch=I\}} + \beta_{HD\times batch} I_{\{HD\times batch=1\}} + \epsilon$

Meta-analysis

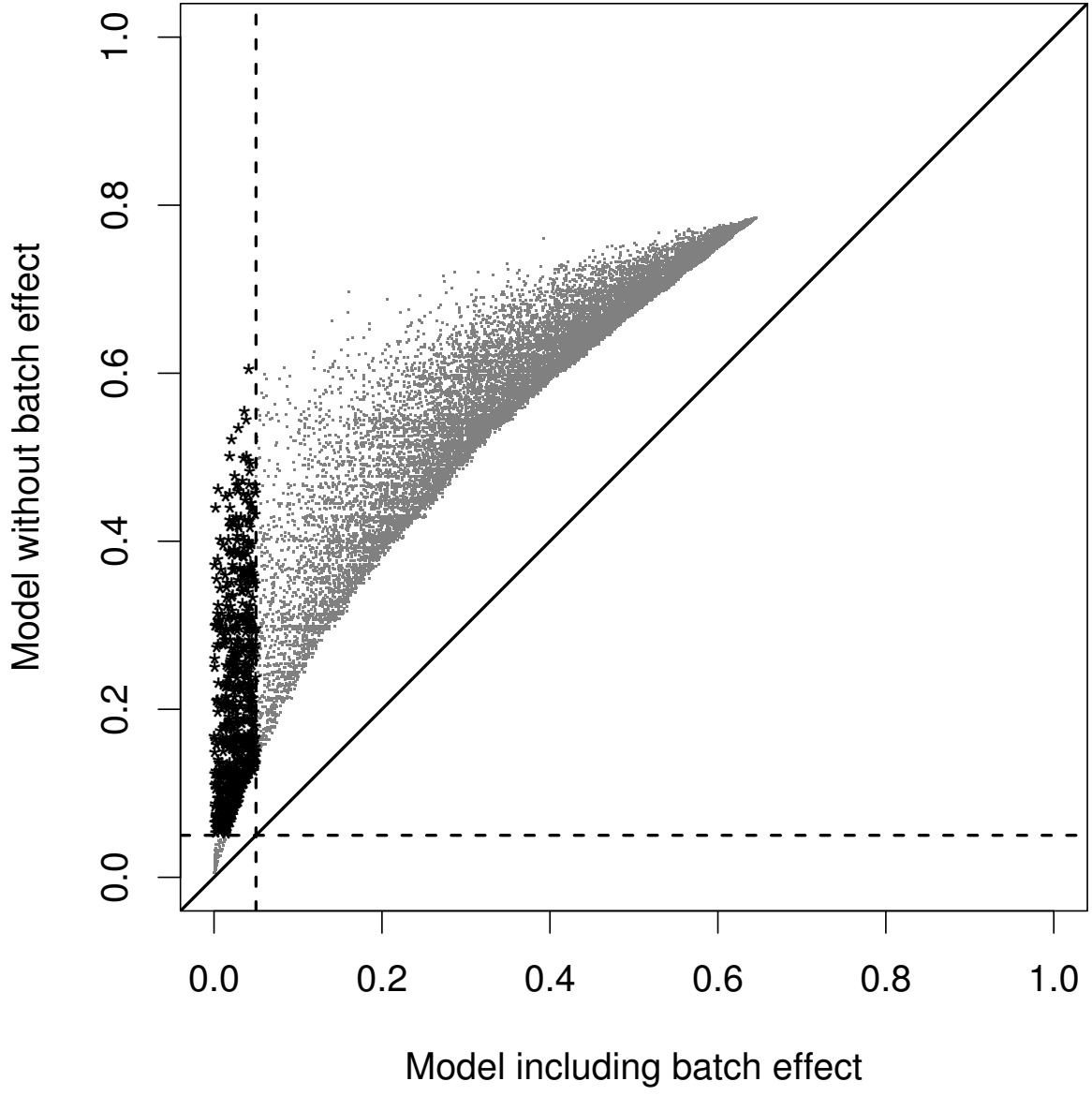
- First analyze each experiment as a separate study
  - *Separately* RMA the 2 sets (8 chips, 6 chips)
  - *Within each study* identify DE genes by the model
    $y = \beta_0 + \beta_{HD} I_{\{HD=1\}} + \epsilon$
- Heterogeneity analysis (χ^2 test)
- If appropriate, combine (via FE, RE, Fisher)

Results: Persistent batch effects



Combined data RMA dendrogram

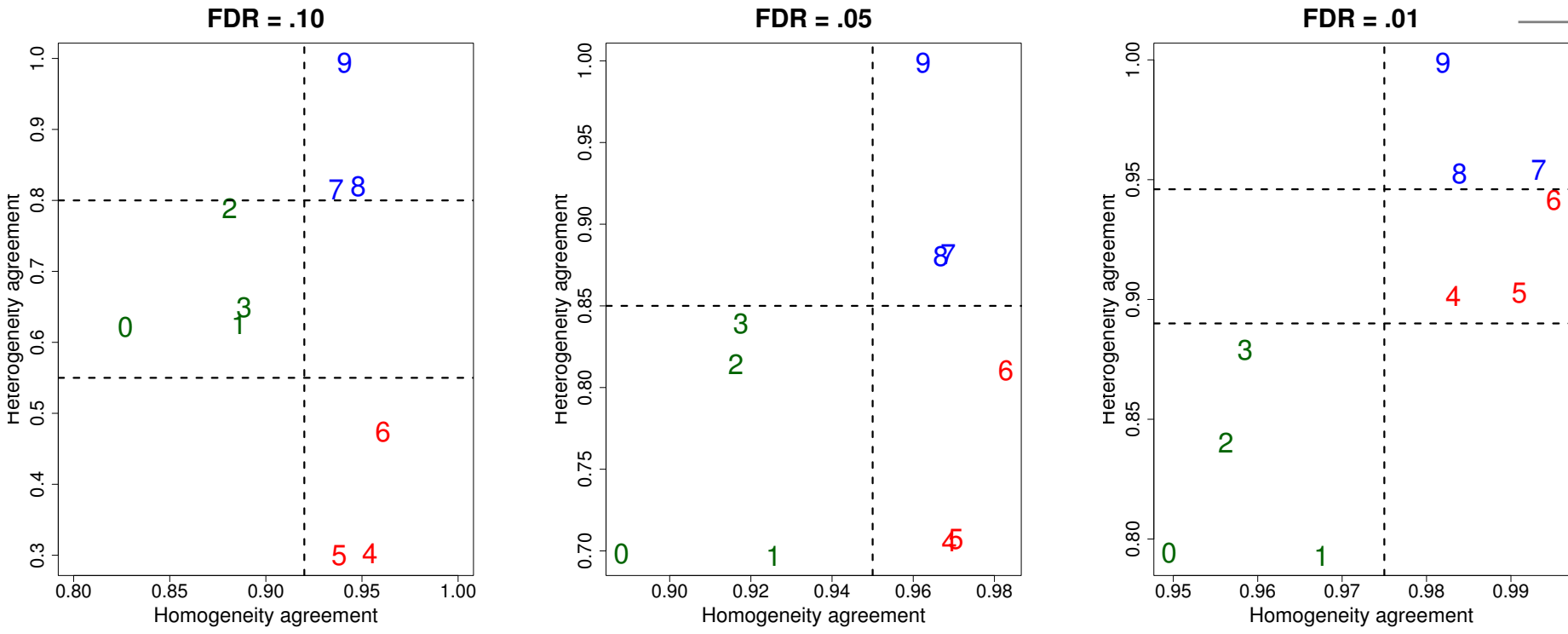
Results: HD effects



Results: Stratification by het status

	Sig. at FDR = .10			Sig. at FDR = .05			Sig. at FDR = .01		
Method	All	Hom.	Het.	All	Hom.	Het.	All	Hom.	Het.
C	0.07	0.06	0.19	0.03	0.03	0.12	0.01	0.01	0.05
FE	0.18	0.17	0.38	0.12	0.11	0.30	0.06	0.05	0.21
RE	0.06	0.06	0.01	0.04	0.04	0.00	0.02	0.02	0.00
FX	0.08	0.06	0.70	0.04	0.03	0.29	0.01	0.01	0.10
FR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FR Number	(5)	(3)	(2)	(3)	(2)	(1)	(3)	(2)	(1)

Results: Pairwise agreement



Symbol	0	1	2	3	4	5	6	7	8	9
Pair	FE	FE	FE	FE	FX	FX	FX	C	C	RE
	FR	RE	C	FX	RE	FR	C	FR	RE	FR

Microarray: worst case

- Irizarry *et al.* (*Nat. Methods* 2005) carried out a *platform comparison* that incorporated an assessment of lab effect using a random effects model
  - Importance of using *relative* measures
  - See also Cox and Solomon *Components of Variance*
- Rhodes *et al.* (*Cancer Res.* 2002) combine permutation t-test p-values for genes measured in all studies
- Choi *et al.* (*Bioinformatics* 2003) consider FE and RE meta-analysis of standardized mean differences between two sample types

Concluding remarks (I)

- ***Caveat:*** Pooling raw data not always possible or sensible
  - Even in the 'simple' cases several decisions required before data analysis can proceed
  - In *sufficiently homogeneous cases* pooling raw data probably best
  - Publicly accessible databases for obtaining raw data
- Value of exploring methods on *simple cases*
  - Can give some insight into performance in the more complicated scenarios

Concluding remarks (II)

- In the microarray simple case study here, we compare *results* from different methods of analysis; *can't assess method performance or robustness* – 'known truth' not available
  - Biological follow-up on many genes (currently in progress)
  - Simulation studies may provide some insight (Stevens and Doerge, *BMC Bioinformatics* 2005)
- Study 'batch' effects
  - *Normalize within study* prior to combining
  - Account for study when pooling, *e.g.* random effect (Irizarry *et al.*, *Nat. Methods* 2005)

Concluding remarks (III)

- Between *and within* lab variability should be examined where possible
  - What covariates might be associated with measurements
  - Can suggest improvements in lab procedures
- Implications for *large single center studies*, where
  - Patients recruited over time
  - Arrays not hybridized at the same time
  - Different technicians carry out the work
  - *etc.* ...

Acknowledgements

Collaborators

- Mauro Delorenzi and Thierry Sengstag
  - Institut Suisse de Recherche Expérimentale sur le Cancer and Swiss Institute of Bioinformatics
- Ruth Luthi-Carter
  - École Polytechnique Fédérale de Lausanne, Laboratoire de neurogénomique fonctionnelle
- R. J. Ferrante, J. K. Kubilus for providing samples

Grant Support

- Swiss National Science Foundation
- US National Institutes of Health