Efficient dimension-reduction technique for the joint analysis of correlated phenotypes

Maxime Turgeon, Karim Oualkacha, Antonio Ciampi, Golsa Dehghan, Brent W. Zanke, Celia M.T. Greenwood and Aurélie Labbe

McGill University

Poster #93
Session A

Multivariate phenotypes

- Complex diseases
  - Interest in analyzing multiple intermediate phenotypes instead of association between genetic variants and disease labels
- Joint analysis
  - Pleiotropy
  - Correlated phenotypes

Goals

- Dimension-reduction technique which can handle high-dimensional phenotypes
- Computational efficiency

Principal components of explained variance

$$\mathbf{Y} = (Y_1, \ldots, Y_p)$$

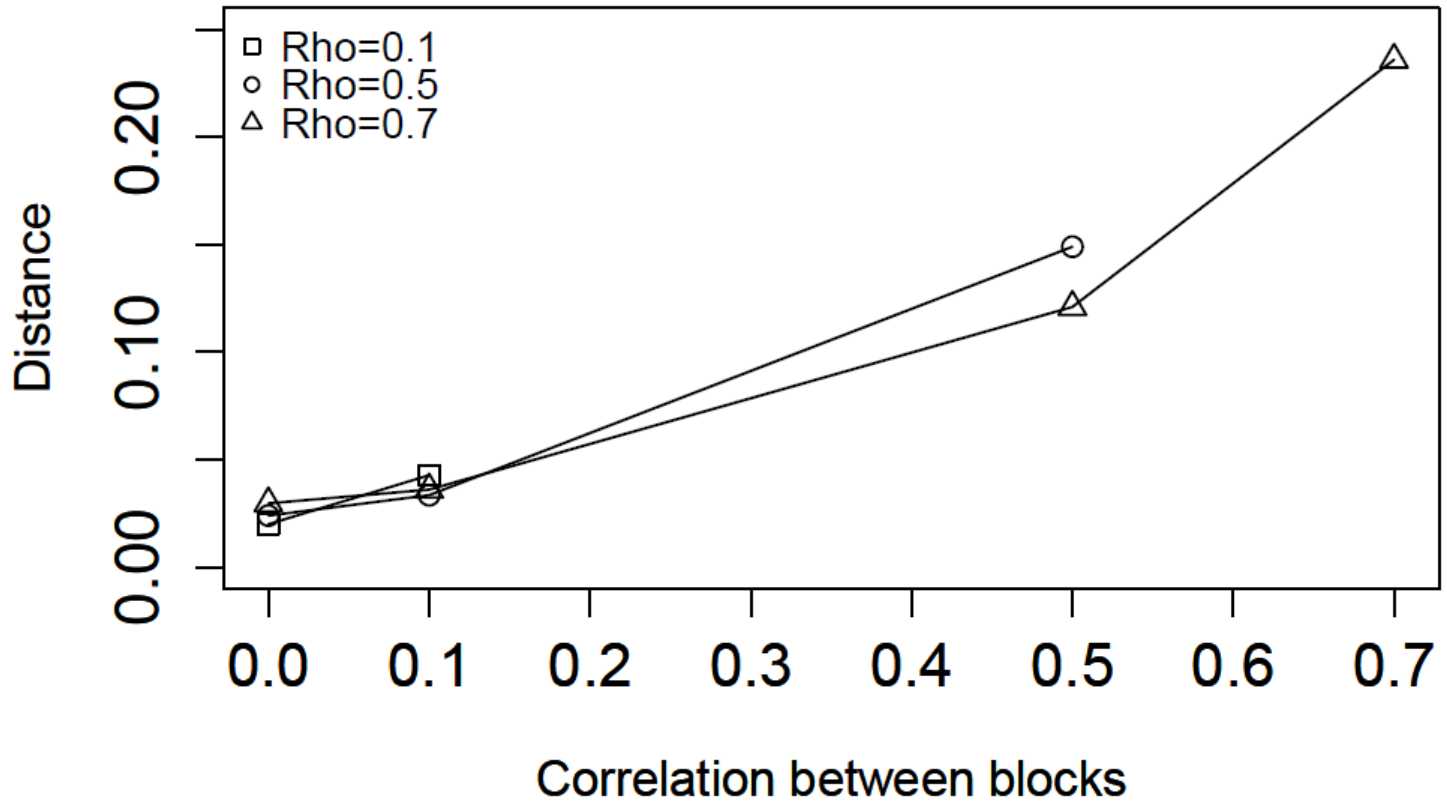
Two-stage approach to PCEV

$$\{Y_{11}, \ldots, Y_{1p_1}\}, \ldots, \{Y_{q1}, \ldots, Y_{qp_q}\} \longleftrightarrow X$$

$$\downarrow \quad \text{Maximize heritability}$$

$$\tilde{Y}_1, \ldots, \tilde{Y}_q \longleftrightarrow X$$

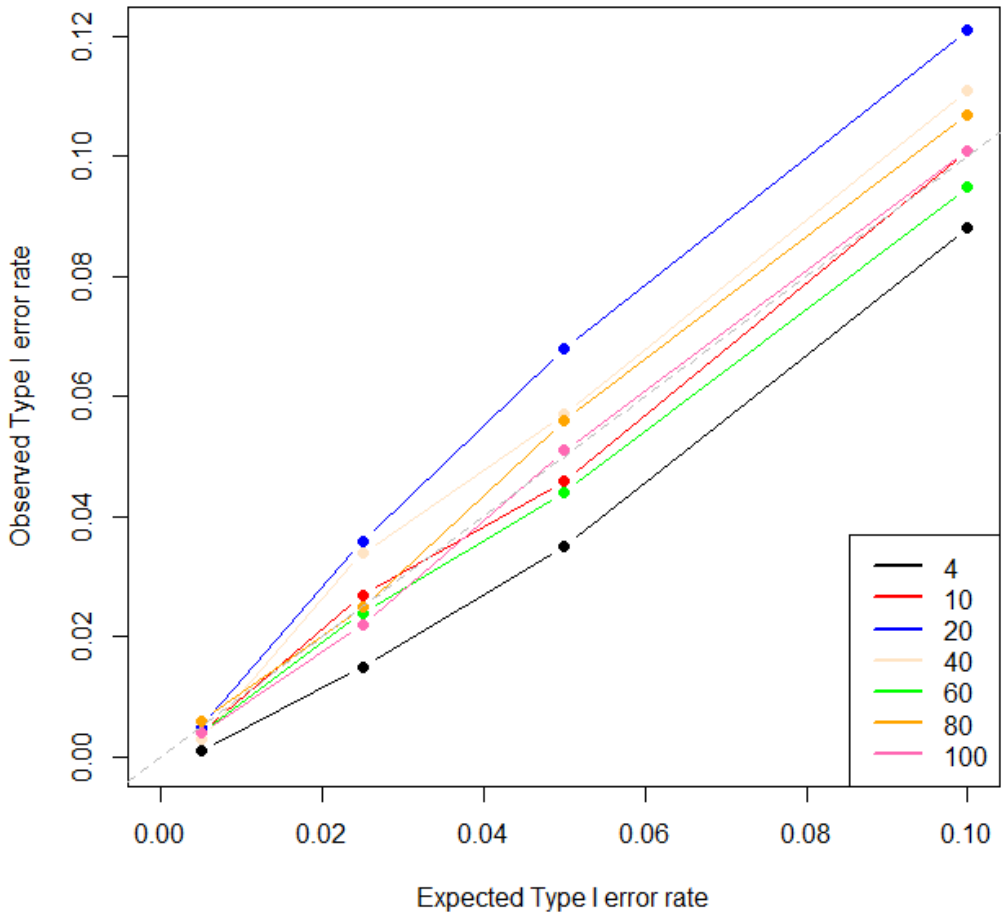
PCEV computation



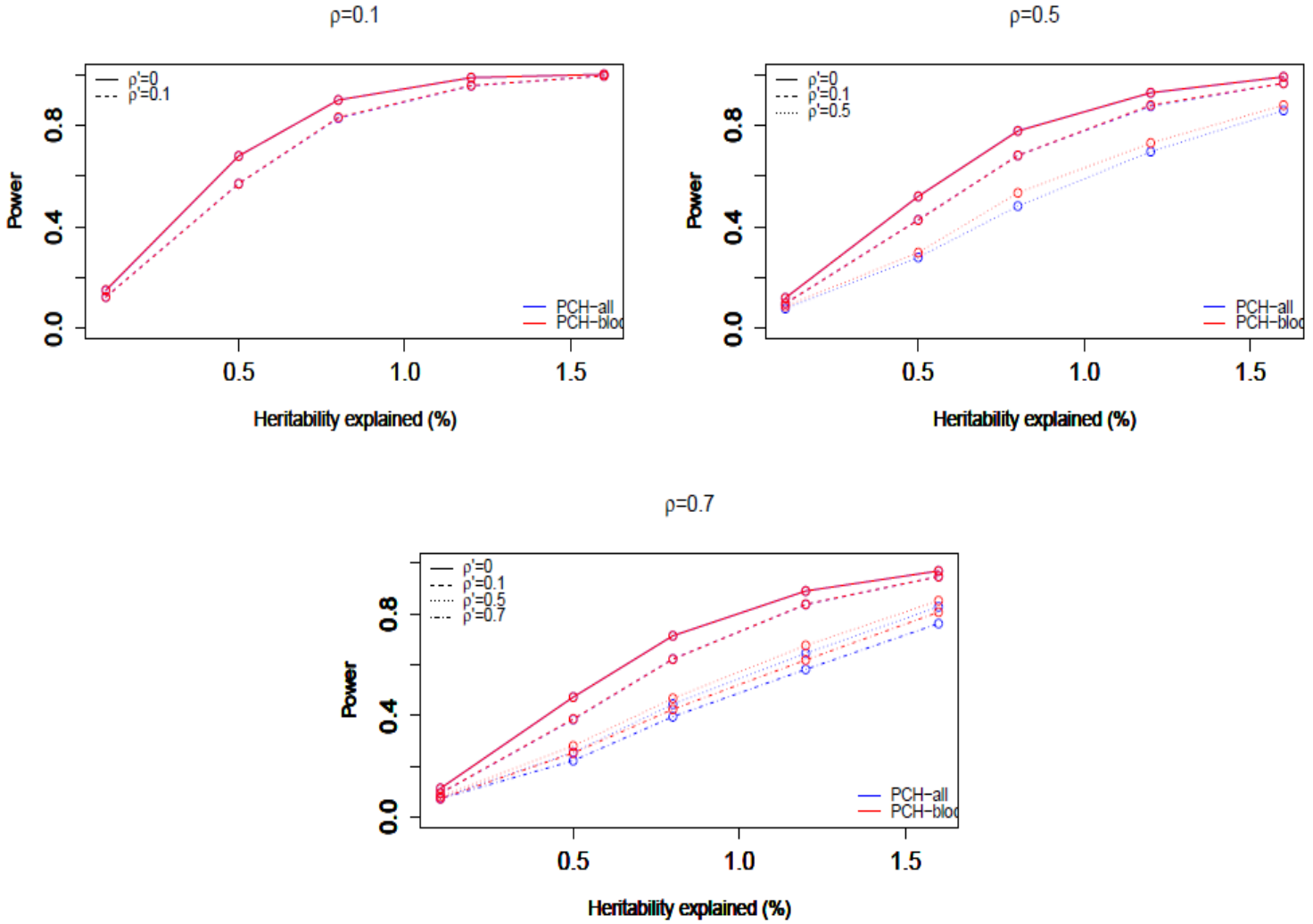
Testing for association between multivariate phenotypes and covariates

- Wilks' test (appropriate only when $p < n$)
- Perform PCA on the phenotypes, retain a certain number of components (less than n), perform PCEV on the components, and use Wilks' test
- Permutation test

Type I error rate



Power

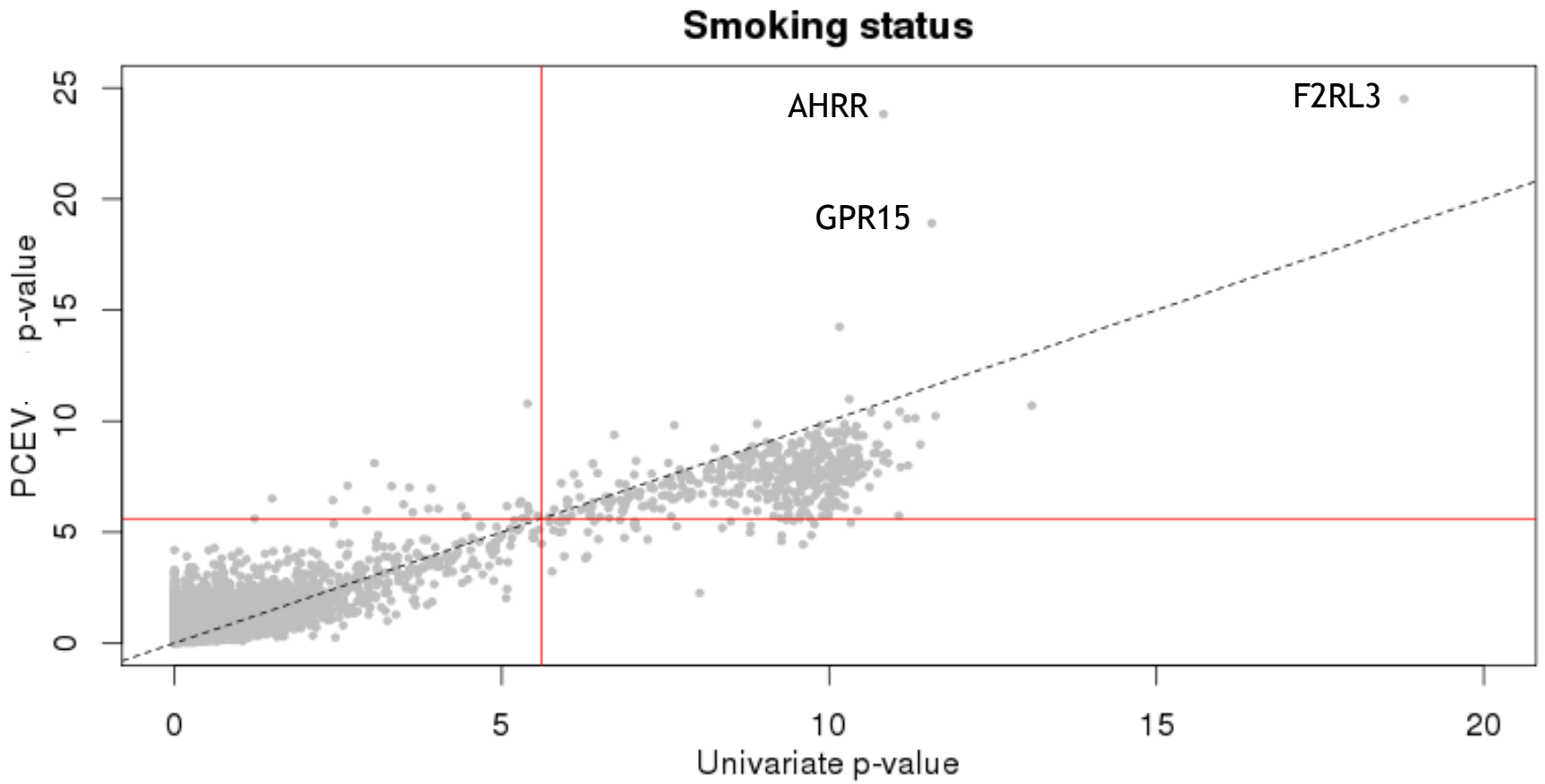


One-stage
Two-stage

Data analysis

- 993 healthy individuals who served as controls for the Assessment of Risk of Colorectal Tumors in Canada (ARCTIC) cohort

- Each gene was analyzed separately
  - Multivariate phenotype: methylation values for CpG sites contained in the gene and in the promoter region
  - Covariate: cigarette smoking (binary)



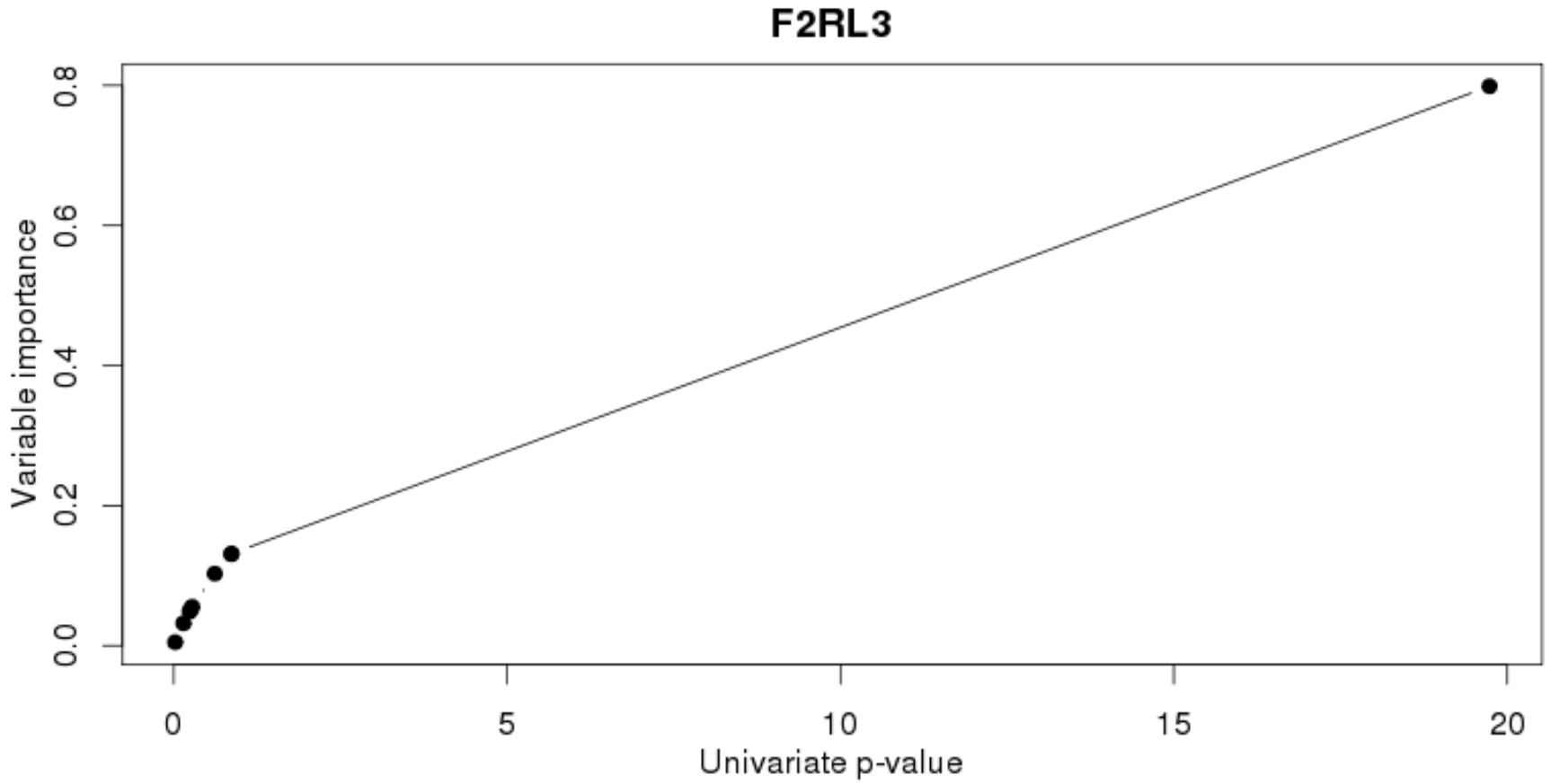
Comparison of univariate approach and PCEV

AHRR, F2RL3 and GPR15 have previously been identified by Bretling et al. (AJHG, 2011)

Variable importance

- We can rank the contribution of each phenotype to the association using *variable importance*:

$$\text{VIMP} := \text{corr}(w^T\mathbf{Y}, Y_j)$$



Variable importance and univariate p-value

The multivariate association of F2RL3 seems to be driven mainly by one CpG site.

Data was provided by:

- Brent Zanke
- Mathieu Lemire
- Thomas Hudson

Thank you

Poster #93
Session A