Multiple SNP analysis with HyperLasso in Pharmacogenetic Studies using Nonlinear Mixed Effects Models

Julie Bertrand and David Balding

University College London Genetics Institute, London, United Kingdom

Context

Pharmacogenetics

Study of the DNA variations on genes coding for proteins involved in drug transport, metabolism, and effect in relation to the inter-individual variability in drug response

- Target
  - selection of metabolic pathways during drug development
  - individualized therapy
  - integration of diversity in population genetics
- Statistical analyses
  - ANOVA-based approach on derived PK parameters
    - loss of information provided by the complete time profile
    - does not account for additional effects or interactions
    - no direct predictions or dosing recommendations

↪ Nonlinear Mixed effect models (NLMEM)

Multiple SNP analysis using NLMEM

- Stepwise-based procedure
  - algorithm proposed by Lehr et al [1]
  - feasibility and potential benefits evaluated in 4 case studies

⇒ classical method with specific features to account for linkage disequilibrium

- HyperLasso (HLASSO)
  - generalisation of the double exponential (or Laplace) prior assumed by the Lasso [2]
  - normal exponential gamma distribution with a shape (λ) and a scale (γ) parameters [3]
  - λ small:

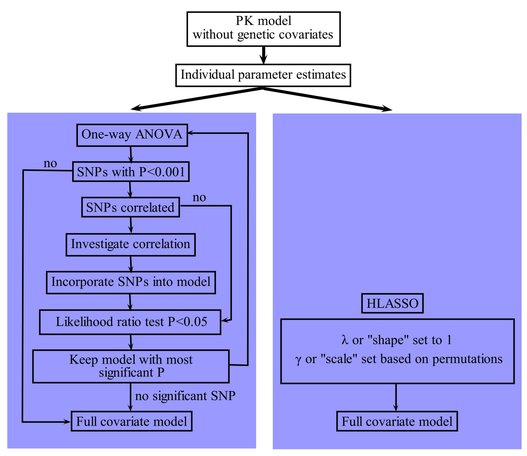
    → sharp peak at zero = sparse solutions

    → heavy tails = variables minimally shrunk once included
  - double exponential recovered with large λ

⇒ statistical method developed in genetics used in conjunction with NLMEM

Objective

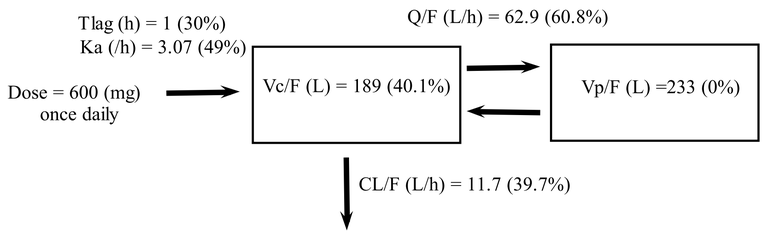
- To assess the power of the stepwise-based procedure and HLASSO for detecting Single Nucleotide Polymorphism (SNP) effects on a pharmacokinetic parameter using NLMEM



Simulation study

Pharmacokinetic settings

- Structural and statistical model
  - inspired from real study [4]



- Phase II-like study design
  - 300 individuals with t= 0.5, 1.25, 2, 4, 9, 24
- Pharmacokinetic modelling performed with SAEM in MONOLIX 3.1

Genetic settings

- Generation of genotypes using HAPGEN [5]
  - HAPMAP caucasian reference haplotypes
  - 1227 snps from the DMET Chip [6]
  - distributed over the 22 autosomes and chromosome X
  - 171 genes with a coverage of 29 [0-804.3] Kb
  - 6 [1-56] snps per gene

- Alternative hypothesis
  - 3 unobserved causal variants with MAF>0.05 randomly chosen
  - SNP_1 and SNP_2
    - decrease in CL/F by 40% associated to the variant allele
  - SNP_3
    - increase in F by 30% associated to the minor allele

  ⇒ decrease by 77% in Vc/F, Vp/F, Q/F and CL/F

Evaluation

- 200 data sets simulated under H_0 and H_1

T = number of simulated data sets
P = number of PK model parameters
SNP = number of causal SNPs
TP = number of True positive
SNPs correlated to the causal variant ($\rho > 5\%$)
FP = number of False positive
SNPs uncorrelated to the causal variant

$$\text{Power} = \frac{\sum_{T=1}^{200}\sum_{SNP=1}^{3}\sum_{P=1}^{3} min(TP,1)}{200}$$

$$\text{False Positive} = \frac{\sum_{T=1}^{200}\sum_{SNP=1}^{3}\sum_{P=1}^{3} FP}{200}$$

Results

A typical simulated dataset

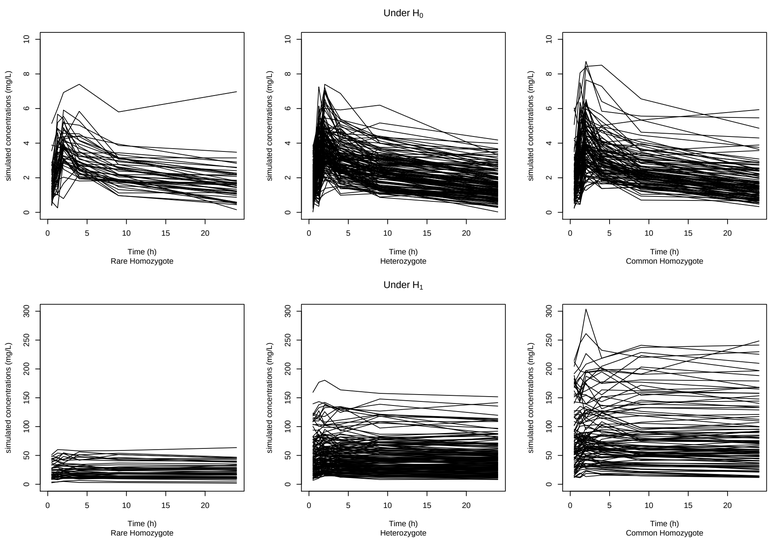


Figure 1: Concentration versus time individual profiles sorted by genotypes for the causal variant SNP_1 under both hypotheses.

- Median=806 and range=[783-834] polymorphic SNPs per data set

Power

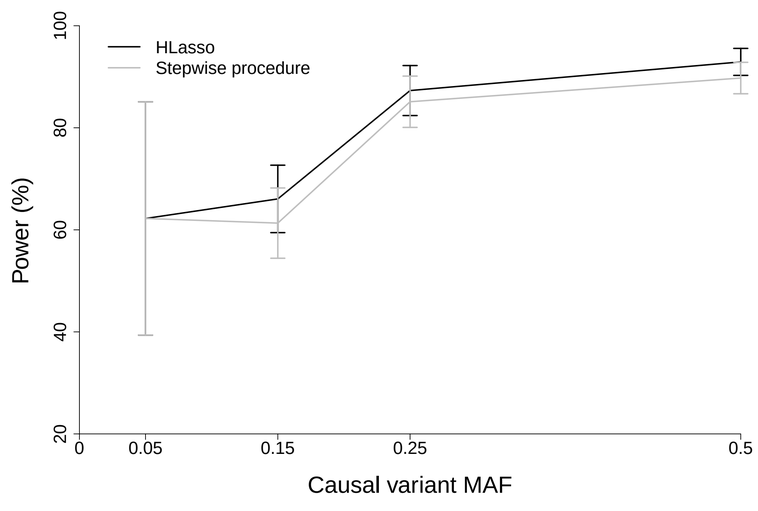


Figure 2: Power estimates and their 95% confidence interval versus the minor allele frequency (MAF) of the causal variant for both algorithms

False positive

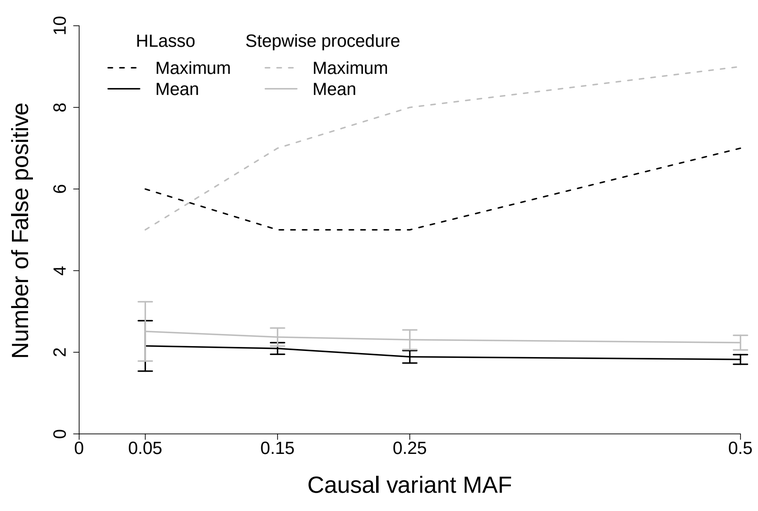


Figure 3: Maximal and mean estimates for the number of false positives (with the 95% confidence interval around the mean) versus the minor allele frequency (MAF) of the causal variant for both algorithms

Discussion

- Similar power of the stepwise-based procedure and HyperLasso
  - increasing with MAF as expected
- Reasonable number of false positives
  - trend of the maximal number of false positive shown with the MAF to explore
- Important gain in computing time with HyperLasso
  - median=0.1 h and range=[0.09-0.21] versus 15 h [0.5-66] for the stepwise procedure under the alternative hypothesis
- On-going work
  - to increase the number of causal variants (10-15)
  - to consider moderate to weak effects (gradient)

References

[1] Thorsten Lehr, Hans-Guenter Schaefer, and Alexander Staab. Integration of high-throughput genotyping data into pharmacometric analyses using nonlinear mixed effects modeling. *Pharmacogenetics and Genomics*, 20(7):442–450, 2010.

[2] R Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B Methodological*, 58(1):267–288, 1996.

[3] Clive J Hoggart, John C Whittaker, Maria De Iorio, and David J Balding. Simultaneous analysis of all SNPs in genome-wide and re-sequencing association studies. *PLoS Genetics*, 4(7):e1000130, 2008.

[4] B. S. Kappelhoff, A. D. Huitema, Z. Yalvac, J. M. Prins, J. W. Mulder, P. L. Meenhorst, and J. H. Beijnen. Population pharmacokinetics of efavirenz in an unselected cohort of HIV-1-infected individuals. *Clin Pharmacokinet*, 44:849–861, 2005.

[5] J Marchini, B Howie, S Myers, G McVean, and P Donnelly. A new multipoint method for genome-wide association studies by imputation of genotypes : Supplementary Methods. *Nature*, 2(2):906–913, 2007.

[6] Thomas M Daly, Carmen M Dumaual, Xin Miao, Mark W Farmen, Reuben K Njau, Dong-Jing Fu, Nancy L Bauer, Sandra Close, Nancy Watanabe, Carsten Bruckner, Paul Hardenbol, and Richard D Hockett. Multiplex assay for comprehensive genotyping of genes involved in drug metabolism, excretion, and transport. *Clinical Chemistry*, 53(7):1222–1230, 2007.

During this work Julie Bertrand was funded by a grant from the French National Research Agency on HIV and Viral Hepatitis (ANRS)