## A BAYESIAN MODEL IN GENETIC ASSOCIATION STUDIES

Juan-Ramón González[1,3], Geòrgia Escaramís[1,2], Rafael De Cid[2], Francesc Castro[1,3], Manolis Kogevinas[1,3], Xavier Estivill[1,2,4,5]

[1] CIBER en Epidemiología y Salud Pública (CIBERESP), Spain,
[2]Genes and Disease Program, Center for Genomic Regulation (CRG), Barcelona, Spain,
[3]Center for Research in Environmental Epidemiology (CREAL), Barcelona, Spain,
[4]Experimental and Health Sciences Department, Pompeu Fabra University, Barcelona, Spain,
[5]National Center of Genotyping (CEGEN), Barcelona Node, Barcelona, Spain

Disease association studies with single-nucleotide polymorphisms (SNPs) have been widely applied in the genetic literature to search for genes that confer susceptibility for the development of a particular disease. In genome-wide association studies usually one evaluates hundreds of thousands of SNPs across the genome and then follows up the most-promising SNPs. Additionally; existing information about the SNPs, such as gene structure, haplotype blocks or linkage disequilibrium patterns, among others, can be added into the analysis for more accurate results.

Conti and White (2003) and Chen and Witte (2007) proposed a hierarchical modelling framework that simultaneously combines various types of a priori information. The authors proposed a first stage using conventional generalized linear models (GLM) to obtain the Generalized Least Squares (GLS) estimates of the specific SNP effects and in a second stage they use the effect estimates as dependent variates to a simple linear function of SNPs specific covariates and giving a certain correlation structure to the resulting covariance matrix. They furthermore obtain a more accurate SNP effect estimate via a variance-weighted average of the first and second-stage estimates. The authors argue the use of Wald statistics for SNP effect inference. However, since the final estimator is not asymptotically unbiased, Wald procedures may lead to wrong conclusions. We propose here a full Bayesian approach via Markov Chain Monte Carlo algorithms to obtain a posterior distribution of the SNPs effect and then construct credible intervals for inference. This method allows in addition to jointly model first and second stages. To accommodate the spatial dependence among SNPs, the second stage model includes a vector of random effects with certain prior distributions. We consider the exponential decay function (Wakefield et al., 2000), also considered in the work of Conti and White (2003), and the conditional autoregressive model (CAR; Besag, York and Mollié, 1991), widely applied in spatial disease mapping. The latter approach is computational convenient but rather appropriate for equally spaced SNPs.

We will illustrate our approach by analyzing a population-based multicentric cohort study including genetic information about five candidate genes related to asthma.

**Bibliography**

Besag, J., York, J. and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Ann. of the Institute of Statistical Mathematics* **43**, 1-21.

Chen G.K. and Witte, J.S. (2007). Enriching the analysis of Genomwide Association Studies with Hierarchical Modeling. *Am. J. Hum. Genet* DOI: 10.1086/519794.

Conti, D.V. and Witte, J.S. (2003). Hierarchical Modeling of linkage Disequilibrium: Genetic Structure and Spatial Relations. *Am. J. Hum. Genet* **72**, 351-363.

Wakefield, J.C., Best, N.G. and Waller, L.A. (2000). Bayesian approaches to disease mapping. In: Elliot, P., Wakefield, J.C., Best, N.G. and Briggs D.J. (eds) *Spatial epidemiology: methods and applications*. Oxford University Press, Oxford, pp 104-127.