1. **Introduction (**Accurate Estimation of Quantitative Trait Locus Effects

with Epistatic by Improved Variational Linear Regression **)**

The goal of quantitative trait loci (QTLs) mapping and association studies is to identify certain regions of the genome that contain genes involved in specifying a quantitative trait, and to estimate the genetic effects of these loci. The relationship between the genetic effects of QTLs and the phenotypic value of quantitative traits can be described by a linear model [1, 2]. There are usually a large number of markers across the whole genome, and most of the markers have very little or even no effect on the phenotypes. So the model is sparse, and in most cases, the number of markers or variables is bigger than the sample size, especially when interactions among markers are considered. Therefore, the model is oversaturated and usually solved by shrinkage or variable selection methods. The variable selection procedure for QTLs mapping can be seen as one of deciding which subset of variables have effects on phenotypes, and identifying out all possible effects of those markers. There are many ways of variable selection. For instance, the indicator selection uses a spike (probability distribution of variables not included in model) and slab (probability distribution of variables included in model) prior to indicate whether each variable is included in one model. Such methods usually adopt Gibbs sampling or importance sampling to calculate the weighted average as the inferred results [3-5]. Shrinkage methods do not use indicators to induce the sparseness of variables, but instead specify a priori directly on variables to approximate the “slab” and “spike” shapes and use a penalty function to shrink most variables toward zeros. Early Bayesian shrinkage methods employed Markov chain Monte Carlo (MCMC) techniques to infer the models [6], which were computationally intensive. Xu et al. proposed an empirical Bayes method that used a carefully chosen prior distribution for the variables, which required much less

**4. Conclusions and discussion** The main goal of this paper is to propose an improved variational linear regression approach for high scale variable selection problems such as the ones arising in epistatic analysis. Variational approximation is a classic method to solve linear model, but its accuracy is limited in the case of there are many more variables than samples. Several improved methods were proposed for the variational approximation, such as hierarchical shrinkage, importance sampling and so on. But these methods still lack of efficiency or accuracy. We proposed a simple but effective method to improve the performance of VLR by dynamically deleting some specific variables, which can greatly improve the accuracy of VLR. The reason that RCVLR is more accurate than VB-EBL, VB-BAS, VB-BL may lie in that it can derive the approximate marginal distributions of all variables in a single process, while hierarchical shrinkage infers each variable separately, therefore some correlation information among variables is lost. The VI-IMP considers the correlation among variables, but it uses important sampling and a weighted average method to improve the accuracy; therefore the computational cost is higher than other variational approaches. RCVLR can obtain good performance in general, and is not sensitive to the choice of hyper parameters, while some other variational approaches, such as VB-EBL, rely heavily on selecting appropriate parameters to get good performance [18]. VI-IMP averages the inferred results under different prior combinations. But how to determine the prior ranges is a problem unsolved [4]. If an inappropriate range is inputted, the performance may be poor. But if a very wide range is set to cover the exact parameters, the algorithm will need many more samples and thus need much more computational time.

**ENAR Abstract**

Knowledge about how changes in gene expression are encoded by expression quantitative trait loci (eQTLs) is a key to construct the genotype-phenotype map for complex traits or diseases. Traditional eQTL mapping is to associate one transcript with a single marker at a time, thereby limiting our inference about a complete picture of the genetic architecture of gene expression. In this talk, I present an innovative application of variable selection approaches to systematically detect main effects and interaction effects among all possible loci on differentiation and function of gene expression. Forward-selection-based procedures were particularly implemented to tackle complex covariance structures of gene-gene interactions. We reanalyzed a published genetic and genomic data collected in a mapping population of *Caenorhabditis elegans*, gaining new discoveries on the genetic origin of gene expression differentiation, which could not be detected by a traditional one-locus/one-transcript analysis approach.