Using GPUs to explore the genetic architecture of complex traits

Phenotypic variation arises partly because of environmental influences, and partly due to genetic differences. Most traits of interest, for example common human diseases, are caused by a complex pattern of both. The human genome is a code consisting of 3 billion base pairs and from individual to individual most of these are identical, differing on average once every thousand bases. A very hot topic in genetics now is trying to map these polymorphisms to phenotypic variation.

A central question relating to this search is how exactly do mutations affect the phenotype? Most studies assume that each mutation acts independently from all others, and this leads to a computationally straightforward task of testing for association at each polymorphism in the genome with the phenotype. But theory provides an alternative hypothesis such that the effect of each causal mutation actually depends on which mutations are present elsewhere in the genome. Such genetic interactions (epistasis) are, in principle, much more difficult to identify than independent mutations because the problem passes into a higher dimensional parameter space. For this reason, although there is theoretical support for the importance of epistasis, it is seldom searched for empirically, even though the data to do so is abundantly available.

If we can understand how mutations are acting - whether that be independently or as part of complex interactions - then this will be useful for many aspects of biology and medicine including the mapping of causal mutations to phenotypes, developing more accurate disease risk predictors, and understanding the mechanisms that underlie evolution.

We recently developed software that searches for epistasis exhaustively by parallelising the analysis on GPUs, using the OpenCL API {\tt epiGPU}. It has increased the search speed by more than 3 orders of magnitude, thus making this type of analysis tractable.

We have a rich dataset of RNA expression profiles for around 1000 individuals. We propose to search for interactions between genetic mutations in these individuals that might account for the differences in the expression of each of these RNA transcripts. In total, there are approximately 15000 RNA transcripts, and each transcript will take approximately 2 hours to analyse using {\tt epiGPU}, so access to a GPU cluster would make such a study possible.

This type of analysis has never been done before, and it would provide an extremely thorough survey of the way in which mutations contribute towards complex traits. If this analysis fails to find much epistasis then this will be an important result in itself - providing empirical evidence in a continuing debate about the architecture of genetic variation. On the other hand if we discover that a large number of the polymorphisms associated with the trait do not act independently and are indeed involved in complex interactions then this will be the first empirical evidence for widespread epistasis.