Contribution ID: **134** Type: **Oral**

Genome Wide Association Studies of human complex diseases with EGEE

Tuesday, February 12, 2008 11:40 AM (20 minutes)

As part of the research conducted at the INSERM U525 laboratory, the THESIAS software was created in order to analyze statistically, associations between gene polymorphisms and diseases. Given a data set containing the genotypes of case and control individuals, THESIAS measures haplotype frequencies combining several polymorphisms and associations with the disease. Until now this kind of analysis was restricted to single genes and a few polymorphisms (<25). The recent availability of DNA chips allowing to genotype hundreds of thousands of polymorphisms across the genome implies a change in scale in the necessary computations. For whole genome haplotype analysis we decided to use the EGEE grid.

3. Impact

Identifying which DNA sequences variations(SNPs) are associated to a disease on the entire human genome has a complexity which increases exponentially with the number of SNPs. Frequencies of combinations of multiple SNPs must be estimated and ideally all the possibilities would be analyzed. However, there are at least 10 millions SNPs on the human genome and calculating all the combinations is hardly imaginable. Fortunately, SNPs located close to each other (for example within a gene) are frequently tightly correlated, they are said to be in linkage disequilibrium (LD) and they define haplotype blocks that can be tagged by a limited number of
marker-SNPs. The most recent genotyping arrays contain 1 million marker-SNPs and are highly informative. Computational burden may be further reduced by investigating haplotypes (sets of closely linked SNPs) in a sliding window. This research can lead to the identification of new causes and mechanisms of disease of potential therapeutic interest.

4. Conclusions / Future plans

As a proof of principle, we have analyzed thousands of SNPs for their association with cardiovascular disease in thousands of individuals. Easy-gLite, a UI on top of the gLite UI has been created to simplify batch job submissions, monitoring and automatic resubmission of failed jobs. We will soon use EGEE on analysing the whole genome, with about 500000 SNPs, which is at least 50 times more important than our last analyses.

Provide a set of generic keywords that define your contribution (e.g. Data Management, Workflows, High Energy Physics)

genome wide, genome, SNP, EGEE, association studies

1. Short overview

Until now, associations analyses between gene polymorphisms and diseases was limited to a few number of polymorphisms because those analyses require much computational power. The EGEE grid provides enough computation power for analysing the whole human genome. The following describes the THESIAS program created for this research, but also how we have used EGEE with this software.

Primary author: Mr MUNTEANU, Alexandru Ionut (INSERM, UMR S 525, Faculté de Médicine Pitié-Salpêtrière, Paris, France)

Co-authors: Mr TRÉGOUËT, David (INSERM, UMR S 525, Faculté de Médicine Pitié-Salpêtrière, Paris, France); Mr CAMBIEN, François (INSERM, UMR S 525, Faculté de Médicine Pitié-Salpêtrière, Paris, France)

Presenter: Mr MUNTEANU, Alexandru Ionut (INSERM, UMR S 525, Faculté de Médicine Pitié-Salpêtrière, Paris, France)

Session Classification: Life Sciences

Track Classification: Application Porting and Deployment