**Small tutorial of KGG for Gene-, Protein-protein interaction- and Pathway-based association analysis**

Miaoxin Li ( mxli@hku.hk)

**Input data:**

1. Meta-analysis test statistics of Crohn's disease originally downloaded from (http://www.broadinstitute.org/~jcbarret/ibd-meta/)

CrohnMetaPValueSetHg17Chr1.txt

1. Known risk gene list of Crohn's disease

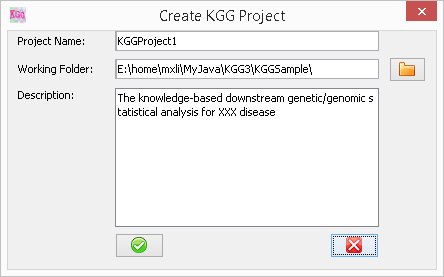
CrohnCandidateGeneSet.txt

**Reference data to account linkage disequilibrium (LD) from 1000 Genomes Project**

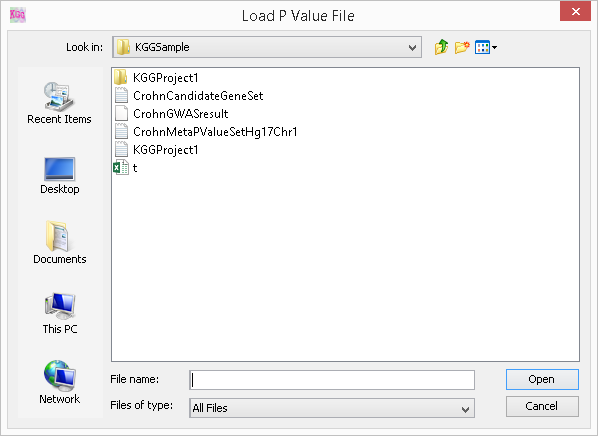
1kg.phase1.v3.shapeit2.eur.hg19.chr1.vcf.gz

**Operations:**

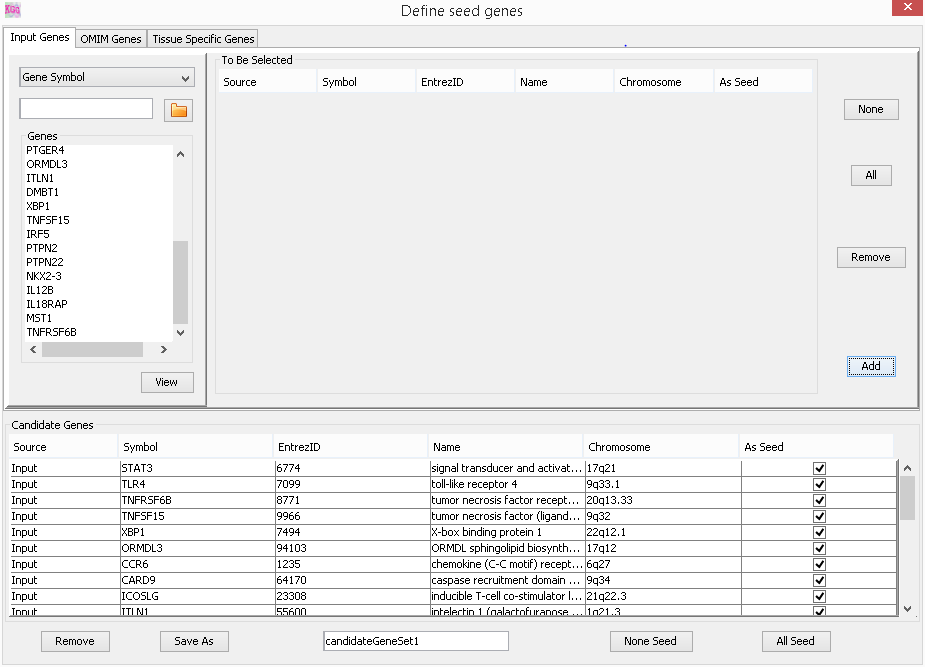
1. Copy the **KGGSample** folder of the miaoxin/IBG2013/KGG/ into your folder.
2. Create a KGG project : **Project->Create Project**



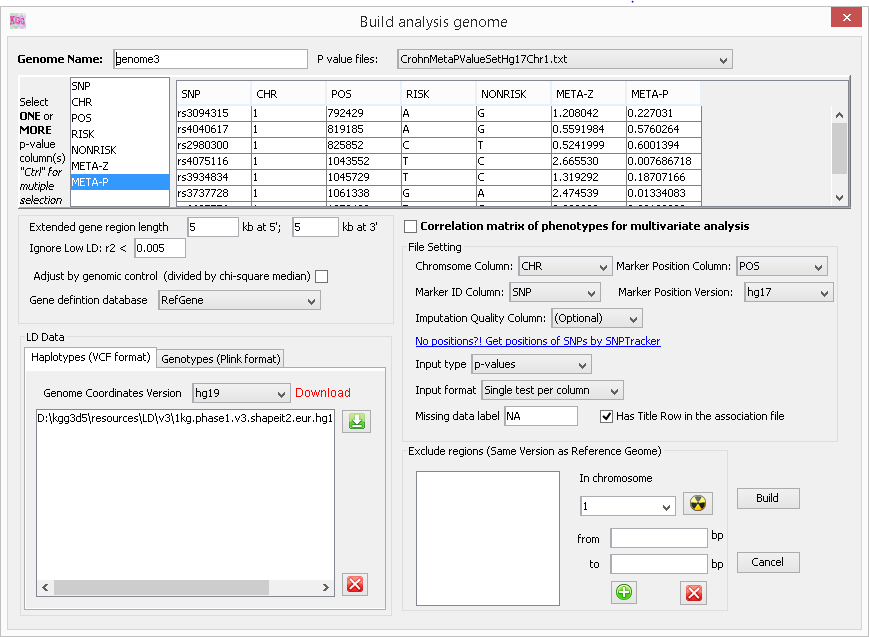
1. Import original association file (load statistics or p values): **Data->Load P value file**



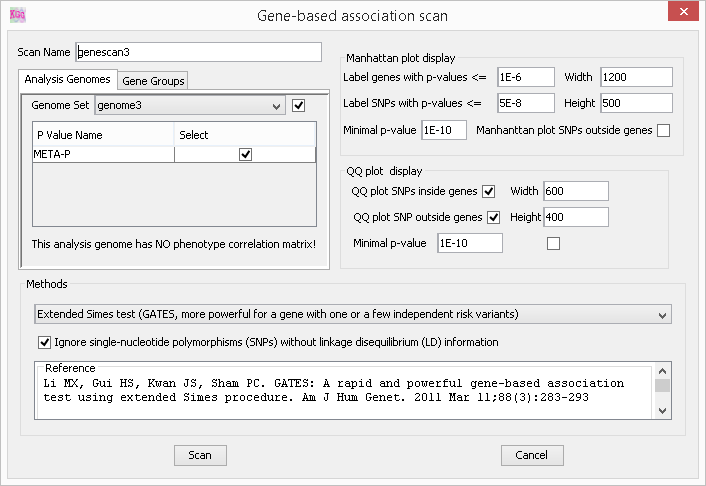
1. Load candidate genes (Susceptibility genes almost established for a disease): **Data->Define Candidate Genes**



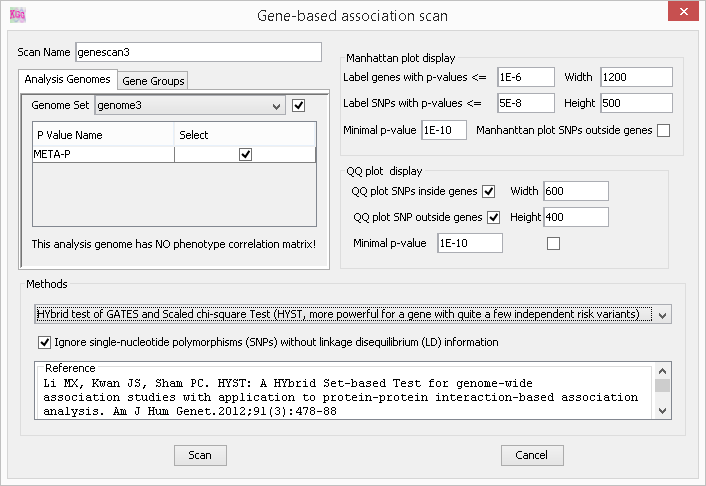
1. Build an analysis Genome(~2min): **Data->Build analysis genome**



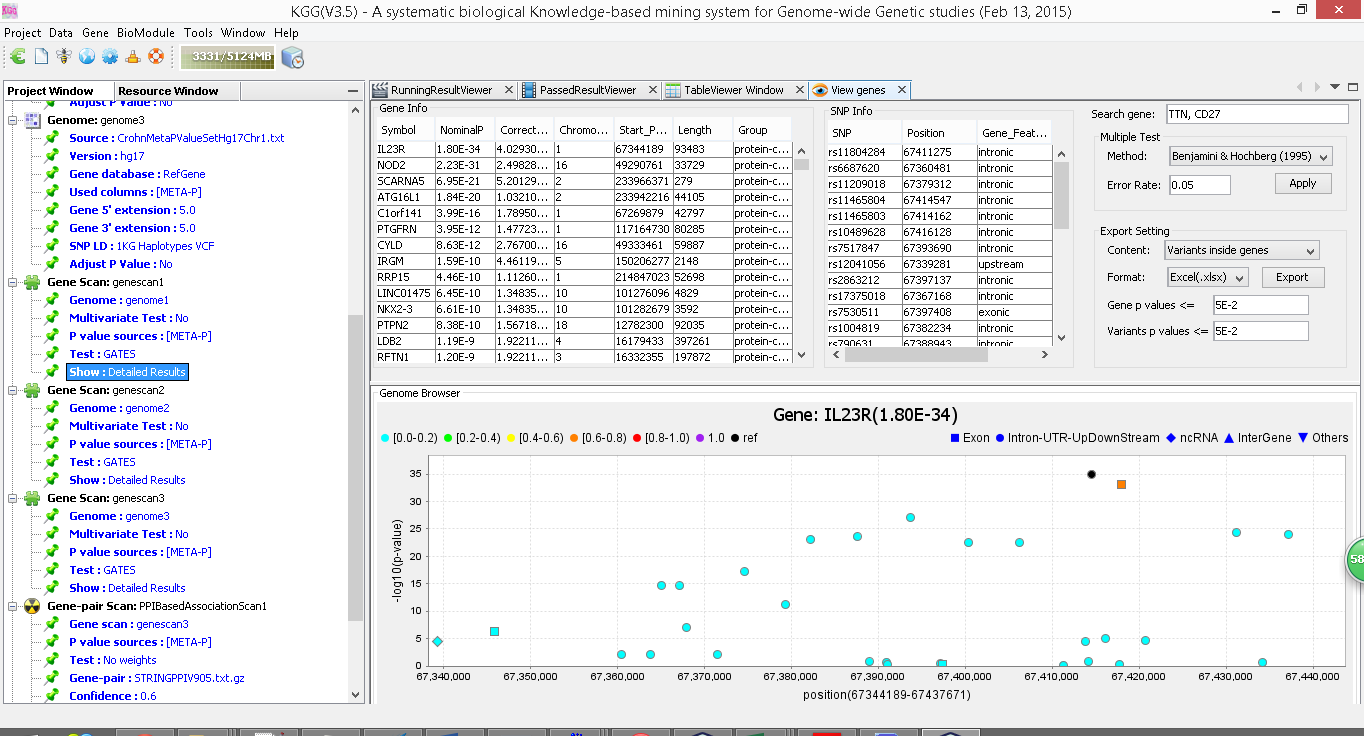
1. Gene-based association analysis by GATES (~10sec.) **Gene->Univariate association test**



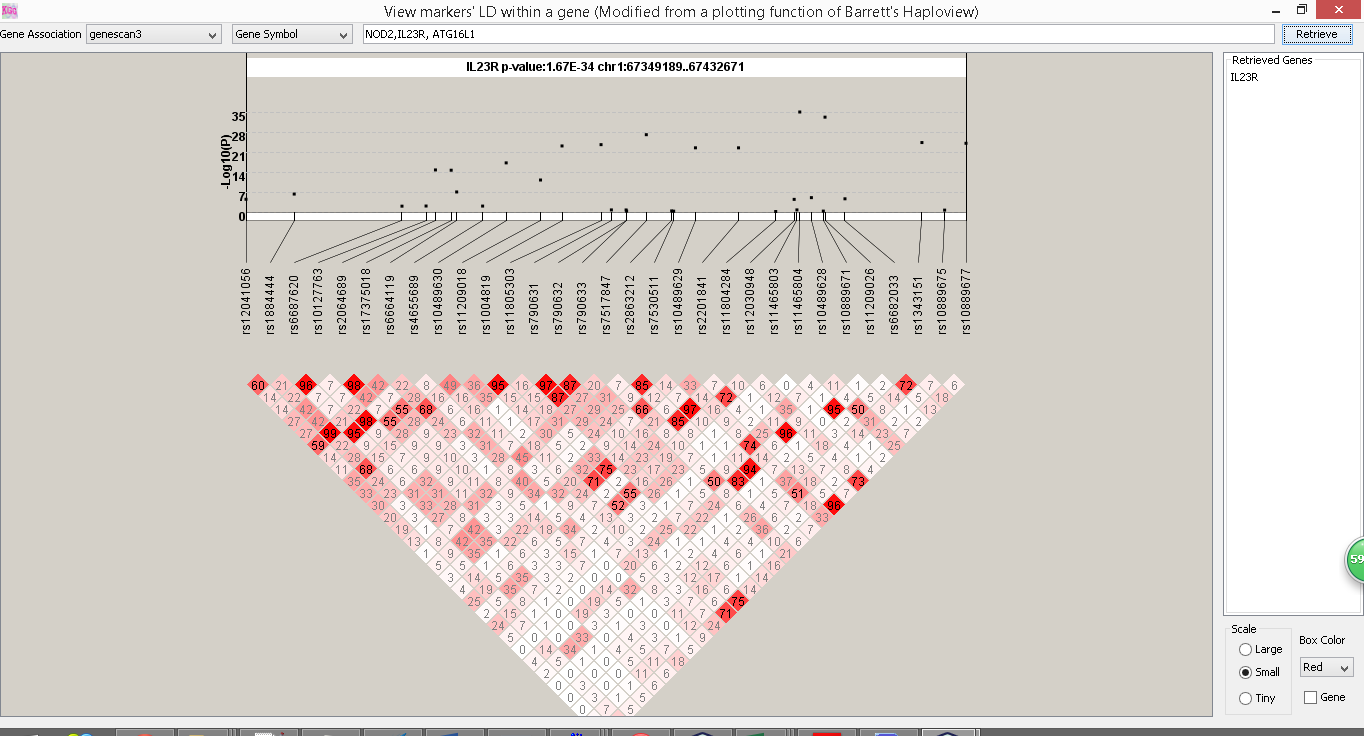
1. Gene-based association analysis by HYST(~10sec.) **Gene-> Univariate association test**



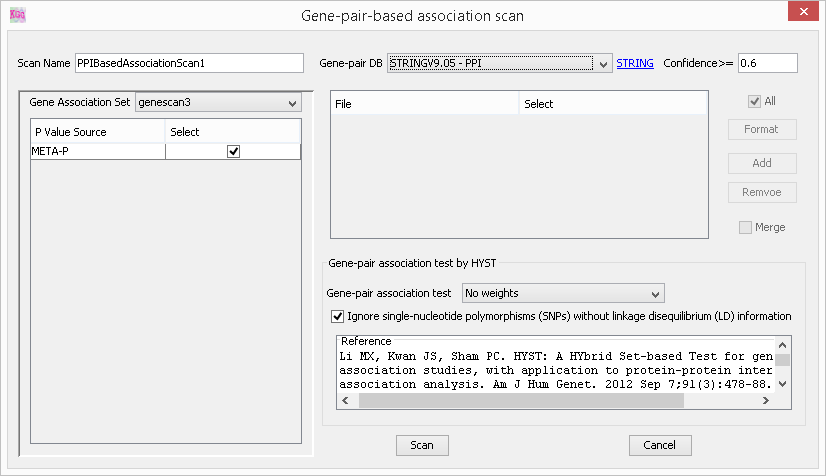
1. View and export gene-based association p-values and their SNPs: ***At each gene scan* -> Show: Detailed Results**



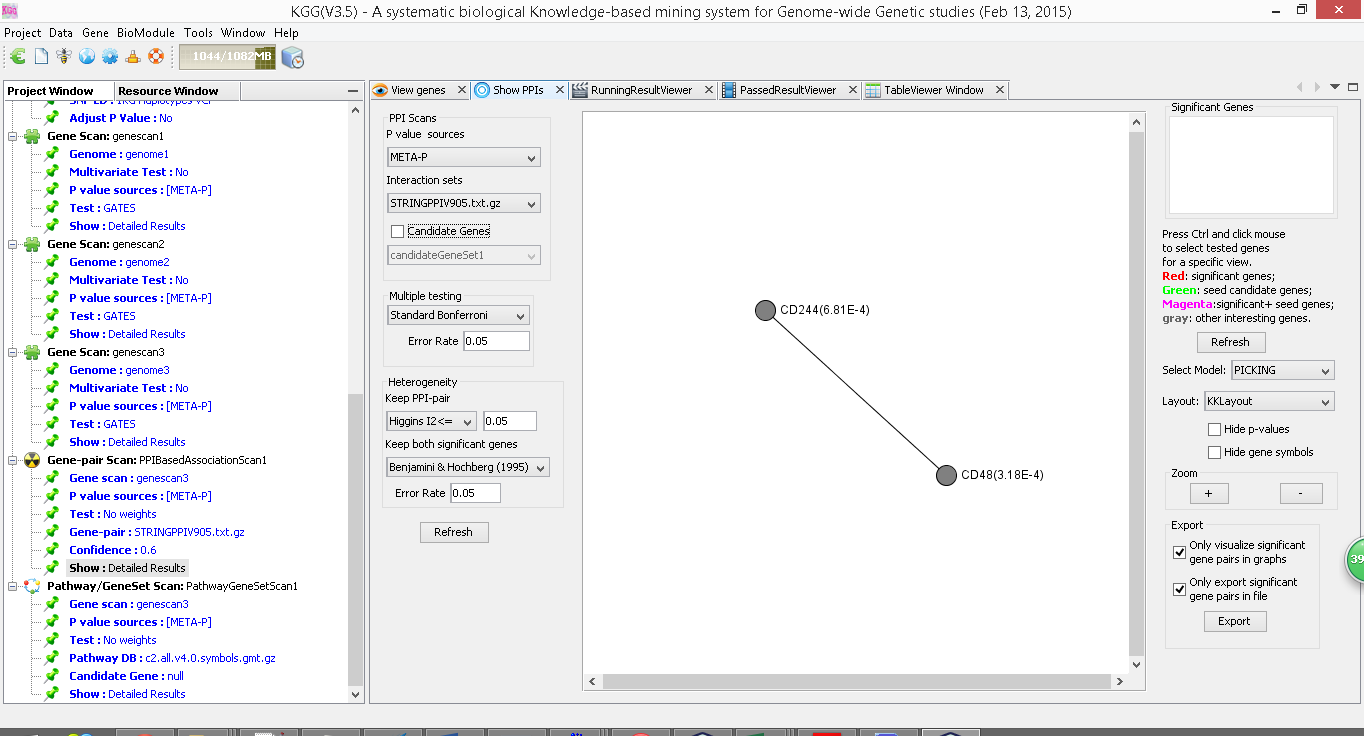
1. View LD structure of SNPs within a gene:  **Gene->LD Plot**



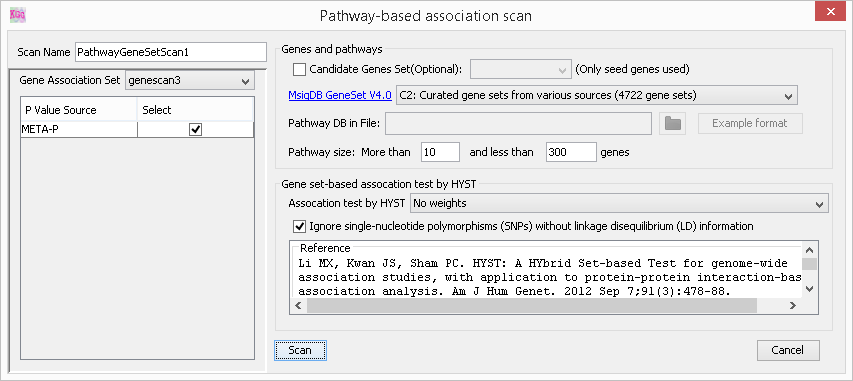
1. Protein-protein interaction-based association analysis by HYST (~1 min.) **BioModule->Gene-pair-based association scan**



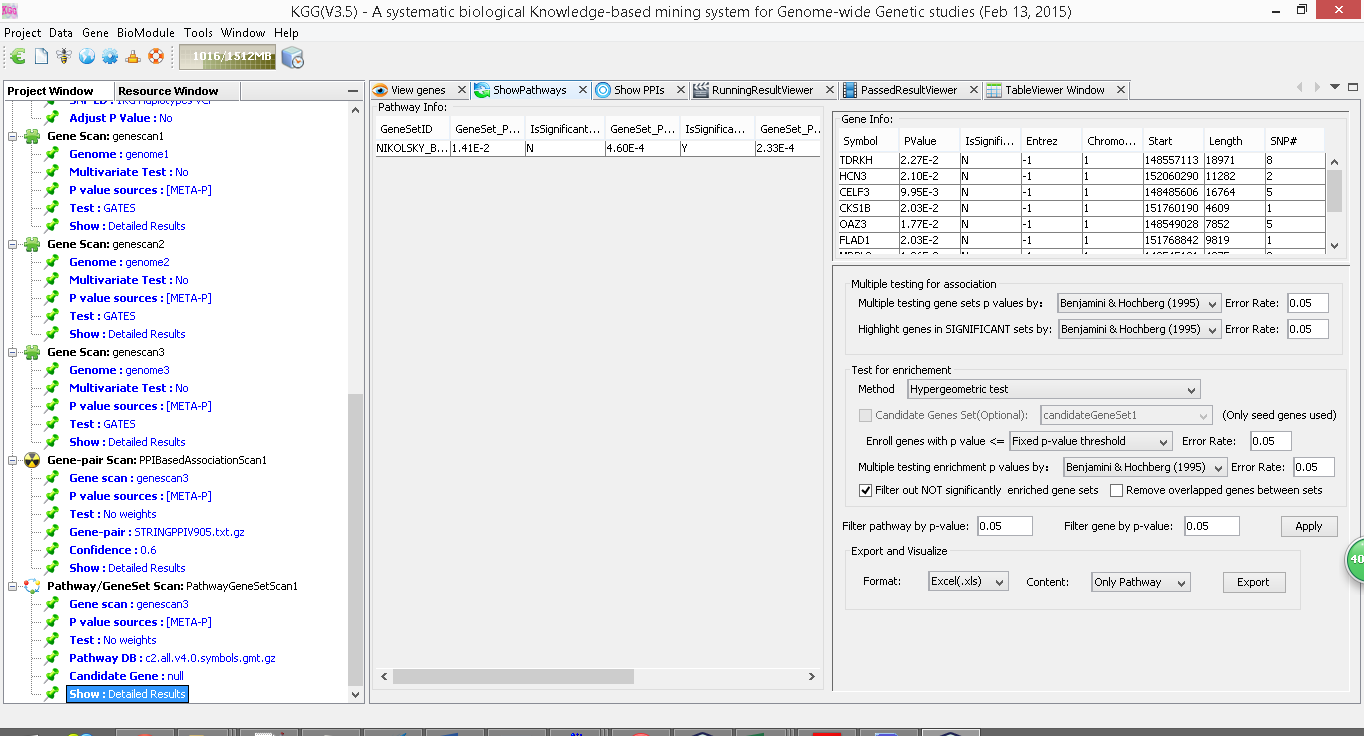
1. View protein-protein interaction-based association results ***At each Gene-pair Scan* -> Show: Detailed Results**



1. Pathway gene set-based association analysis by HYST (~1 min.) **BioModule ->** **Pathway-based association scan**



1. View pathway-based association results ***At each Pathway/GeneSet scan* -> Show: Detailed Results**



**Multivariate gene-based association test**

**Input data:**

Nine quantitative metabolic traits measured in the population-based NFBC1966 (N=4763; Sabatti et al., 2009), raw genotypes obtained from dbGAP were to calculate p-values of each SNP for each trait.

1. 9MetabolicPhenotypesPhg19.txt

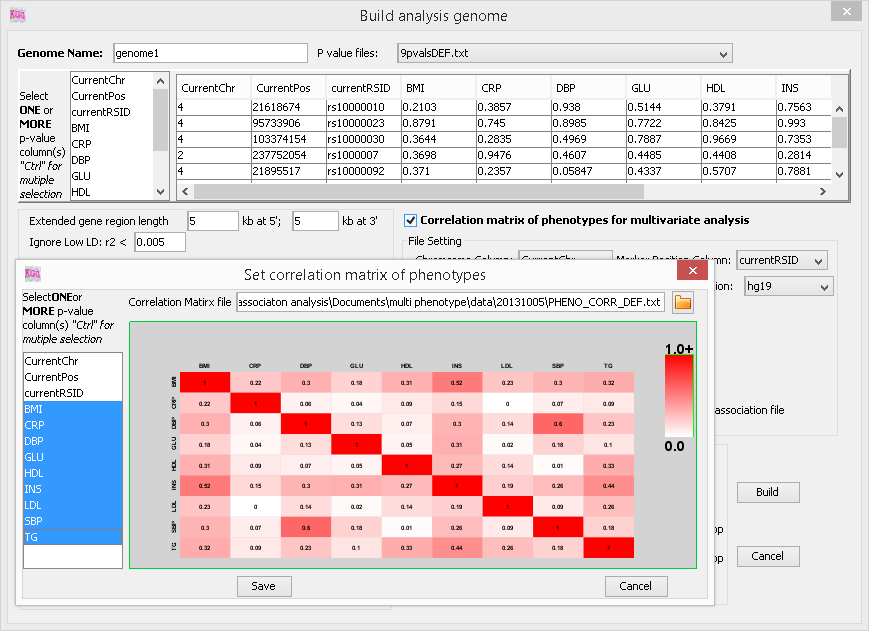
Correlation matrix of the 9 traits

1. 9MetabolicPhenotypesCorr.txt

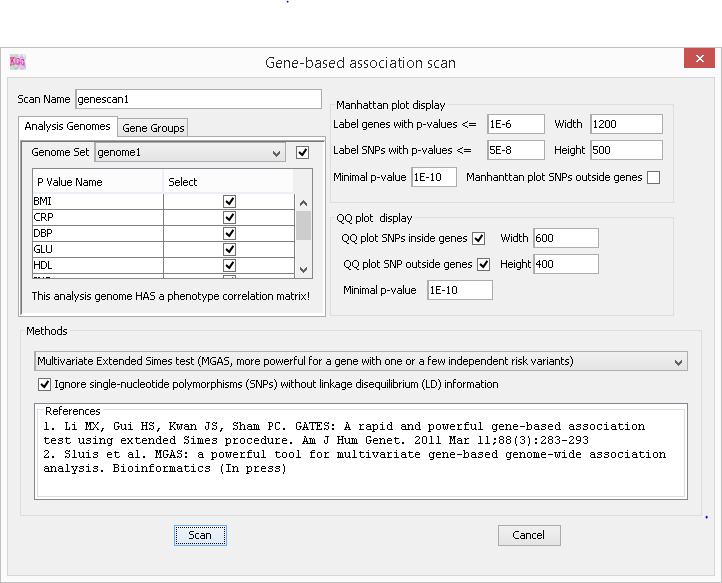
**Reference data to account linkage disequilibrium (LD) from 1000 Genomes Project**

1kg.phase1.v3.shapeit2.eur.hg19.chr[1..X].vcf.gz

1. Build analysis genome with phonotype correlation **Data->Build analysis genome**



1. Multivariate gene-based scan **Gene->Univariate association test**



**Estimate the power of set-based tests**

1. Estimate the power of set-based tests: **Tools->Gene-pair-based association scan**

