SNP Data Analysis for Study of Complex Diseases Caused by Multi-Gene Disorders (YW2-16)

YUNG, Wai Pan

Supervisor:
Professor Yu Weichuan

Introduction

To increase the utilization of CPU and thus improve the CPU efficiency throughout the analysis.

To adopt a non-iterative pre-screening search method to filter out non-significant SNPs.

To propose a Cloud-Computing theory as a future work to speed up the whole analysis process.

Objective

- To increase the utilization of CPU and thus improve the CPU efficiency throughout the analysis.
- To adopt a non-iterative pre-screening search method to filter out non-significant SNPs interaction.
- To propose a Cloud-Computing theory as a future work to speed up the whole analysis process.

Methodology

1. Boolean Representation
A Boolean representation of the genotype of a sample is used as it allows more efficient utilization of the CPU. For example, a 64-bit CPU can perform 64 AND operations just in one instruction.

2. KSA pre-screening
Kirkwood Superposition Approximation (KSA) can act as a quick pre-screening to filter out all non-significant SNPs as it has a closed-form solution which can consistently overestimate the true result.

3. Likelihood Ratio Test
Supposed: \( \log \frac{p}{1-p} = \beta_0 + \beta_1 A + \beta_2 B \)
Reality: \( \log \frac{p}{1-p} = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 AB \)

Interaction = Reality - Supposed

System Block Diagram

Result

To compare our performance with other existing Gene-Gene Interaction detection methods. Our method is approximately 100 times faster than INTERSNP; 60 times faster than PLINK and 170 times faster than BEAM. The result becomes more significant when a larger sample size is considered.

Future Work

To better cope with a genome-scaled project, our method can be further implemented on the Cloud by the MapReduce mechanism with an Infrastructure-as-a-Service Provider (IaaS).