SNP-Analyzer: 
Analysis of SNP Microarray Data

Guzzi PH, Agapito G, Cannataro M
Department of Experimental Medicine and Clinic, University Magna Graecia of Catanzaro

Motivation
Motivation: Each individual has a unique sequence of DNA that determines his/her characteristics. Differences among individuals or populations can be measured in terms of substitutions of bases in the same position. Researchers focused in particular on the substitution of a single base that occurs in a small subset of the population. These mutations, also referred to as single nucleotide polymorphisms (SNP’s), are usually defined as a stable substitution of a single base with a frequency of more than 1% in at least one population. For instance, let us consider the short sequences ATGT and ACGT, a base change occurs at position 2 and is denoted by T/G. Many works demonstrated a correlation between the presence of SNPs and the development of diseases, as well as the effectiveness of drugs. Thus the presence (or the absence) of specific SNPs may be used as a clinical marker for the prediction of drug effectiveness, foreseeing the response of individuals with different SNPs to drugs. Pharmacogenomics experiments involve the gene sequencing and the individuation of SNPs by using microarray technology and computational analysis. The DMET (drug metabolism enzymes and transporters) Plus Premier Pack is a novel microarray assay developed by Affymetrix for gene profiling designed specifically to test drug metabolism. DMET is able to genotypize function variant in a defined set that comprise 225 ADME-related genes, i.e. genes known to be related to drug absorption, distribution, metabolism and excretion (ADME). Different recent works demonstrated the roles of genetic variations in ADME genes in association with the heterogeneity in drug treatment effects.

Methods
Data produced by DMET platform must be preprocessed and analyzed in order to find correlation between the presence/absence of SNPs and the status of samples (e.g. type of drug treatment/response). To the best of our knowledge, existing software tools, e.g. the DMET Console platform, generally allow only the preprocessing of binary data and simple data analysis operations, but do not allow to test the association of the presence of SNPs with the response to drugs (or in general with sample conditions). Consequently, researches have to export and manually process SNPs tables produced by the DMET Console. Another challenge in analyzing such data is the huge volume of raw data involved in typical clinical studies where samples belonging at least into two classes (e.g. diseased vs health) are analyzed. Each DMET chip, used to analyze a sample, has about two thousand gene-specific probes each one is able to detect all possible SNPs.
(about thirty values). In studies involving hundreds or thousand patients, this results in huge datasets. Researchers need to test, for each probe and for each couple of SNPs, if any difference among SNPs distributions in the two classes is statistically significant (e.g. by using the well known Fisher test).

**Results**

We developed SNP-Analyzer, a novel algorithm and its first software prototype for the automatic statistics test of the association between SNPs and sample conditions. The goals of the proposed system are: (i) to automatize the workflow of analysis of SNP data avoiding the use of multiple tools; (ii) to optimize the execution of statistical test on large datasets. SNP-Analyzer is a platform-independent software built in Java that supports the statistical analysis of DMET-based pharmacogenomics studies. It has a simple graphical user interface that allows users (doctors/biologists) to upload and analyse DMET files produced by DMETConsole in an interactive way. The proposed system adopts an efficient ad-hoc data structure to manage the huge volume of SNPs data produced in typical clinical experiments and to reduce the number of statistics tests. The system has been tested and validated on a clinical SNP dataset produced in the University of Catanzaro University Hospital and is available on-line for non commercial use.

**Availability**

http://bioingegneria.unicz.it/SDDA

Contact email
hguzzi@unicz.it