



Multigenic Dissection of Complex Traits

Blanca Himes^{1,2}, Amanda Sedgewick², Scott Weiss³, Marco F Ramoni^{1,2}

¹. Harvard Partners Center for Genetics and Genomics, Harvard Medical School; ²Division of Health Sciences and Technology, Harvard University and Massachusetts Institute of Technology; ³. Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School;

ABSTRACT

With the emerging genome-wide perspective, today genetic research is shifting its focus from the traditional high-penetrance/low-incidence diseases to the far-more challenging low-penetrance/high incidence ones. This shift holds the promise to uncover the genetic bases of traits shared by large portions of the population and to change the way in which medical practice is able forecast, prevent and manage chronic and late-onset diseases. This new perspective, however, challenges traditional genetic methods and require the development of novel design and analysis techniques, able to tackle complex phenotypes, epistatic interactions, and multigenic traits. This poster describes these emerging methods to design genomic studies and analyze their results using two examples: a late-onset disease – stroke – and a chronic disease – asthma. In particular, it will describe the novel biomedical informatics methods required for the design and analysis of a study to identify the complex genomic profile underpinning the risk of stroke in sickle cell anemia patients and the integration of multiple genes to track down the genomic profile of adult-onset asthma. It describes the novel mechanism for sharing and distributing genomic data using as an example the Human Variation Omnibus, a comprehensive genome/phenome database under development at Harvard Medical School.

Supported by grant U54LM008748.

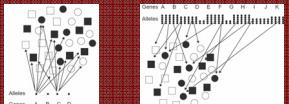
OVERVIEW

The identification of the genetic bases of common diseases requires the development of novel analytical tools for genetic data. Common diseases, by definition, defy some of basic rules of genetic analysis: they tend to be late on-set, posing limited burden on the reproductive changes of an individual; they have high incidence, and therefore their genetic bases are regarded as having low penetrance.

The Common diseases/common variants hypothesis states that, since chronic and late-onset diseases do not affect fitness, a variant will reach some equilibrium in the population and remains in the population. This hypothesis saves the idea that we can actually still search for one single gene, because each gene will individually account for some portion of the disease in the population.

The goal of this project is to develop a methodology able to account for interacting genes and epistatic phenomena in the identification of the genetic bases of common diseases.

The methodological framework we use are Bayesian networks and the applications we describe here are to stroke and asthma.



METHODS

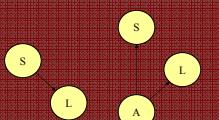
The basic methodology of our approach is known as Bayesian networks. Bayesian networks are a representation formalism born at the intersection of Artificial Intelligence and statistics.

A Bayesian network is defined by a directed acyclic graph and a set of conditional probability distributions.

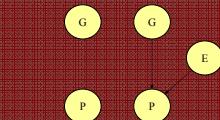
A Bayesian network can be extracted directly from data (both in its topology and in its distributions).

A Bayesian network can be used to compute the posterior probability of any node given the values of the others: in this way, we can use a Bayesian network to predict the disease status of a patient given some of its features (such as genetic variations).

Spurious association



Missed association

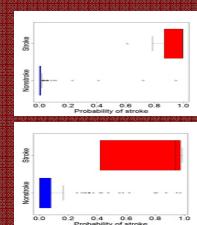
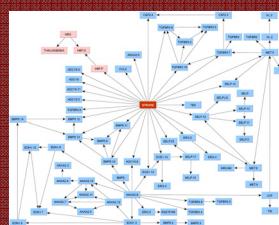


STROKE

We have applied this approach to the identification of the genetic risk factors of stroke in sickle cell anemia patients. We learned a network from a cohort of 1398 subjects with sickle cell anemia analyzing 108 SNPs in 39 candidate genes.

We validated the model by predicting the occurrence of stroke in an independent population of 114 patients with an accuracy of 98.2%.

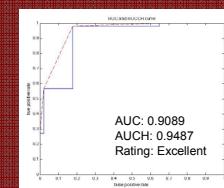
Risk	ANXA2.6	BMP6.10	BMP6.12	SELP.14	TGFBR3.10	ERG.2	N
0.007 (0.003)	hCV26910500	n267190	rs40505	rs3917733	rs284675	rs989554	
0.020 (0.038)	AG	TT	TT	CT	CT	AG	1
0.193 (0.193-0.20)	AG	TT	TT	CT	CC	AG	4
0.727 (0.610-0.83)	AA	TT	CT	CC	CC	AA	56
0.868 (0.70-0.97)	GG	TT	CC	CC	CC	AA	64
0.988 (0.79-1)	GG	TT	CC	CT	CC	AA	21



Sebastiani P, Ramoni MF, et al, Nat Genet, 2005 Apr;37(4):435-40

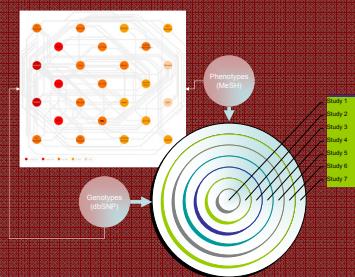
ASTHMA

We are using the same approach to identify the genetic bases of adult-onset asthma. We are currently using a case-control experiment comprising 377 asthma cases and 371 controls. We have validated the current model by predicting the occurrence of asthma in an independent set of 51 cases and 51 controls with an Area under the ROC curve of 0.94.



DISTRIBUTION

In collaboration with Nature Publishing Group we are developing the Human Variation Omnibus (HVO), an on-line databases of human variation studies. The goal of this project is to produce a system able to collect, distributed and integrate genotype studies and allow investigators to explore and combine studies across common phenotypes. HVO uses MeSH to integrate studies across different phenotypes and allows users to explore the distribution of a SNP across different studies sharing a common set of phenotypes. The system if fully integrate within the NCBI network of resources and leverages on these resources to annotate and index SNPs and genes.



i2b2