Project Collaboration between the Mathematics for Real-World Systems Centre for Doctoral Training, University of Warwick & Polymaths

Developing a Model for Genotype-Phenotype Mapping to Predict Risks of Heart-Related Conditions

Summer 2017

polymaths

1. Introduction

Genomics is disrupting medicine. The cost of sequencing a genome has dramatically dropped. When the human genome was first sequenced and published in 2003, it took about 18 years, and cost an estimated US$2.7 billion. Today we can sequence a genome in a few days for less than US$1,000, and the costs will continue to fall with advances in sequencing technologies. The era of personalised genomics is already upon us, but not fully yet with personalised medicine.

Already, direct-to-consumer businesses have been set up to profile genomes and feed information back to individuals, providing advice and counselling where needed. The first of these was 23andMe. They work with genome profiles based on around 0.5 to 2 million base positions out of the 3.2 billion in the human genome. These profiles have been identified to be involved in various functions. Based upon these profiles, 23andMe provide reports about health risks, how a person's genetics may affect their response to drugs, genetic traits and their ancestry.

However, the complexity of human health means that increased specialisation is required to understand the genetic causes of disease and the predisposition to disease. Which is why, specialised genetic testing is needed for specific health areas.

This project will tackle genetic testing of heart-related conditions.

2. Project Background

A simple saliva-based genotyping test has been developed to specifically identify a broad range of unique genetic markers (SNPs), which influence a broad range of heart-related conditions. Based upon this genetic test, reports are generated to help individuals understand how their genetics are impacting upon the health of their heart. The test searches for specific genetic markers associated with a broad range of heart related conditions, including cholesterol levels, risks for hypertension and arterial fibrillation.

The report also examines an individual's response to drugs for their heart condition. It examines eight classes of drugs that affect the cardiovascular system; anti-platelets, anti-

coagulants, statins, stimulants, beta-blockers, ACE inhibitors, calcium channel blockers and hormone therapies.

3. The Challenge

Genome-Wide Association Studies (GWAS) have identified single nucleotide polymorphisms (SNPs) associated with a range of heart related conditions. This list of SNPs have been grouped into allele variations in some cases, and included within the standardized guidelines of a number of drug prescription labels, such as Warfarin, CLopidogrel and Metroprolol etc. Hence, in such drug response reports, the full adherence to such clearly defined allele variations are sought in order to give a corresponding phenotype interpretation.

However, in the case of establishing associated genetic risks with conditions outside the pharmacogenomics arena, the guidelines are vague and sometimes non-existent, regardless of how strong and definite the SNPs association are with their correlating conditions. All of which can be traced to a large number of genome wide association studies.

Nevertheless, when considering a specific condition, rather than a specific SNP, the challenge exists in grouping all these different SNPs, which in some cases might have been studied as independent variables, and to give them an overall interpretational system that would take into consideration the varying influence for each SNP. In other words, to evaluate the overall increasing or decreasing effect of SNPs upon an individual, and to what extent, and to what degree.

The challenge of this project is to establish an accurate and precise genotype-phenotype interpretation model for each of the following conditions:

- Coronary Artery Disease
- Myocardial Infraction
- Genetic risk for decreased HDL levels
- Genetic risk for elevated LDL levels
- Genetic risk for elevated Triglycerides levels

4. Methods

The project will explore models for genotype-phenotype mapping.

Exploring a range of models such as logistical regression (LR), Support Vector Machines (SVM), which have been used to aid in the prediction and diagnosis of health conditions, and risks in individuals, to methods in network science, the project will seek to develop a novel solution to this challenge.

4.1. Data

The data consists of:

- Myocardial Infraction: 11 SNPs
- Coronary Artery Disease: 12 SNPs
- Genetic Risk for Elevated LDL Cholesterol: 10 SNPs
- Genetic risk for decreased HDL level: 14 SNPs
- Genetic risk for elevated Triglyceride levels: 11 SNPs

which will be presented as summary tables.

Each gene and its tested SNP(s) will be presented alongside its possible genotype results, which, consists of 3 possible genotypes for each SNP, and the associated phenotype summary - Decreased, Increased or Typical. An odds ratio value will be given for the inherited condition, and values will also be given for the average concentration change to the genetic risk of elevated/decreased levels of blood nutrients, such as cholesterol and triglycerides.

Examples of the Summary Tables

Coronary Artery Disease Odds Ratio

Gene	Genotype	Summary	Odds Ratio
Gene: *CDH13* SNP: rs8055236	**(T;T)**	Typical risk	1.00
	(T;G)	Increased risk	1.91
	(G;G)	Increased risk	2.23

Genetic risk for decreased HDL level average concentration change

Gene	Genotype	Summary	Average concentration change in mg/dL
Gene: ABCA1 SNP: rs1883025	**(C;C)**	Typical	0
	(C;T)	Decreased	-0.94 mg/dL
	(T;T)	Decreased	-0.94 mg/dL

While summary tables for all SNPs have been generated, they are still subject to review as research is still in progress. So some values may change during the project, and this will be discussed as the project develops.

All data on SNPs is available through research papers and GWAS studies, and from a number of databases, which we will be referring to in this project:

GWAS Catalog

https://www.ebi.ac.uk/gwas/

Phenotype-Genotype Integrator on NCBI

https://www.ncbi.nlm.nih.gov/gap/phegeni

4.2 Supervisors

Javier Cuervo PhD
Polymaths
javier@polymathsconsulting.com

Farooq Khan
Polymaths
farooq@polymathsconsulting.com

5. About Polymaths

Polymaths is an R&D consulting firm that helps organisations discover innovative solutions to complex problems, through complexity science thinking and the computational sciences.

Modelling is at the heart of what we do. We use modelling approaches from across disciplines, whether it is the application of statistical physics to economics, using game theory to understand cancer, applying quantum mechanics in epidemiology, natural selection in astrophysics, or biology in design; we hybridize multiple fields, embracing concepts and methodologies from across disciplines to create innovation. This is central to our vision for creating innovation.

5.1 The Opportunity

This is an exciting opportunity to work on a project that could have significant impact on developing a technology in the life sciences.

Polymaths
Institute of Research & Development
Birmingham Research Park
Vincent Drive
Edgbaston
Birmingham
B15 2SQ

+44 (121) 412 8454
@polymathsglobal

www.polymathsconsulting.com