SNP Data Curation

Client Overview

The client is one of the world's leading research and teaching institutions. The institute has a multidisciplinary group of scientists and scholars, from mathematicians and engineers, biologists to computer scientists and physicists, in an interactive and collaborative environment.

Challenge

As part of an international collaborative project to annotate the genome of an individual, the client required a comprehensive database of association data of single nucleotide polymorphisms (SNPs) and human disease phenotypes. A new database was required because no single public resource offered the capabilities necessary for the clinical annotation of a whole genome. The ultimate aim of the data is to aid the creation of new analytical tools and statistical approaches geared towards clinical recommendations that would allow physicians to select evidence parameters of their choice

Solution

Optra Systems built a new database of SNP-disease associations culled from a number of publicly available sources. For each SNP it captured disease phenotypes, study and sample information, statistical data for association parameters and frequencies of the variants all under a single effort.

- A literature survey of published papers on SNP-disease association was carried out through PubMed to extract disease associations from tables, full text, and supplementary data.

- Broad level disease name or condition under study along with the corresponding MeSH qualifier was collected. Under this broad level, the individual associations between the SNPs and the narrow phenotypes were recorded.
- Data was recorded for both case-control and cohort studies. The sample description included the sample size, gender, ethnicity and other information pertaining to the particular sample under study. The ethnicity was later mapped to the HapMap phase III populations.
- While recording the association data, various statistical parameters were considered. A preliminary threshold of p value less than or equal to 0.05 was used as a cutoff. Odds ratios, relative risks and the hazards ratios gave the strength of association and effect size

Individual SNP information consisted of major and minor alleles, risk alleles for the diseases, genetic models for association and the frequencies of the alleles in the populations.

Benefits

- Optra System's domain expertise assisted the client to create a comprehensive and novel database of disease-SNP associations as an initial step towards deriving clinically meaningful information about disease risk and response to drugs in patients with whole genome sequence data.
- The extensive manual curation involved in the project (4000+ papers and ongoing) enabled the capturing of all relevant information onto a single schema for downstream automated usage given the lack of standardized reporting format for association studies.
- An important benefit of this new database over the preexisting ones is that it indicates the risk and protective allele for each disease – SNP association.

Technologies

- Literature mining was implemented using manual data curation from PubMed
- The recorded data was stored in a relational database – sql

OptraHealth LLC.
530 Lakeside Drive, Ste 250, Sunnyvale, CA 94085
Tel: +1-408-524-5300, Fax: +1-408-524-5302, Email: info@optrahealth.com

About Optra Health

Optra Health is an ISO-certified global organization with deep domain expertise in medical devices, lab automation, life science informatics and healthcare IT solutions. The company provides a fully-scalable, cost-effective OptiShore™ delivery model. This enables customers to choose the optimal balance between on-site, on-shore, and off-shore development that will best address their budget and collaboration requirements. With Optra Health, customers are able to shrink their time-to-market by leveraging practical, building-block based solutions. Committed to clear communication and total transparency, the company consistently meets or exceeds its clients' expectations. Offering a full complement of expert engineering and consulting services, Optra Health is aligned to real business needs applied over the entire product development lifecycle. The robust, scalable and efficient IT infrastructure of the company, together with its outstanding project management team, consistently ensures superior results. Optra Health's global delivery model helps its customers cut costs by about 50% without compromising on quality and realize a 200% improved production cycle.

Visit Optra Health today: http://www.optrahealth.com

Contact Optra Health Today

OptraHealth LLC.
530 Lakeside Drive, Ste 250
Sunnyvale, CA 94085
Tel: +1-408-524-5300; Fax: +1-408-524-5302
Email: info@optrahealth.com