Journal of Glycomics & Lipidomics

Extended Abstract

Identification of clinically useful genomic and epigenomic variants

Xiong Momiao

Abstract

Next generation sequencing technologies will generate unprecedentedly massive (thousands or even ten thousands of individuals) and highly-dimensional (up to hundreds of millions) genomic and epigenomic variation data. A fundamental question is how to efficiently extract genomic and epigenomic information of clinical significance. Traditional paradigm for identifying variants of clinical validity is to test association of the variants. However, significantly associated genetic variants may or may not be usefulness for diagnosis and prognosis of diseases. Alternative to association studies for finding genetic variants of predictive utility is to systematically search variants that contain sufficient information for phenotype prediction. To achieve this, we introduce concepts of sufficient dimension reduction which project the original high dimensional data to very low dimensional space while preserving all information on response phenotypes. We then formulate clinically significant genetic and epigenetic variant discovery problem into sparse SDR problem and develop algorithms that can select significant genetic variants from up to or even ten millions of predictors with the aid of dividing SDR for whole genome into a number of sub-SDR problems defined for genomic regions. The sparse SDR is in turn formulated as sparse optimal scoring problem. To speed up computation, we apply the alternating direction method for multipliers to solving the sparse optimal scoring problem which can easily be implemented in parallel. To illustrate its application, the proposed method is applied to the TCGA overall cancer dataset.

The American College of Medical Genetics and Genomics (ACMG) previously developed guidance for the interpretation of sequence variants.1 In the past decade, sequencing technology has evolved rapidly with the advent of high-throughput next generation sequencing. By adopting and leveraging next generation sequencing, clinical laboratories are now performing an ever increasing catalogue of genetic testing spanning genotyping, single genes, gene panels, exomes, genomes, transcriptomes and epigenetic assays for genetic disorders. By virtue of increased complexity, this paradigm shift in genetic testing has been accompanied by new challenges in sequence interpretation. In this context, the ACMG convened a workgroup in 2013 comprised of representatives from the ACMG, the Association for Molecular Pathology (AMP) and the College of American Pathologists (CAP) to revisit and revise the standards and guidelines for the interpretation of sequence variants. The group consisted of clinical laboratory directors and clinicians. This report represents expert opinion of the workgroup with input from ACMG, AMP and CAP stakeholders. These recommendations primarily apply to the breadth of genetic tests used in clinical laboratories including genotyping, single genes, panels, exomes and genomes. This report recommends the use of specific standard terminology: 'pathogenic', 'likely pathogenic', 'uncertain significance', 'likely benign', and 'benign' to describe variants identified in Mendelian disorders. Moreover, this recommendation describes a process for classification of variants into these five categories based on criteria using typical types of variant evidence (e.g. population data, computational data, functional data, segregation data, etc.). Because of the increased complexity of analysis and interpretation of clinical genetic testing described in this report, the ACMG strongly recommends that clinical molecular genetic testing should be performed in a CLIA-approved laboratory with results interpreted by a board-certified clinical molecular geneticist or molecular genetic pathologist or equivalent.

In 2013 a workgroup consisting of ACMG, AMP, and CAP members, representing clinical laboratory directors and clinicians, was formed with the goal of developing a recommendation for the use of standard terminology for classifying sequence variants using available evidence weighted according to a system developed through expert opinion, workgroup consensus and community input. In order to assess the views of the clinical laboratory community, surveys were sent to over 100 sequencing laboratories, from the United States (US) and Canada that were listed in GeneTests.org, requesting input on terminology preferences and evaluation of evidence for classifying variants. Laboratory testing experience included rare disease as well as pharmacogenomics and somatic cancer testing. The first survey, aimed at assessing terminology preferences, was sent in February 2013 and the results presented in an open forum at the 2013 ACMG annual meeting including over 75 attendees. Survey respondents represented over 45 laboratories in North America. The outcome of the survey and open forum indicated that: (1) a five-tier terminology system using the terms pathogenic, likely pathogenic, uncertain significance, likely benign, and benign was preferred and already in use by a majority of laboratories, and (2) the first effort of the workgroup should focus on Mendelian and mitochondrial variants.

In the first survey, laboratories were also asked to provide their protocols for variant assessment, and eleven shared their methods. By analyzing all the protocols submitted, the

workgroup developed a set of criteria to weight variant evidence and a set of rules for combining criteria to arrive at one of the five classification tiers. Workgroup members tested the scheme within their laboratories for several weeks using variants already classified in their laboratories and/or by the broader community. In addition, typical examples of variants harboring the most common types of evidence were tested for classification assignment to ensure the system would classify those variants according to current approaches consistently applied by workgroup members. A second survey was sent out to the same laboratories identified through GeneTests.org as well as through AMP's list serve of approximately 2000 members in August of 2013 with the proposed classification scheme and a detailed supplement describing how to use each of the criteria. Laboratories were asked to use the scheme and to provide feedback as to the suitability and relative weighting of each criteria, the ease of use of the classification system, and whether they would adopt such a system in their own laboratory. Responses from over 33 laboratories indicated majority support for the proposed approach and feedback further guided the development of the proposed standards and guidelines.

The workgroup also evaluated the literature for recommendations from other professional societies and working groups that have developed variant classification guidelines for well-studied genes in breast cancer, colon cancer, and cystic fibrosis and statistical analysis programs for quantitative evaluation of variants in select diseases.2–5 While those variant analysis guidelines are useful in a specific setting, it was difficult to apply their proposed criteria to all genes and in different laboratory settings. The variant classification approach described in this paper is meant to be applicable to variants in all Mendelian genes whether identified by single gene tests, multi-gene panels, exome sequencing or genome sequencing. We expect that this variant classification approach will evolve as technology and knowledge improves. We should also note that those working in specific disease groups should continue to develop more focused guidance regarding the classification of variants in specific genes given that the applicability and weight assigned to certain criteria may vary by gene and disease.

This work is partly presented at 2nd International Conference on Big Data Analysis and Data Mining 30-December 01, 2015 San Antonio, USA

Xiong Momiao,
The University of Texas School of Public Health, USA, E-mail: momiao.xiong@uth.tmc.edu