Molecular diagnosis of severe combined immunodeficiency using whole-exome sequencing

WL Yang*, YL Lau, PPW Lee, BHY Chung, J Yang

KEY MESSAGES

1. We studied 50 cases of severe combined immunodeficiency from 33 families using a combination of genetic and genomic approaches.
2. Molecular diagnosis was successfully made in 19 of the 33 families.
3. A novel primary immunodeficiency disease gene—*RASGRP1*—was identified.
4. In a number of cases, we gained new understanding of genotype-phenotype correlations, such as mutations in *TTC7A* and in desmoplakin.
5. We gained valuable experience in making a molecular diagnosis of severe combined immunodeficiency; such diagnosis may help determine the causal mutations in this group of patients.

Hong Kong Med J 2018;24(Suppl 3):S15-7
HMRF project number: 01120846

WL Yang, YL Lau, PPW Lee, BHY Chung, J Yang

Department of Paediatrics and Adolescent Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong

* Principal applicant and corresponding author: yangwl@hku.hk

This summary is partly based on studies first reported in: (1) Yang W, Lee PP, Thong MK, et al. Compound heterozygous mutations in *TTC7A* cause familial multiple intestinal atresias and severe combined immunodeficiency. Clin Genet 2015;88:542-9. (2) Mao H, Yang W, Latour S, et al. *RASGRP1* mutation in autoimmune lymphoproliferative syndrome-like disease. J Allergy Clin Immunol 2017. pii: S0091-6749(17)31756-6 [Epub ahead of print].

Introduction

This study aimed to provide a molecular diagnosis for patients with severe combined immunodeficiency (SCID)—a severe form of primary immunodeficiency diseases (PID)—who were negative in candidate gene screening, by using a combination of genetic and genomic approaches. This study also aimed to discover new genes or signal transduction pathways involved in SCID, to survey the spectrum of causal mutations for SCID in a Chinese population, and to develop algorithms and bioinformatics pipelines/tools for whole-exome sequencing (WES) data analysis that can be applied to the clinical diagnosis of rare Mendelian diseases. We successfully conducted WES on the study patients and identified novel SCID genes. Our approach facilitates an understanding of the disease and enables the provision of definitive genetic counselling for affected families. We gained valuable experience in WES data analysis for the diagnosis of monogenic diseases. We have also trained graduate students in the analysis of relevant genetic and genomic data.

Methods

We performed WES on genomic DNA extracted from peripheral blood leukocytes of 50 cases of severe combined immunodeficiency from 33 families. Coding exons of almost 20 000 genes in the human genome were enriched before being sequenced using next-generation sequencing technology, mainly with the HiSeq 2500 System (Illumina, San Diego [CA], United States) on 2x150-bp paired reads. The effect of the genetic variants detected on the encoded proteins was analysed, as well as their population frequency based on data from public and proprietary domains, the function of the genes, and the assumed genetic inheritance mode in the affected families. Potential causal mutations were confirmed by Sanger sequencing, and their frequency in the local population was examined by Sequenom Massarray, a MALDI-TOF mass spectrometry–based genomics analytical method (San Diego [CA], United States). For a number of samples, RNA-seq (RNA sequencing) was used to aid the identification of causal mutations by detecting crucial gene expression changes and alternative splicing aberrations. Owing to the inadequacy of analysis tools for next-generation sequencing platform data, we developed an in-house tool and a genetic variant database that was more relevant to the local population. For exome sequencing data analysis, we have an established pipeline in the laboratory.

For RNA-seq data analysis, quality control with the FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) for raw reads included checks of sequence quality, GC content, overrepresented *k*-mers, and duplicated reads in order to detect sequencing errors, polymerase chain reaction (PCR) artefacts, or contaminations. Reads were mapped to a reference genome with a gapped mapper STAR.[1] Quantification was primarily based

on the number of reads that mapped to each transcript sequence using HTSeq-count.[2] As we did not have exactly matched controls, RNA-seq data from a number of samples from NCBI Gene Expression Ominibus were used as controls. Those samples were derived from peripheral blood mononuclear cells, the same cell source as our patient RNA-seq data. Differential expression analysis was performed using DESeq2,[3] which takes raw read counts as input and introduces possible bias sources into the statistical model to perform an integrated normalisation as well as a differential expression analysis.

The functional analyses of the RNA-seq data were divided into three major components: expression level analysis to detect deletion of a candidate gene or disturbance of a major signal transduction pathway; variant calls to detect mutations that might be missed by WES; and detection of aberrant splicing variant or absence of major functional transcripts. Studying differentially expressed genes allowed the detection of complete loss of expression of a gene or a major transcript for a gene. Complete loss could indicate a homozygous deletion of a locus that might not be reflected by WES data, as variations in coverage depth are intrinsic to exome sequencing data. Furthermore, upregulation or downregulation of a group of genes might indicate aberration of a major signalling transduction pathway. Thus, the differentially expressed genes were assessed using pathway analysis tools such as ToppGene (https://toppgene.cchmc.org/) and David (https://david.ncifcrf.gov/). The variants identified from RNA-seq data were systematically compared with those from the WES data to try to detect variants that might be missed. Different transcripts from alternative splicing for a given gene were evaluated to detect genes with potential mutations that could affect splicing, thereby resulting in either the expression of an aberrant splicing transcript or the absence of a major functional transcript. For the latter, we focused on genes without detection of a major constitutive coding exon(s) or genes that were missing the established major transcript.

Results and Discussion

In this study, we discovered a novel PID gene, *RASGRP1*, which will aid future efforts in diagnosis for this group of patients.[4] In addition, we identified a de novo mutation in desmoplakin, and identified causal mutations in *TTC7A* in a patient with immunodeficiency and intestinal atresias.[5] We identified known causal genes in 19 cases among 33 PID families, thus providing a molecular diagnosis for these families. Furthermore, we developed a novel tool for WES analysis that specifically focuses on SCID, and we provided training to students in genomic data analysis during this process.

Using WES in combination with in-depth analysis and other genetic and genomic approaches, we identified novel genes for SCID and also made a molecular diagnosis in 19 affected patients. These novel findings add to our understanding of SCID and will help our future work in molecular diagnosis for this group of diseases. This approach is important to enable individualised treatment of the disease and for developing new intervention paradigms and new drug targets for SCID. Although identifying genes known to cause SCID is not itself novel, the findings inform us about the mutation spectrum in the local population for this group of diseases—a vital piece of information that will help future genetic screening. Definitive molecular diagnosis will ensure accurate genetic counselling and provide valuable information for prenatal diagnosis if needed. There were a number of cases in which the molecular diagnosis prompted us to re-examine the patient's clinical phenotype. This has enabled a new understanding of genotype-phenotype correlations.

There were a number of families in whom no causal mutations could be identified, even after applying state-of-the-art technologies including WES and RNA-seq. This finding demonstrates the limitations of the technology and the lack of an in-depth understanding of relevant genes and disease mechanisms. Nonetheless, we have been able to build on our experience of next-generation sequencing analysis for molecular diagnosis and to train graduate students to handle data analysis and to make a diagnosis.

From this study and related work, we conclude that WES is a cost-effective approach for molecular diagnosis of PID and other monogenic diseases. A number of candidate genes have usually been screened prior to application of WES. Nonetheless, owing to the enormous heterogeneity of SCID, candidate gene screening is inefficient and time-consuming. With its power and ever-decreasing cost, WES is becoming a mainstream cost-effective approach to the molecular diagnosis of SCID. Among our cases was a case of radiation-sensitive SCID, and mutations in *DCLRE1C* were correctly suspected. Despite this finding, genetic screening using PCR and Sanger sequencing was negative. Yet, homozygous deletion of multiple exons of *DCLRE1C* and adjacent genes was detected by WES. We speculate that the reason for the lack of detection by Sanger sequencing was because of trace contamination of normal DNA that was enormously amplified by the 40-plus cycles of PCR. The issue of trace contamination is well tolerated by next-generation sequencing technology, because only about 10 cycles of PCR amplification were applied during library construction.

Acknowledgements

This study was supported by the Health and Medical

Research Fund, Food and Health Bureau, Hong Kong SAR Government (#01120846). We thank the patients and the families for participating in this research.

Ethical Approval

This study was approved by Institutional Research Board of The University of Hong Kong / Hospital Authority Hong Kong West Cluster (UW12-211). Informed consent was obtained from each participant.

Declaration

The authors have no conflicts of interest to disclose.

References

1. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15-21.
2. Anders S, Pyl PT, Huber W. HTSeq - a Python framework to work with high-throughput sequencing data. Bioinformatics 2015;31:166-9.
3. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 2014;15:550.
4. Mao H, Yang W, Latour S, et al. *RASGRP1* mutation in autoimmune lymphoproliferative syndrome-like disease. J Allergy Clin Immunol 2017. pii: S0091-6749(17)31756-6 [Epub ahead of print].
5. Yang W, Lee PP, Thong MK, et al. Compound heterozygous mutations in TTC7A cause familial multiple intestinal atresias and severe combined immunodeficiency. Clin Genet 2015;88:542-9.