**Asia Pacific Society for Immunodeficiencies**

# Bioinformatics Workshop

**Function Room 1, 2/F, Hong Kong Academy of Medicine**

**1:30pm – 3:00pm**

Dr Wanling Yang

Associate Professor, Department of Paediatrics and Adolescent Medicine, The University of Hong Kong

**Synopsis**

Recent advances in next-generation sequencing (NGS) technology provide a cost-effective approach to large-scale resequencing of human samples for medical application. In recent years, whole exome sequencing, the targeted sequencing for protein coding regions, has become a powerful and regularly-used tool for dissecting the genetic basis of Mendelian diseases. In this work shop, we will briefly introduce the common practice for NGS data analysis, including data manipulation, variant calling, annotation and in-depth analysis for drawing genotype-phenotype correlations. Finally we will use some real cases to demonstrate its usage in clinical diagnosis. The typical framework for DNA sequencing data analysis was elucidated by DePristo et al. in 2011 (Nat Genet. 2011 May;43(5):491-8).

When the variants are collected for each sample, relevant information is evaluated to help identify the causal mutations. First, variants are evaluated by their population frequencies. Since most monogenic diseases are rare, variants that are common in a population are excluded as they are unlikely to be causal for very rare diseases. In this case, data from 1000 Genomes Project, HapMap Project and ExAC Project with whole exome sequencing data on 65,000 samples, and often an in-house database are used to filter out the common variants. The functional impact of the variants on the corresponding genes is then evaluated. Loss of function mutations (nonsense mutations, frameshift indels, splicing aberrations) and missense mutations are preferentially considered since their functional impact is better known. The functional annotations of the genes are also considered. Databases such as NCBI Gene, GeneCard, OMIM, Mouse Genome Informatics etc. often provide valuable information for evaluating the function of the gene and its potential role in the phenotype. Usually, known causal mutations for a disease provide the most definitive information on the evaluation. Thus OMIM, HGMD, ClinVar and literature search should be consulted. For groups of diseases such as PIDs, specifically designated databases often provide more detailed information, such as RAPID and Infevers.

Mendelian inheritance modes are important for the evaluation of the mutations. For example, whether it is autosomal dominant, autosomal recessive, X-linked disease makes much difference in the data analysis. For autosomal recessive diseases, whether the parents are consanguineous determines whether we should focus more on homozygous mutations located in a long homozygous stretch in the patient's genome or we should pay more attention to compound heterozygous is an important consideration. When parental data are available, detection of de novo mutations helps greatly in the search for causal mutations. Information such as gain of function mutation, mutation hotspots are important in cases of dominant inheritance. For novel mutations, guidelines for evaluating mutations in medical applications should be followed: Matthijs G, et al, EJHG. 2016;24(1):2-5).

**Useful web sources**

Population variant frequencies

1. 1000 genome project http://www.1000genomes.org/category/phase-1/
2. NHLBI GO Exome Sequencing Project (ESP): http://evs.gs.washington.edu/EVS/
3. Exome Aggregation Consortium (ExAC): http://exac.broadinstitute.org/

The tools for evaluating mutations:

1. ANNOVAR : http://annovar.openbioinformatics.org/en/latest/user-guide/region/
2. BLAST: http://blast.ncbi.nlm.nih.gov/Blast.cgi
3. Exonic Variants evaluation: http://research.nhgri.nih.gov/skippy/input.shtml
4. Domain Mapping of Disease Mutations: http://bioinf.umbc.edu/dmdm/

The tools for evaluating gene function:

1. BioGPS: http://biogps.org/#goto=welcome
2. NCBI: http://www.ncbi.nlm.nih.gov/gene/
3. OMIM: http://www.ncbi.nlm.nih.gov/omim/
4. RAPID: Resource of Asian Primary Immunodeficiency Diseases
5. Clinical Genomic Database: http://research.nhgri.nih.gov/CGD/search/
6. Immunome : http://structure.bmc.lu.se/idbase/Immunome/index.php
7. Infevers: http://fmf.igh.cnrs.fr/ISSAID/infevers/
8. ClinVar: http://www.ncbi.nlm.nih.gov/clinvar/
9. Mouse genome informatics: http://www.informatics.jax.org/
10. HGMD: http://www.hgmd.cf.ac.uk/ac/index.php

**On-site registration from 6-8th October 2017**

*Priority will be given to paid-up APSID members, otherwise everyone is welcome!*
Please go to Lim Por Yen Lecture Theatre, G/F, Hong Kong Academy of Medicine for the APSID Symposia and get registered!

Light lunch will be provided after the ASPR Congress closing ceremony.
This is a practical session. Please bring along your own laptop computer.