THE HUMAN VARIOME PROJECT
an NGO official partner of UNESCO
sharing data • reducing disease

3rd Variant Effect Prediction Training Course

27 - 30 August 2018

Newcastle University Medicine Malaysia

Islander Puteri, Johor, Malaysia

COURSE BOOK

Scientific Programme Committee

Prof. Johan T. den Dunnen (Leiden, Nederland) CHAIR
Prof. Chris Baldwin (NUMed, Malaysia) LOCAL ORGANISER
Dr Andreas Laner (Munich, Germany)
Prof. Poh San Lai (NUS, Singapore)

Previous Courses

2016 Heraklion, Greece
2017 Prague, Czech Republic

The Human Variome Project and the Scientific Organising Committee wishes to express its gratitude to the following sponsors for their generous support of this event, helping to make this event possible.

In Kind

Gold

Newcastle University Medicine Malaysia

Agilent

Trusted Answers

THE HUMAN VARIOME PROJECT
an NGO official partner of UNESCO
sharing data - reducing disease

3rd Variant Effect Prediction Training Course

27 - 30 August 2018

Newcastle University Medicine Malaysia
Islander Puteri, Johor, Malaysia

PROGRAM

Monday 27th August

10:00 - 11:00	**REGISTRATION**
11:00 - 12:45	**PLENARY SESSION 1** **Iskandar Demonstration Theatre**
11:00 - 11:15	**Welcome & Introduction** Johan den Dunnen Leiden Univ. Medical Center, Leiden, Netherlands & Chris Baldwin Deputy CEO, Dean of Biomedical and Foundation Sciences, Newcastle University Medicine Malaysia
11:15 - 12:00	**Variants in the genome, position and possible consequences** Johan den Dunnen Leiden Univ. Medical Center, Leiden, Netherlands
12:00 - 12:45	**Calling DNA variants using different tools** Raman Sethi National University of Singapore, Singapore
12:45 - 14:00	**Lunch Break - *Fame Cafe***

14:00 - 16:00 **PLENARY SESSION 2**
Iskandar Demonstration Theatre

14:00 - 14:45 **General databases: OMIM, dbSNP, gnomAD (ExAC), etc.**

Robert Kuhn
UCSC Genome Browser, UC Santa Cruz Genomics Institute, Santa Cruz, CA, USA

14:45 - 16:00 **HGVS Nomenclature Workshop**

Johan den Dunnen
Leiden Univ. Medical Center, Leiden, Netherlands

16:00 - 16:30 **Coffee Break - *Fame Cafe***

16:30 - 18:00 **PLENARY SESSION 3**
Iskandar Demonstration Theatre

Human Phenotype Ontology (HPO) Workshop

Andreas Laner
MGZ - Medical Genetics Centre, Munich, Germany
&
Johan den Dunnen
Leiden Univ. Medical Center, Leiden, Netherlands

18:00 **END OF DAY**

Tuesday 28th August

8:30 - 10:45	**PLENARY SESSION 4** **Iskandar Demonstration Theatre**	
8:30 - 9:00	**Gene Variant Databases (LSDBs): HGMD, LOVD, ClinVar, etc.** Johan den Dunnen Leiden Univ. Medical Center, Leiden, Netherlands	
9:00 - 9:45	**Variant Classification: ACMG recommendations** Andreas Laner & Anna Benet-Pages MGZ - Medical Genetics Centre, Munich, Germany	
9:45 - 10:30	**Viewing the Data: the Ensembl browser and its possibilities** Ben Moore Ensembl, EMBL - EBI, Cambridge, UK	
10:30 - 11:00	**Coffee Break - *Fame Cafe***	
11:00 - 11:30	**Samples to answer: Developing a cost-effective and robust targeted NGS workflow** LEE Chee Yang Senior Field Application Scientist – Genomics Solution Division, Diagnostics & Genomics Group, South Asia Pacific, Agilent Technologies Singapore *Agilent Technologies Singapore is a sponsor of the 3rd Variant Effect Prediction Training Course*	
11:30 - 13:00	**WORKSHOP STREAM A** **IT Cluster 1** **Variant Interpretation Using ACMG Guidelines** Andreas Laner & Anna Benet-Pages	**WORKSHOP STREAM B** **IT Cluster 2** **Ensembl Genome Browser Workshop** Ben Moore
13:00 - 14:15	**Lunch Break - *Fame Cafe***	
14:15 - 15:45	**WORKSHOP STREAM A** **IT Cluster 2** **Ensembl Genome Browser Workshop** Ben Moore	**WORKSHOP STREAM B** **IT Cluster 1** **LOVD+ Demonstration** Johan den Dunnen
15:45 - 16:30	**Coffee Break & Poster Session 1 - *Fame Cafe***	

16:30 - 18:00	**WORKSHOP STREAM A** **IT Cluster 1** **LOVD+** **Demonstration** Johan den Dunnen	**WORKSHOP STREAM B** **IT Cluster 2** **Variant Interpretation** **Using ACMG Guidelines** Andreas Laner & Anna Benet-Pages
18:00	**END OF DAY**	

Wednesday 29th August

8:30 - 10:30	**PLENARY SESSION 5** **Iskandar Demonstration Theatre**	
8:30 - 9:15	**Viewing the Data: the UCSC browser and its possibilities** Robert Kuhn UCSC Genome Browser, UC Santa Cruz Genomics Institute, Santa Cruz, CA, USA	
9:15 - 10:00	**Variant Annotation: VEP** Ben Moore Ensembl, EMBL - EBI, Cambridge, UK	
10:00 - 10:30	**Copy-number variation detection from exon capture data** Anna Benet-Pages MGZ - Medical Genetics Centre, Munich, Germany	
10:30 - 10:45	**Assemble for Group Photograph** Go to “Newcastle University Medicine Malaysia” Arch at Bells Court	
10:45 - 11:15	**Coffee Break - *Fame Cafe***	
11:15 - 12:45	**PLENARY SESSION 6** **Iskandar Demonstration Theatre**	
11:15 - 12:00	**Potential Consequences on the RNA Level and using prediction tools** Andreas Laner MGZ - Medical Genetics Centre, Munich, Germany	
12:00 - 12:45	**Potential Consequences on Protein Level and using prediction tools** Poh San Lai National University of Singapore	
12:45 - 14:00	**Lunch Break - *Fame Cafe***	
14:00 - 15:30	**WORKSHOP STREAM A** **IT Cluster 1** **UCSC Genome Browser** Robert Kuhn	**WORKSHOP STREAM B** **IT Cluster 2** **Variant Annotation using VEP** Ben Moore
15:30 - 16:15	**Coffee Break & Poster Session 2 - *Fame Cafe***	
16:15 - 17:45	**WORKSHOP STREAM A** **IT Cluster 2** **Variant Annotation using VEP** Ben Moore	**WORKSHOP STREAM B** **IT Cluster 1** **UCSC Genome Browser** Robert Kuhn
17:45	**END OF DAY**	

Thursday 30th August

8:30 - 10:30	**PLENARY SESSION 7** **Iskandar Demonstration Theatre**
8:30 - 8:50	**Functional Testing from lab tests to animal models** Johan den Dunnen Leiden Univ. Medical Center, Leiden, Netherlands
8:50 - 9:35	**UCSC Variant Annotation Integrator** Robert Kuhn UCSC Genome Browser, UC Santa Cruz Genomics Institute, Santa Cruz, CA, USA
9:35 - 10:30	**NGS in diagnostics: where things can go wrong** Anna Benet-Pages MGZ - Medical Genetics Centre, Munich, Germany & Johan den Dunnen Leiden Univ. Medical Center, Leiden, Netherlands
10:30 - 11:00	**Coffee Break - *Fame Cafe***
11:00 - 13:00	**PLENARY SESSION 8** **Iskandar Demonstration Theatre**
11:00 - 11:45	**Setting up diagnostic NGS application in Malaysia** Zilfalil bin Alwi Universiti Sains Malaysia, Kelantan, Malaysia
11:45 - 12:15	**Future developments** Johan den Dunnen Leiden Univ. Medical Center, Leiden, Netherlands
12:15 - 13:00	**Q&A, Course Evaluation and Close**
13:00	**COURSE END**

Abstracts

Plenary Session 1

Variants in the genome, position and possible consequences

Johan T den Dunnen

Clinical Genetics & Human Genetics, Leiden University Medical Center (LUMC), Leiden, Nederland

The number of variants observed in the human genome has by now passed the 50 million mark. While most of these variants have no known phenotypic consequences, some are deleterious, having severe consequences for the health of the individual.

Variants are divided in three basic types; variants affecting the sequence itself (substitutions and small deletions / duplication / insertions), variants affecting the number of copies (large multi-exon or gene deletions, duplications) and variants affecting the position of a sequence (inversions, translocations, transpositions). Depending on the technology used to analyse a DNA sequence, not all variant types can be detected. Some technologies have been specifically designed to detect specific variants, e.g. MLPA and arrays to detect copy number changes (deletions, duplications). In theory all variant types can be detected using sequencing technology yet the technology (short/ long read) and analysis methods (NGS pipeline) used determine whether all types will be detected and with what efficiency.

Besides the type of variant also its size and location has a large influence on its consequences. Variants involving larger segments of DNA are more likely to contain functionally essential elements and are more likely to have deleterious consequences. Similarly, variants in the coding region of a gene are more likely to affect gene function then variants in introns or in intergenic regions. While the focus is on variants in the coding region, variants affecting regulatory elements (incl. poly-A addition, splicing, expression (promotors/enhancers)) should not be neglected esp. when with a clear phenotype no variant can be found in the obvious candidate gene(s).

The presentation will discuss the three basic variant types and give a brief overview of the most frequent consequences observed in practice. In addition, examples of unusual and rare variants will be presented, highlighting the complexities associated with variant interpretation. The take home message is: everything is possible. Everything which theoretically can go wrong will at some point will go wrong., also in DNA.

Calling DNA variants - SNV, CNV & SV

Raman Sethi, Dept of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore

Next generation DNA sequencing is a powerful tool that can generate high quality sequence data from the entire genome in limited time at a very low cost. Owing to the quality, cost and robustness of the sequence generated by DNA Sequencing technologies, a tremendous growth in genomics research has been witnessed in the past few years. Presently, different types of variants ranging from Single Nucleotide Variants or Point Mutations (single base or nucleotide change from the reference genome), Insertions and Deletions (insertion or deletion of nucleotides/ bases w.r.t reference sequence of size less than 50 bp), Structural Variations (genomic segments over 50 bp in length and includes deletions, duplications, insertions, translocations, inversions and complex rearrangements), Copy Number Variations (DNA segment that is 1kb or larger and present at variable copy number in comparison with a reference genome) can be detected using DNA Sequencing methods. A typical DNA Sequencing workflow involves five main steps: 1) Performing Quality Check of the raw sequencing reads, 2) Alignment of the raw sequencing reads to the human reference genome using genome alignment tools, 3) Calling of the variants from the aligned reads using variant calling tools, 4) Annotation of the called variants to predict the effect or function using annotation tools, and 5) Visualization of the aligned reads and called variants using visualization software. To perform each of the step involved in the analysis, a diversity of tools has been developed. However, the different tools vary from each other in sensitivity and specificity and show low concordance. The output generated by different genome alignment or variant calling method depends largely on the algorithm or detection strategy employed by these methods and can influence variant detection. Moreover, the variant calls output also depends on the genome alignment generated by different genome alignment methods. At present, there are more than 100 variant calling methods or tools developed for the identification of various classes of variants. In this presentation, the different tools used for the identification of Single Nucleotide Variants (SNVs), Structural Variations (SVs) and Copy Number Variations (CNVs) and the algorithm underlying these tools will be discussed. In addition to this, the limitations involved in identification for each class of variants will be explained.

Plenary Session 2

General databases: OMIM, dbSNP, gnomAD, etc.

Robert Kuhn

UCSC Genome Browser, UC Santa Cruz Genomics Institute, Santa Cruz, CA, USA

Many databases have been developed to assist in the interpretation of human variation. Variants that have been determined to be benign, often because of their minor allele frequency (MAF) in the population or their presence in healthy individuals, can be found in dbSNP, GnomAD and the Database of Genomic Variants (DGV). Pathogenic variants are distributed into many different databases: dbSNP, OMIM, DECIPHER, UniProt, ClinVar and LOVD.

The HGVS recommendations to describe sequence variants

See abstract in “Workshops” section.

Plenary Session 3

Human Phenotype Ontology (HPO) Workshop

See abstract in “Workshops” section.

Plenary Session 4

Gene variant databases & sharing information

Johan T den Dunnen

Clinical Genetics & Human Genetics, Leiden University Medical Center (LUMC), Leiden, Nederland

It seems so simple: DNA diagnostics is based on sharing data on genes, variants and phenotypes. Without sharing DNA diagnostics is not possible. When we do not share, we do not offer optimal care to the patients and their families. One would therefore expect that (i) sharing is the standard, and (ii) excellent well-funded databases are available displaying all information known.

Unfortunately, reality is quite different. Sharing is not the standard, far from. Although funding agencies and journals try to force improvements developments are slow. Many databases, especially the gene variant databases, struggle to survive caused by lacking funding and/or the availability of active database curators. As a consequence, available knowledge is spread over a range of databases. On one hand there are the general databases (likeHGMD, dbSNP, EVA, ExAC, COSMIC, OMIM, etc), “a mile wide / an inch deep”, collecting large-scale genome-wide information. On the other hand there are the gene variant databases (like ClinVar, LOVD, UMD, etc), “an inch wide / a mile deep”, collecting individual genetic information including detailed phenotype data.

In my presentation I will focus on the so called gene variant databases, or LSDBs, locus specific databases. I will show examples of the major databases, briefly discuss their origin, their differences, the type of information they contain and the quality to expect. Based on the “Global Variome shared LOVD” I will show how simple it is to submit data and stress the importance to have active database curators. Don't forget the importance of sharing data; the focus should be on the interest of the patient, not on personal interest.

THE ACMG-AMP Classification System

Andreas Laner

MGZ - Medical Genetics Centre, Munich, Germany

Keywords: Variant interpretation, classification, ACMG Guidelines

The ever-growing number of laboratories serving more clinicians and testing more patients for more genes than ever before is a direct result of the implementation of modern, high-throughput DNA sequencing methods in routine diagnostics.

The concept behind this development is the promise that modern clinical molecular diagnostics will have a tremendous impact on human health and disease and play a central role in determining the consequences of genetic variation. That said, the interpretation of sequence variants in terms of pathogenicity and/or clinical consequence is a notoriously difficult task and, owing to the complexity of gathering and interpreting genetic data, individual scientists may disagree on the clinical interpretation of genetic test results.

One major public effort to promote consistency and accuracy in variant interpretation among laboratories was launched in 2015 by the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP), the so-called ACMG-AMP guidelines for the interpretation of sequence variants. While the previous ACMG recommendations from 2007 did not provide defined terms or detailed variant classification guidelines, the revised guidelines from 2015 describe a 5-tier classification system using different types of variant evidence and strength based on informed expert opinion and empirical data.

This presentation will cover the basic characteristics and applications of the ACMG-AMP guidelines as well as some important limitations and pitfalls associated with the use of these guidelines in certain circumstances. Other classification systems will also be briefly discussed. Since recent publications report conflicting results in terms of consistent variant classification using the ACMG-AMP guidelines, the problem of discordant inter-laboratory variant classification will also be addressed.

Viewing the data: the Ensembl browser and its possibilities

Benjamin Moore

Ensembl, EMBL - EBI, Cambridge, UK

The Ensembl genome browser [1] provides visualisation and analysis of integrated genomic data, including genes, variants, comparative genomics and gene regulation, for over 100 species.

The variation data presented in Ensembl can be categorised as either small sequence variants with specific well-defined changes or larger structural variants, and includes population frequency and phenotype data, where available. This variation data is integrated from a large number of reference databases such as ClinVar, dbSNP, DGVa, the NHGRI-EBI GWAS catalogue and OMIM as well as major biological projects including the 1000 Genomes project and GnomAD.

In addition to presenting variation data from reference databases, Ensembl annotates variants with linkage disequilibrium values and predicted effects on transcripts, proteins and regulatory regions, including functional consequences and pathogenicity predictions.

These data can be accessed through our web browser, APIs (Perl and REST), MySQL and FTP dumps. As well as presenting these variation data, Ensembl also provides the Variant Effect Predictor (VEP) toolkit for the analysis and interpretation of variation data.

[1] The Ensembl Genome Browser: www.ensembl.org

Samples to answer: Developing a cost-effective and robust targeted NGS workflow

LEE Chee Yang, Ph.D.

Senior Field Application Scientist – Genomics Solution Division, Diagnostics and Genomics Group, South Asia Pacific, Agilent Technologies Singapore

The advancement of high-throughput technologies, such as next-generation sequencing (NGS) is providing novel insights into the molecular basis of disease, including complex diseases such as Cancer. This have revolutionized the discovery and understanding of rare polymorphisms, structural variants, novel transcripts and methylation states for a wide scope of studies on cancer, mendelian disorders and so on. The Agilent SureSelect platform is a target enrichment system that allows next-generation sequencing users to analyze specific genomic regions or loci with unprecedented depth and accuracy with better throughput and scalability. In addition, with the integration of the Alissa analysis platform (Align & Call and Interpret) into the workflow, it allows for seamless data analysis and interpretation of sequencing data. Here, we discuss the Agilent target enrichment solution from library prep, target enrichment, sample QC, automation, as well as data analysis (Alissa) that enables a streamlined workflow tailored to meet specific needs for target coverage, throughput and turn-around time for comprehensive profiling of variants.

Biography:

Lee Chee Yang joined Agilent since 2010 and is part of the South Asia Pacific (SAP) Applications Team for the Diagnostics and Genomics Group. His work as a Senior Field Applications Scientist includes providing applications support for the Genomics portfolio of products, such as targeted NGS and microarray for research and clinical research in cancer and human health. He has also been involved in many projects, one of which was utilizing systems that combines liquid chromatography-mass spectrometry with PCR-based assays for viral/microbial detection. Chee Yang completed his Ph.D. at Curtin University in Western Australia and was involved in a wide range of research projects. These include the characterization of the ovine MHC class II region and its association with gastrointestinal parasite resistance, the effects of IgA on parasite resistance in sheep, identification of microsatellites in pedigree testing in Alpacas, as well identifying microsatellites in several crustacean species in environmental studies.

Plenary Session 5

Viewing the data: the UCSC Genome Browser and its possibilities

Robert Kuhn

UCSC Genome Browser, UC Santa Cruz Genomics Institute, Santa Cruz, CA, USA

The UCSC Genome Browser is a display platform for visualizing genomic data of many kinds. In the context of variant analysis, it is possible to load your own data into the Browser and compare it to resident datasets displaying large variants (CNVs) and smaller substitutions and indels (SNV/SNPs) with both pathogenic and benign designations. Data from all the large projects recommeded by ACMG can be viewed at the same time with active links to the original data source.

For variants that have not been seen before, Conservation data as well as regulatory information from several large projects are available: ENCODE and Roadmap Epigenomics, enhancers, Ensembl Regulatory Build and many other tracks. The Genome Browser allows the viewing of all these data at one time, in one place, with quick access to the original data at the site of the contributing database. You can move seemlessly between hg19 and hg38 assemblies, find data about mouse knockouts in many human genes, export DNA sequences, test PCR primers and much more.

Variant Annotation: VEP

Benjamin Moore

Ensembl, EMBL - EBI, Cambridge, UK

The Ensembl Variant Effect Predictor (VEP) [1] is an open source, free to use tool for the annotation of genomic variants [2]. It is available as an easy-to-use web interface, as a standalone perl script and can also be accessed through the Ensembl REST API.

The VEP supports the annotation of both sequence variants with specific and well-defined changes (including Single Nucleotide Variants (SNVs), insertions, deletions, multiple base pair substitutions, microsatellites, and tandem repeats); and larger structural variants, including those with changes in copy number or insertions and deletions of DNA. Annotation of variants can be performed for data submitted in a number of formats, including: HGVS notation, VCF and variant IDs. Therefore, the VEP is suitable for variant interpretation in a wide range of study designs, from the analysis of a single variant to the annotation of millions of variants identified in whole-genome or whole-exome variant calls.

For all input variants, the VEP returns detailed annotation for predicted effects on transcripts, proteins and regulatory regions, including functional consequences and pathogenicity predictions. For known or overlapping variants, allele frequencies, phenotype information and literature citations can also be retrieved from the Ensembl databases.

Recently, we have also developed Haplosaurus [3]; a tool that annotates consequences taking multiple variants into account using phased genotypes from a VCF file. This approach offers an advantage over the VEP analysis, which treats each input variant independently. By considering the combined change contributed by all the variant alleles across a transcript, the compound effects the variants may have are correctly annotated, giving a personalised reference proteome.

[1] The Ensembl Variant Effect Predictor: http://www.ensembl.org/Tools/VEP

[2] McLaren, W. et al. “The Ensembl Variant Effect Predictor” Genome Biology 2016, 17:122 doi: 10.1186/s13059-016-0974-4

[3] Haplosaurus: https://github.com/Ensembl/ensembl-vep

Copy-number variation detection from exon capture data

Anna Benet-Pagès

Medical Genetics Center, MGZ, Munich, Germany

Gene dosage abnormalities account for a significant proportion of pathogenic mutations in rare genetic disease related genes. In times of next generation sequencing (NGS), a single analysis approach to detect SNVs and CNVs from the same data source would be of great benefit for routine diagnostics. However, CNV detection from exon-captured NGS data has no standard methods or quality measures so far. The primary strategy of the current bioinformatics methods is based on the read depth of coverage (DOC). The underlying approach is to compare the differences of DOC in particular genomic regions between case and control samples. The DOC-based methods can detect arbitrarily large CNVs and can be effectively used with paired-end, single-end, and mixed read data. Numerous standalone and web-based tools are currently available to detect CNVs based on different features of NGS data, resulting in variation in the prediction of CNVs. The advantages of incorporating more than one method for reliable prediction of CNVs, in addition to the key factors which affect enormous the sensitivity and specificity of CNV pipelines (i.e. size of the reference set, kit performance, normalization approach, single exon calls, variability in the capture efficiency of nearby genomic regions, and low complexity sequences) will be here discussed. Furthermore, the experiences of copy number analysis in 4000 patients with hereditary cancer or rare Mendelian diseases will be presented. Overall parallel analysis of SNVs and CNVs from NGS capture data within a routine diagnostic setting increased the diagnostic yield between 5% and 10% depending on the associated phenotype.

Plenary Session 6

Potential consequences on RNA level

Andreas Laner

MGZ - Medical Genetics Centre, Munich, Germany

Keywords: Splice acceptor and splice donor sites; branch sites; ESE / ESS, mRNA stability; micro-RNA binding; translational folding / codon usage

Pathogenic DNA variants are classically thought of as truncating variants such as PTCs, indels, single or multi exon deletions/duplications, canonical +/- 1, 2 splice sites or missense substitutions that change the biological function of the gene product. However, studies during the last two decades suggest that more than 15 % of disease-causing variants exert their impact by altering the RNA structure and/ or function (mainly splicing)[1, 4]. This is a rather conservative estimate, as research has only recently begun to routinely assess e.g. non-canonical splicing abnormalities, and there is evidence that many unclassified genetic variants might turn out to result in splicing aberrations or other consequences on the RNA level.

Variants of the canonical +/- 1, 2 splice sites are relatively easy to interpret and in most cases can be regarded as pathogenic, although there are important exceptions like small in-frame skipped exons containing no functional important domain. Changes in branch sites and in exonic or intronic splicing enhancers and silencers (ESE/ESS and ISE/ISS) are harder to identify and to interpret even though more elaborate analysis tools have been developed in the last years. Variants affecting one of these afore mentioned classes alter the RNA structure by influencing pre-mRNA splicing.

On the other hand, even variants which do not influence the correct mRNA splicing can have a vast biological effect on RNA level for example by altering the mRNA stability either directly or by changing binding sites of micro-RNAs. Finally, altering the speed of translational folding via codon usage can influence the protein structure and subsequent protein function.

This presentation provides an overview of different functional mechanisms leading to potentially deleterious consequences on the RNA level.

[1] Baralle et al. 2009. EMBO Rep 10(8):810-816

[2] Diederichs et al.; 2016. EMBO Molecular Medicine 8(5):442-457

[3] Scotti et al.; 2015. Nat. Rev. Genet. 17(5): 19-32

[4] Ward & Cooper; 2010. J Pathol. 220(2): 152-163

Potential Consequences on Protein level

Poh-San Lai 1,2,3,4

1 Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228

2 Khoo Teck Puat - National University Children's Medical Institute, National University Health System, Singapore 119228

3 Genome Institute of Singapore, A*STAR, 60 Biopolis St, Singapore 138672

4 Defence Medical & Environmental Research Institute, DSO National Laboratories (Kent Ridge), Singapore 117510

Whole exome sequencing (WES) is increasingly been used for clinical sequencing to identify disease-causing variants for resolving diagnosis in patients. Typically, tens of thousands of nucleotide variants (SNVs) can be generated from an exome of an individual. SNVs in gene coding regions can be either synonymous (code for same amino acids with no effect on translated protein products) or non-synonymous (change the amino acids with potential damaging effect on protein functions). However, not all "damaging effect" on a protein may lead to a phenotypic effect. Nonetheless, determining the consequence of non-synonymous SNVs is important in order to filter out the background variants to prioritize the number of candidate variants for further curation and interpretation. Computational predictive tools are thus important as each individual can carry between 8,000 to 10,000 non-synonymous SNVs, many of which may be rare, or have unknown significance. Experimental validation of the effect of the amino acid change is not technically or resource feasible for every variant due to the large numbers involved. Additionally, appropriate functional assays may not exist for all proteins. Hence, protein prediction tools are used to help identify variants that are likely to cause disease. Most of these tools have been developed based on features such as amino acid or nucleotide conservation, changes in biochemical properties, functional domains affected, structural effects, etc. Some also incorporate algorithm training and weight predictions from various platforms. Hence, the prediction outputs from the tools may sometimes disagree due to the different predictive features used. This talk will share on the different protein prediction tools that are commonly used in WES datasets and discuss on how to predict the effect of protein changes from observed non-synonymous SNVs.

Plenary Session 7

Functional testing: from lab tests to animal models

Johan T den Dunnen

Human Genetics, Leiden University Medical Center (LUMC), Leiden, Nederland

When a variant has been identified, after sharing/submission to a public repository, the first question is “Has the variant been reported before?”, and if so “What were the associated consequences, is the function of the gene (RNA, protein) affected or not?”. This presentation will focus on what one can do when the variant has not been reported before or when the associated consequences are not sure. What possibilities are their to test the functional consequences and how reliable are the results obtained.

One of the most valuable sources of information that can be used is of course material from the individual analysed. Depending on the material available (tissue, cells, protein, RNA, etc.) a range of experiments can be performed to check the functional consequences incl. a detailed RNA (Northern blot, RT-PCR) or protein (enzyme activity, Western blot, immuno histochemistry) analysis. When living cells are available complementation assays can be performed or tissue-specific processes studied after generating IPS cells and lineage specific culturing. When no living cells are available expression constructs can be generated and used to test the functional consequences in different cell systems, incl. cells where the endogenous gene copy was knocked out. A more demanding approach is the use and construction of animal models (mouse/rat, zebra fish, Drosophila, C.elegans, yeast, etc.). Finally, some general public repositories will be mentioned, incl. the Protein Atlas (www.proteinatlas.org), which can be used helping to address the question: “What are the consequences of the variant”.

UCSC Variant Annotation Integrator

Robert Kuhn

UCSC Genome Browser, UC Santa Cruz Genomics Institute, Santa Cruz, CA, USA

The Variant Annotation Integrator is a tool which takes input in the form of a dataset of variants, including the genomic coordinates, the reference nucleotide and the variant and exports a prediction of the biochemical consequences of the change: missense, splice variant, etc, as well as SIFT and PolyPhen (and other) predictions. Conservation score and intersections with regulatory elements can also be exported in the output. This presentation will outline how to get data into the VAI and how to configure and interpret the output. A recent addition to the VAI is support for HGVS nomenclature at both input and output.

NGS in diagnostics: where things can go wrong

Anna Benet-Pages, Johan T den Dunnen

MGZ - Medical Genetics Centre, Munich, Germany; Human Genetics, Leiden University Medical Center (LUMC), Leiden, Nederland

Although the next-generation sequencing (NGS) technology has been around for a number of years, it is still a new technology in continuous development. As a consequence we still get occasionally surprised by errors which can occur at any step of the process. Because there is currently no platform for laboratories to share such issues, every laboratory has to identify and correct its own errors. The variable amounts and types of damage, if ignored, can negatively affect the final results. The impact on downstream applications such as sequencing can be profound: from simple library failures to libraries that produce spurious data, leading to misinterpretation of the results. Samples may get swapped, samples may get contaminated and during sample preparation sequences might get lost/amplified. The sequencing technology may have systematic errors and when the data go through analysis pipelines things may go wrong during mapping against a reference sequence, variant calling, variant annotation, variant interpretation. As a consequence both false positive (non-exiting variants called) and false negative (true variants missed) calls will be made.

In this workshop we will illustrate the problems by showing some of the errors that were encountered in the past years, discuss how these can be prevented, and most importantly, how errors can be detected. At the end there will be time for discussion where people may share their own experiences; learning from each other's mistakes will help to improve the opportunities and simplify the challenges of the daily NGS routine.

Workshop programme:

- sources of error (from sample prep to bioinformatic analysis)
- evaluation and validation of NGS workflows
- the value of External Quality Assessment (EQA)
- discussion with the audience

Plenary Session 8

Setting Up Diagnostic NGS Application in Malaysia

Professor Dr. Zilfalil Bin Alwi

University Sains Malaysia, Kelantan, Malaysia

With the advancement of next-generation sequencing (NGS) technologies, a new era of genomics and molecular biology can now be applied to provide higher throughput data at lower cost, enabling population-scale genome research to obtain genomic information of patients. Over the years, many academic institutions and companies have expanded NGS applications from research to the clinic. Implementation of NGS in diagnosing genetic anomaly and diseases can potentially assist in clinical and public health decisions based on discoveries of undescribed mutations mostly involved in the regulation of gene expression to understand the complex landscape of genetic-related diseases such as cancer, and determination of causative agent of infectious diseases and the epidemiology, and evolution of infecting pathogens. At present, genomics diagnostic services are no longer limited to developed countries. In Malaysia, while there has been an increase in the use of NGS in the field of research, the applications of NGS in a clinical diagnostic setting is still at the infancy stage especially for government institutions. The implementation of NGS is challenging in terms of equipment cost and lack of bioinformaticians and trained personnel to administer the system. However, in the next decade the application of NGS is expected to progressively replace the conventional Sanger sequencing in diagnostic setting. This lecture will identify the current gaps and possible corrective actions to improve the quality of genomics and predictive analytics utilising NGS in Malaysian clinical diagnostic setting.

Future developments

Johan T. den Dunnen

Leiden Genome Technology Center (LGTC), Human Genetics, Leiden University Medical Center, Leiden, Nederland

It is very difficult to predict the future, yet we will give it a try.

DNA sequencing and analysis technology is developing at an incredible speed. The option to sequence a complete human genome trying to resolve a potentially genetic health problem is a very attractive option. Looking ahead, with the notion that sequence cost will drop further, we can foresee that genome sequencing will become a standard analysis method and that everything alive will be sequenced.

In a medical setting this will mean that cancer treatment will be based on a thorough analysis of the tumour, sequencing both DNA and RNA to find which pathways are deregulated, to pick the best drugs for treatment. For standard hospital visits, before treatment starts, every patient will be offered a genome sequence with as first result the prevention of adverse drug effects. In addition, diagnosis will include bio-molecular measurements (RNAseq, metabolomics , proteomics, etc.) and the effects of treatment will be followed using the same tools.

When cost drops further, regular health checks will be performed based on a gene expression profiles from sequencing a blood RNA sample. Everybody will be able to know their DNA sequence with the information it contains regarding one's talents and health risks. At some point, you will be able to sequence your genome at home, with a small extension of your computer or mobile phone You will control, use and share your DNA data through your personal locker at the DNA-bank.

So the future?, ..sequencing!

Practical Workshops

The HGVS recommendations to describe sequence variants

We encourage you to bring any variants you are having trouble naming with you to the workshop.

Johan T den Dunnen

representing the HGVS/HVP/HUGO Sequence Variant Description Working Group (SVD-WG)

The HGVS recommendations for the description of sequence variants have grown into a standard accepted world-wide. The recommendations can be applied to describe sequence variants on the level of DNA, RNA and protein in any organism.

The recommendations originate from the activities of a group of scientist interested in the collection and sharing of variants that are associated with disease, the HUGO Mutation Database Initiative (HUGO-MDI). Currently the recommendations are under the auspices of the Human Genome Variation Society (HGVS), the Human Variome Project (HVP) and the HUman Genome Organisation (HuGO) and are an IRDiRC “Recognized Resource”. A committee with expert representatives from diverse groups of users, the HGVS/HVP/HUGO Sequence Variant Description Working Group (SVD-WG), collects all requests for modifications/additions. The SVD-WG decides whether changes are required and when appropriate prepare a proposal which is set out for community consultation (open for 2-3 months). To allow users to track changes the HGVS recommendations, available through http://www.HGVS.org/varnomen, work with version numbers. When questions arise they can be mailed to VarNomen@HGVS.org. Where necessary the pages will be modified to correct errors or to clarify issues (e.g. by giving examples), with summaries listed on facebook (www.facebook.com/HGVSmutnomen).

In this workshop the HGVS recommendations will be discussed, briefly. Examples will be shown for the description of the basic variant types; substitution, deletions, duplication, insertion, deletion-insertion. Internet tools will be shown that can be used to generate and/or check HGVS variant descriptions. Specific attention will be given to frequently occurring problems and the current issues/hot topics which were recently open for community consultation. During the workshop participants will test their knowledge of HGVS nomenclature by answering specific nomenclature questions.

Human Phenotype Ontology (HPO) Workshop

Andreas Laner, Johan T den Dunnen

MGZ - Medical Genetics Centre, Munich, Germany; Human Genetics, Leiden University Medical Center (LUMC), Leiden, Nederland

In sequencing-based diagnostics, standards are essential when we want to understand each other and exchange information on the phenotype (disease) studied and variants found. While the standard to describe variants is clear, HGVS nomenclature, phenotype descriptions have for long been more variable hampering computational analysis. The Human Phenotype Ontology (HPO) provides a tree-based ontology of medically relevant phenotypic features based on adequate computational data structures, standardising phenotype descriptions. Based on the logical, computer understandable descriptions of clinical features similarities can be identified between different genetic diseases and across different species. HPO's usability and impact are increased by linking the medical terms to lay term synonyms and by translating HPO into several languages using a crowdsourcing approach.

HPO is increasingly adopted as a standard for phenotypic abnormalities by diverse groups such as international rare disease organizations, registries, clinical labs, biomedical resources, and clinical software tools and will thereby contribute toward nascent efforts at global data exchange for identifying disease etiologies (Köhler et al, 2017). HPO currently contains over 13,000 terms and over 156,000 annotations to hereditary diseases. HPO-based phenotype descriptions can now be used to support differential diagnostics using a NGS-based approach to determine the cause of undiagnosed disease, including international phenotype matching queries through the Matchmaker exchange project.

In the workshop following topics will be addressed:

- what are HPO terms
- why are they important
- how can I find the correct HPO term
- how to generate HPO-based phenotype descriptions
- how can I use HPO in NGS diagnostics

If you are unclear about how to describe a phenotype using HPO, bring it with you to the workshop.

Ensembl Genome Browser Workshop

The Ensembl Course materials are available for download on the VEPTC Website:

- **Ensembl Course book**
- **Ensembl Answer book**

Download here: http://veptc.variome.org/course-materials.html

Variant Interpretation Using ACMG Guidelines Workshop

In addition to reading the below, please download:

- **ACMG Workshop Course book**
- **The ACMG recommendations**

Download Here: http://veptc.variome.org/course-materials.html

The dramatic progress in sequence technology, lab automatization, and bio-IT data processing in the last decade have made high-throughput sequencing applications the standard method in molecular diagnostics. Especially since the development of benchtop NGS machines, almost every lab is able to create vast amounts of high-quality sequence data. However, there are still some important hurdles to overcome, especially the interpretation of sequence variants with a view to providing accurate clinical recommendations, a process that is considered a major bottleneck. Evaluating the pathogenicity of a variant is challenging given the plethora of types of genetic evidence that laboratories need to consider. Deciding how to weigh each type of evidence is difficult, and standards have been set. In 2015, the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) published guidelines for the assessment of variants in genes associated with Mendelian diseases.[1]

The workshop will be divided into a practical part and a short demonstration.

Part A: Practical variant classification (75 min)

In the first part of the workshop, variants from real cases will be discussed collectively. The variants have been selected to represent ACMG-AMP categories that are known to be challenging in the classification process.[2, 3, 4, 5] These variants will been sent to workshop participants in advance of the VEPTC.

To make the workshop more interactive, participants are asked to collect evidence for and against pathogenicity for each case and to prepare possible questions and remarks in advance. During this practical section, differences, difficulties, and discrepancies in variant classification will be discussed for all five cases.

Furthermore, we encourage participants to bring their own "difficult to classify" variants to the workshop. We will collectively try to solve obstacles and problems in variant classification using the ACMG guidelines.

Part B: Demonstration of inter-laboratory concordance in variant classification (15 min)

In the second part of the workshop, we will discuss recent publications that reported conflicting results with regard to consistent variant classification using ACMG-AMP guidelines. Furthermore, the problem of discordant inter-laboratory variant classification will also be addressed.

Aim:

In this workshop, participants will be familiarized with the basic application of ACMG-AMP classification guidelines as well as with the limitations and pitfalls inherent in working with these guidelines on a daily basis. The following points will be addressed during the presentation:

- Familiarization with ACMG-AMP guidelines and their basic application
- Identification of classes of variants not covered by ACMG-AMP guidelines or which must be considered cautiously (e.g. variants with low/moderate penetrance)
- Identification of top error-prone sources of information (e.g. ClinVar OMIM entries, old data sources, research submissions, disease areas, etc.) [5]
- Awareness of various possible errors in variant interpretation
- Awareness of the fact that a considerable number of inter-laboratory discrepancies in variant classification are the result of a lack of published internal data, special biology, and old or invalid data sources.[2, 3, 4, 5]

Recommended Literature:

(1) Richards et al.; Genet. Med. 17, 405–424, 2015
(2) Amendola et al.; Am J Hum Genet 98, 1067–1076, June 2, 2016
(3) Harrison et al.; Genet. Med. Mar 16 (PMID: 28301460)
(4) Pepin et al.; Genet Med. Jan; 18(1) 20-4 (PMID: 25834947)
(5) Yang et al.; Genet. Med. Jun 1 2017 (PMID: 28569743)

LOVD+ WORKSHOP

Johan T den Dunnen

Human Genetics, Leiden University Medical Center (LUMC), Leiden, Nederland

The LOVD+ variant analysis platform is an open source software tool built on top of LOVD that is designed to facilitate the analysis and interpretation of data from next-generation sequencing studies. LOVD+ uses open file standards and formats, making integration with a modern NGS infrastructure (usually built on open source tools) straightforward. The open nature of the architecture allows full access and control over the source code, facilitating full integration with third party systems (LIMS, ELN, etc.). LOVD+ is a fully packaged system tested on multiple operating systems, typically hosted on a web server running a Linux flavor, with instances running on local machines as well as on cloud-based virtual machines. LOVD+ has been developed and is used by the Clinical Genetics' DNA laboratory, Leiden University Medical Center (Leiden, Nederland, working under ISO15189 certification) and the Melbourne Genomics Health Alliance (Melbourne, Australia).

LOVD+ facilitates gene panel and whole exome sequencing (WES) and supports singl patient and trio-based analysis (healthy parents and affected child sequenced) as well s more complex family structures. The platform has a set of preconfigured variant filtering analyses which can be set per disease group and/or inheritance pattern (de novo, autosomal recessive, X-linked, etc.), allowing users to very quickly focus on those variants that are clinically relevant. An LOVD+ analysis starts with uploading data from annotated VCF files (variant annotation performed using tools like VEP or SeattleSeq). Given its open design, data from other sources (e.g. SV/CNV analysis, RNA-seq, methylation) can be added to refine variant filtering. Variants classified in LOVD+ are automatically stored in an in-house database directly supporting future analysis.

While LOVD+ is a freely available open source platform, support might be desired to install the system, train users and guarantee smooth operation. Such support is available from www.GeneticReports.org, an organization established to help fund the Global Variome/Human Variome Project shared LOVD.

For the workshop introducing LOVD+ please go to http://courses.lovd.nl/LOVD+. The Introduction manual is available from the Documentation tabs.

UCSC Genome Browser

The UCSC Genome Browser Course materials are available for download on the VEPTC Website:

- **UCSC Genome Browser Course book**

Download here: http://veptc.variome.org/course-materials.html

Variant Annotation Using VEP Workshop

The VEP Course materials are available for download on the VEPTC Website Here:

- **VEP Course book**
- **VEP Answer book**

Download here: http://veptc.variome.org/course-materials.html

POSTER ABSTRACTS

POS 01 - Tuesday 15:45 - 16:30

A comprehensive automated computer-aided discovery pipeline from genomes to hit molecules

Ruchika Bhat[1,2], Rahul Kaushik[2,3], Ankita Singh[2], Debarati DasGupta[1,2], Abhilash Jayaraj[1,2], Anjali Soni[1,2], Ashutosh Shandilya[1,2], Vandana Shekhar[2], Shashank Shekhar[2] and B. Jayaram[1,2,3*]
[1]Department of Chemistry, Indian Institute of Technology, Delhi, Hauz Khas, New Delhi-110016, India, [2]Supercomputing Facility for Bioinformatics & Computational Biology, Indian Institute of Technology, Delhi, Hauz Khas, New Delhi-110016, India and [3]Kusuma School of Biological Sciences, Indian Institute of Technology, Delhi, Hauz Khas, New Delhi-110016, India.
Email: bjayaram@chemistry.iitd.ac.in; ruchika@scfbio-iitd.res.in
KEYWORDS: genomics, proteomics, structure-based drug discovery, genome to hit pipeline, web suite

Big data generation through sequencing of genomes and proteomes has led to over 2800 whole genomes and ~84 million protein sequences[1]. However utilization of this data to generate lead molecules for curing diseases remains a challenge. We propose here, Dhanvantari[2-19], a comprehensive software suite which automates the computational journey from genome to hit molecules via its various genomics, proteomics and drug designing modules with possible entry at any module. The proposed software suite offers new opportunities and insights to “genome-based” drug discovery along with classical protein-targeted structure-based approaches to discover new drug-like molecules. The pipeline helps in exploring new potential drug targets from genomic/proteomic data which were earlier inaccessible, and helps to find novel hits via screening against million compounds or natural products or FDA approved drugs or even customized molecule libraries. Case studies on Hepatitis B Virus and Hepatitis A Virus against their druggable proteins via this pipeline have led to potent and novel inhibitors with micro molar range inhibitions in in vitro studies. The entire protocol proposed in Dhanvantari requires ~ 6 to12 hours, however individual steps get completed within minutes. The software suite is freely accessible as an online resource at http://www.scfbio-iitd.res.in/software/dhanvantari_new/Home.html with no additional dependencies. Presently, there is no fully-automated open source similar to Dhanvantari which can mine the information about potential drug like molecules from genomic/proteomic information.

REFERENCES

1) Coordinators, NCBI Resource. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 44, D7-D19 (2016).
2) Soni, A. et al. Genomes to Hits In Silico - A Country Path Today, A Highway Tomorrow: A Case Study of Chikungunya. Curr. Pharma. Des. 19, 4687 – 4700 (2013).
3) Khandelwal, G. & Jayaram, B. A phenomenological model for predicting melting temperatures of DNA sequences. PLoS ONE 5, e12433 (2010).
4) Dutta, S. et al. A Physico-chemical model for analyzing DNA sequences. J Chem Inf Model. 46, 78–85 (2006).
5) Singhal, P. et al. Prokaryotic gene finding based on physicochemical characteristics of codons calculated from molecular dynamics simulations. Biophys J. 94, 4173–4183 (2008).

6) Khandelwal, G. et al. DNA energetics based analyses suggest additional genes in prokaryotes. J Bio Sc. 37, 433–444 (2012).
7) Khandelwal, G. DNA-water interactions distinguish messenger RNA genes from transfer RNA genes. J. Am. Chem. Soc. 134, 8814–8816 (2012).
8) Jayaram, B. Beyond the wobble: the rule of conjugates. J. Mol. Evol. 45, 704-705 (1997).
9) Singh, A. et al. Physico-chemical fingerprinting of RNA Genes. Nucleic Acids Res. 45, e47 (2016).
10) Jayaram, B. et al. Bhageerath : An Energy Based Web Enabled Computer Software Suite for Limiting the Search Space of Tertiary Structures of Small Globular Proteins. Nucl. Acids Res. 34,6195-6204 (2006).
11) Narang, P. et al. Protein structure evaluation using an all-atom energy based empirical scoring function. J. Biomol. Str. Dyn. 23, 385-406 (2006).
12) Narang, P. et al. A computational pathway for bracketing native-like structures for small alpha helical globular proteins. Phys. Chem. Chem. Phys. 7, 2364-2375 (2005).
13) Jayaram, B. et al. Bhageerath-H: A homology ab initio hybrid server for predicting tertiary structures of monomeric soluble proteins. BMC Bioinformatics 15, S7 (2014).
14) Dhingra, P & Jayaram, B. A homology/ab initio hybrid algorithm for sampling near-native protein conformations. J Comput Chem. 34, 1925-1936 (2013).
15) DasGupta, D. et al. From Ramachandran Maps to Tertiary Structures of Proteins. J. Phys. Chem. B. 119, 11136–11145 (2015).
16) Singh, A. et al. ProTSAV: A Protein Tertiary Structure Analysis and Validation Server. BBA Proteins and Proteomics 1864, 11-19 (2015).
17) Singh, T. et al. AADS- An automated active site identification, docking, and scoring protocol for protein targets based on physicochemical descriptors. J. Chem. Inf. Model. 51, 2515–2527 (2011).
18) Mukherjee, G. & Jayaram, B. A Rapid Identification of Hit Molecules for Target Proteins via Physico-Chemical Descriptors. Phys. Chem. Chem. Phys. 15, 9107-9116 (2013).
19) T. Jain, and B. Jayaram, "An all atom energy based computational protocol for predicting binding affinities of protein-ligand complexes", FEBS Letters, 2005, 579, 6659-6666.
20) Gupta, A. et al. ParDOCK: An All Atom Energy Based Monte Carlo Docking Protocol for Protein-Ligand Complexes. Protein and Peptide Letters. 14, 632-646 (2007).
21) B. Jayaram, Tanya Singh, Goutam Mukherjee, Abhinav Mathur, Shashank Shekhar, and Vandana Shekhar, "Sanjeevini: a freely accessible web-server for target directed lead molecule discovery", BMC Bioinformatics, 2012, 13, S7. http://www.biomedcentral.com/1471-2105/13/S17/S7

POS 02 - Wednesday 15:30 - 16:15

Association of variant rs1531939 of GRM7 gene with the risk of Schizophrenia in the Bangladeshi Population

Mohammad Safiqul Islam[1, *], Md. Saddam Hussain[1], Md. Shalahuddin Millat[1], Md. Atikur Rahman[1]

[1](Department of Pharmacy, Noakhali Science and Technology University, Noakhali-3814, Chittagong, Bangladesh)

Corresponding author:

Dr. Mohammad Safiqul Islam

Professor and Chairman

Department of Pharmacy

Noakhali Science and Technology University

Noakhali-3814, Bangladesh

Contact: +8801727658650

Email: research_safiq@yahoo.com

Among mental disorders, schizophrenia (SCZ) is the most disabling disease. The genetic polymorphisms of GRM7 gene were reported to be related to the development of several mental diseases. The purpose of this study was to examine the prevalence of rs1531939 allele of GRM7 gene in the Bangladeshi schizophrenic patient and also to detect the association of this disease with this allele. Due to the absence of restriction site of the commercially available restriction enzyme in the rs1531939 allele, we designed an allele-specific polymerase chain reaction (ASPCR) for the first time to detect the mutation in the GRM7 gene. Genotyping was done by extracting DNA from 101 schizophrenic patients and 101 healthy volunteers and performed the ASPCR for amplification of DNA. Among the 101 schizophrenic patients, 12.87, 45.55, 41.58 and 87.13% patients carried CC, CG, GG and CG+GG genotypes respectively whereas, among the 101 controls, 27.72, 26.73, 45.55 and 72.28 % controls carried CC, CG, GG and CG+GG genotypes respectively. In this study, we found that patients with CG genotype had 3.67 times higher risk for the development of schizophrenia in compared to patients carrying CC genotype and this result is statistically significant ($p<0.05$). Whereas patients carrying GG genotype possessed 1.97 times higher risk for the development of schizophrenia in compared to patients carrying CC genotype and this result is not statistically significant ($p>0.05$). Combined model i.e. patients carrying CG +GG genotype had 2.60 times higher risk for the development of schizophrenia in compared to patients carrying CC genotype and this result is statistically significant ($p<0.05$). The frequency of G allele is 64.36% and 58.91% in the patient and control groups respectively. Our results indicate that the rs1531939 allele of GRM7 gene is associated with the higher risk of schizophrenia in our study population. We were also successful to develop an ASPCR method to detect the genotype of rs1531939 allele. As we have identified the genetic basis of the Bangladeshi schizophrenic patients, we hope it will be helpful for the treatment selection of schizophrenic patients.

POS 03 - Tuesday 15:45 - 16:30

WGS Analysis in a patient with suspected Perrault Syndrome

Kunaphas Kongkitimanon [1, 2] ,Wanna Thongnoppakhun [3,4,7], Ekkapong Roothumnong [5,7], Manop Pithukpakorn [3,5,7] , Bhoom Suktitipat* [1,2,3,5,7], Siriraj Neurogenetic clinic network
[1] Integrative Computational BioScience Center, Mahidol University
[2] Cancer Precision Medicine Research Program -- Thailand Research University Network
[3] Siriraj Center of Research Excellence in Precision Medicine
[4] Department of Research and Development
[5] Division of Medical Genetics, Department of Medicine
[6] Department of Biochemistry
[7] Faculty of Medicine Siriraj Hospital, Mahidol University
* Correspondence: bhoom.suk@mahidol.edu

In this report, we studied a Thai female patient suspected of Perrault syndrome (PS), presented with chronic sensorimotor axonal polyneuropathy, followed by primary ovarian failure,Grave's disease, and mild intellectual disability. Whole-genome sequencing (WGS) was performed to confirm the diagnosis and identified other potential causes.
Methods

DNA sample was sent for WGS on Illumina's HiSeq-X platform (Macrogen Inc, Korea). The average depth of coverage was 42.2 (90.8% of the genome > 30x). GATKv4-0-3 workflow was used for data processing and variant discovery. Structural variants were checked using Lumpy. We used Variant Effect Predictor (ensembl-VEP version 90.1) to annotate these discovered variants. The data was analyzed on Firecloud using GRCh38 for alignment and dbNSFP.v3.5a as the main annotation resource. First, all high impact variants as predicted from VEP's gene model with known clinical significance from ClinVar were retained. Then, we kept the moderate impact variants predicted to be deleterious by metaSVM and Combined Annotation Dependent Depletion (CADD) score > 15. Furthermore, we kept only variants with minor allele frequency (MAF) under 1% from GnomAD, 1000 Genomes, ESP6500, and ExAC databases.
Results.

We found 4,918,737 SNVs and indels, with 1,315 high and 14,309 moderate impact from VEP prediction. Among the known PS genes (TWNK, CLPP, HARS2, LARS2, and HSD17B4), only two heterozygous missense variants (NM_000414.3:c.317G>A (p.Arg106His) and c.1675A>G (p.Ile559Val)) in HSD17B4 were identified. Both variants were predicted to be benign. Ten other high impact variants were identified in genes with known clinical syndromes in ClinVar database (NFU1, ATP6V1B1, CC2D2A, MTTP, PAX4, LPL, ACTN3, ALDH3A2, UPK3A, and PDHA1). Additional 162 variants were predicted to be deleterious by MetaSVM; 12 likely benign variants were filtered out based on CADD score < 15 along with 65 variants with MAF > 0.01. Among the 85 variants left, one variant in HBB was reported in Beta-thalassemia and the other variant in ALX4 was reported in enlarged parietal foramina. We identified 311 SV regions; however, none of these SV overlapped with known PS genes.

POS 04 - Wednesday 15:30 - 16:15

Providing molecular diagnosis to patients with mitochondrial oxidative phosphorylation disease using functional studies following next generation sequencing

Sze Chern Lim[1], Yoshihito Kishita[1], Masakazu Kohda[1], Tomoko Hirata[2], Yukiko Yatsuka[1], Hiroko Harashima[3], Kei Murayama[4], Akira Ohtake[3] and Yasushi Okazaki[1,*].

[1]Intractable Disease Research Center, Graduate School of Medicine, Juntendo University, Tokyo, Japan.[2]RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.
[3]Department of Pediatrics, Saitama Medical University, Saitama, Japan.[4]Department of Metabolism, Chiba Children's Hospital, Chiba, Japan.
*ya-okazaki@juntendo.ac.jp

Background: To fulfill energy requirements for cellular activities, cells generate the majority of energy via oxidative phosphorylation (OXPHOS) in the mitochondria. The key enzymes in OXPHOS are complexes I, II, III, IV and V. Defects in OXPHOS can cause human disease and affect ~1:5,000 live births. Mutations causing OXPHOS disease have been identified in more than 280 genes. However, ~60% of OXPHOS patients remain without molecular diagnosis. In addition, the wide spectrum of clinical presentations and poor phenotype-genotype correlation in OXPHOS disease patients attribute to the challenge in confirming molecular diagnosis.

Methods: Patients with early-onset OXPHOS disease who have not received molecular diagnosis were recruited from Saitama Medical University Hospital and Chiba Children's Hospital. Patient DNA samples were analysed by next generation sequencing, and subsequent bioinformatic and functional analyses have been performed to confirm pathogenicity of novel candidate mutations.

Results: In one patient with Leigh Syndrome which is one of the best known childhood presentation for OXPHOS disease, we identified two novel compound heterozygous variants in UQCRC2, which encodes a subunit of OXPHOS complex III. Mutations in UQCRC2 were reported to cause OXPHOS disease. Bioinformatic analysis alone could not confirm the pathogenicity of the novel candidate mutations – one intronic sequence variant predicted to affect intron splicing and one exonic variant predicted to result in a substitution mutation at the C-terminal of the protein. Biochemistry and molecular genetic analyses have been performed to identify any evidence of a defective UQCRC2. Blue Native Polyacrylamide Gel Electrophoresis (BN-PAGE) western blot analysis showed increased level of OXPHOS complex III assembly intermediates in the patient fibroblasts. cDNA analysis is being performed to investigate the consequence of the intronic sequence variant.

Conclusions: This study highlights the important role of functional studies to confirm the pathogenicity of novel candidate mutations identified by next generation sequencing. Nevertheless, functional studies are often constricted by the unavailability of patient tissue samples with disease phenotype. In such case, gene knock-out/knock-in cell models generated using the CRISPR/Cas9 system can be instrumental.

Acknowledgements: We acknowledge the patient and his family, as well as physicians and collaborators for their participation and contribution in this study. This study is partly funded by the Japan Society for the Promotion of Science (JSPS) Kakenhi Grant and JSPS Postdoctoral Fellowship.

POS 05 - Tuesday 15:45 - 16:30

ChemGenome: An ab initio method for prokaryotic genome annotation

Akhilesh Mishra[a,b], Priyanka Siwach[a,c], B. Jayaram[a,b,d*]

[a] Supercomputing Facility for Bioinformatics & Computational Biology, Indian Institute of Technology Delhi, India, [b] Kusuma School of Biological Sciences, Indian Institute of Technology, Delhi, India, [c] Department of Biotechnology, Chaudhary Devi Lal University, Sirsa, Haryana, India, [d] Department of Chemistry, Indian Institute of Technology, Delhi, India,
Email: akhilesh@scfbio-iitd.res.in; bjayaram@chemistry.iitd.ac.in

Gene prediction, also known as gene identification, gene finding, gene recognition or gene discovery, is among one of the important problems of molecular biology and is receiving increasing attention due to the advent of large scale genome sequencing projects. We designed an ab initio model (called ChemGenome) for gene prediction in prokaryotic genomes based on physico-chemical characteristics of codons calculated from molecular dynamics (MD) simulations. The first module of the protocol (DNA space), requires a specification of three calculated quantities for each codon: the double-helical trinucleotide base pairing energy, the base pair stacking energy, and an index of the propensity of a codon for protein-nucleic acid interactions. As this three-dimensional vector moves along any genome, the net orientation of the resultant vector should differ significantly for gene and non-genic regions to make a distinction feasible. The predicted putative protein coding genes from above parameters are passed through a second module of the protocol (Protein space analysis) which reduces the number of false positives by utilizing a filter based on stereochemical properties of protein sequences. The chemical properties of amino acid side chains taken into consideration were the presence of sp^3 hybridized γ carbon atom, hydrogen bond donor ability, short / absence of δ carbon and linearity of the side chains / non-occurrence of bidentate forks with terminal hydrogens in the side chain. The final prediction of the potential protein coding genes are based on frequency of occurrence of amino acids in the predicted protein sequences and their deviation from the frequency values of Swiss-Prot protein sequences (Swissprot space). Thus, starting from the genome sequence of a given species, potential protein-coding genes are predicted by the ChemGenome protocol. The validation of the methodology has been done on 372 prokaryotic genome and the results are better than any known knowledge based computational protocols for genome annotation.

Acknowledgements:
Support from the Department of Biotechnology, Govt. of India to the Supercomputing Facility for Bioinformatics and Computational Biology (SCFBio), IIT Delhi, is gratefully acknowledged. PS extends thanks to Chaudhary Devi Lal University for granting Sabbatical Leave to her. AM is a recipient of UGC-SRF.

References:

1. Ankita Singh, Akhilesh Mishra, Ali Khosravi, Garima Khandelwal, B. Jayaram, "Physico-chemical fingerprinting of RNA Genes", Nucleic Acids Research, 2016.
2. G. Khandelwal, and B. Jayaram, "DNA-water interactions distinguish messenger RNA genes from transfer RNA genes", J. Am. Chem. Soc., 2012, 134, 8814-8816.

3. P. Singhal, B. Jayaram, S. B. Dixit and D. L. Beveridge, "Prokaryotic gene finding based on physicochemical characteristics of codons calculated from molecular dynamics simulations", Biophys J., 2008, 94, 4173-4183.
4. S. Dutta, P. Singhal, P. Agrawal, R. Tomer, Kritee, E. Khurana, and B. Jayaram, "A Physico-chemical model for analyzing DNA sequences", J. Chem. Inf. Model., 2006, 46, 78-85.
5. B. Jayaram, "Beyond the wobble: the rule of conjugates", J. Mol. Evol., 1997, 45, 704-705.

POS 06 - Wednesday 15:30 - 16:15

A rapid throughput computational screening to identify sequence specific DNA minor groove binders via physico-chemical descriptors

Pradeep Pant[1, 2] and B. Jayaram[1, 2, 3, *]

[1]Department of Chemistry, [2]Supercomputing Facility for Bioinformatics & Computational Biology, and [3]Kusuma School of Biological Sciences, Indian Institute of Technology Delhi, Hauz Khas, New Delhi-110016, India

We report here a rapid throughput screening method (RASDD) for identifying potential sequence specific DNA minor groove binders from a million molecule database. Physico-chemical properties such as Wiener index, molar refractivity, H-bond donor(s), H-bond acceptor(s), partition coefficient and curvature of 20 DNA-drug complexes with known binding energies were utilized to set up a QSAR-type equation in order to evaluate the binding free energy of the resultant complex without actually docking the drug in the DNA minor groove. A test set of 25 DNA-drug complexes showed that the predicted binding free energies correlate well with the experimental binding free energies (R = 0.85) with an rms error of 1.2 kcal mol^{-1}. The minor groove/AT rich region of any specified DNA sequence can be screened against a million compound repository within a few minutes, thus speeding up the process towards DNA based drug discovery. RASDD methodology is freely accessible at http://www.scfbio-iitd.res.in/software/drugdesign/rasdd.jsp.

POS 07 - Tuesday 15:45 - 16:30

S2F: protein function prediction tool from its sequence

Amita Pathak1,2, Sahil2, Arti Yadav1, B. Jayaram1, 2, 3, *

1Department of Chemistry, 2Supercomputing Facility for Bioinformatics & Computational Biology, 3Kusuma School of Biological Sciences,

Indian Institute of Technology, Hauz Khas, New Delhi-110016, India

*Corresponding author

amita@scfbio-iitd.res.in, bjayaram@chemistry.iitd.ac.in

Characterizing the functions of proteins coded by the genomes is one of the key challenges of the current era. The rate of in-flow of sequence data has far exceeded the speed with which protein functions are annotated. It is not a trivial task to experimentally determine the function of all the proteins. With ever increasing protein sequences in UniProtKb database i.e ~557,275 total number of manually curated sequences, only ~71,229 are annotated experimentally. Similarly, out of ~20,328 human proteins, only ~15005 have been annotated [1]. Hence, a fast yet accurate method of function annotation is a squeezing need. Computational methods can be useful in predicting functions of proteins using known sequence information. Various computational approaches are available for protein function annotation but these are wanting in specificity and precision. Here, we present S2F, a meta-server which integrates some popular protein annotation programs with different approaches [2] and furnishes the user with a set of top consensus functional terms at a single platform. The meta-server approach has enabled the annotation of complete human proteome including ~5,323 proteins which were hitherto unannotated. The S2F meta-server is available for free access at www.scfbio-iitd.res.in/proteins/S2F.jsp/.

References

1. UniProt Consortium. "UniProt: a hub for protein information." Nucleic acids research 43.D1 (2014): D204-D212
2. Rentzsch, Robert, and Christine A. Orengo. "Protein function prediction–the power of multiplicity." Trends in biotechnology 27.4 (2009): 210-219

POS 08 - Wednesday 15:30 - 16:15

Pattern of germline genetic variants in cancer predisposing genes in a Sri Lankan cohort with inherited cancer syndromes

Nirmala D. Sirisena*, Nilaksha Neththikumara, Sajeewani Pathirana, Dineshani Hettiarachchi, Kalum Wetthasinghe, Vajira H. W. Dissanayake
Human Genetics Unit, Faculty of Medicine, University of Colombo, Colombo 08, Sri Lanka
nirmala@anat.cmb.ac.lk

Background: Even though cancer incidence is on the rise in Sri Lanka, only a limited number of studies have described the genetic variations in selected cancer predisposition genes (CPGs) in the Sri Lankan population. So far, all reported local studies were based on germline variations in the BRCA1 and BRCA2 genes using conventional Sanger sequencing technology. The pattern of germline genetic variations in CPGs in the Sri Lankan population using Next Generation Sequencing (NGS)-based testing has not previously been studied. This study aims to describe the frequency and spectrum of germline genetic variants in CPGs in a Sri Lankan cohort with inherited cancer syndromes.

Methods: 97 consecutive patients with hereditary cancer predisposition who underwent whole exome sequencing using the Illumina® MiSeq® NGS-based platform are reported. Bioinformatics analysis was performed through an in-house developed pipeline to detect germline genetic variants in CPGs. In order to arrive at a final variant classification, all variants underwent thorough assessment and review of available evidence such as population frequency information, published case reports, case/control and functional studies, internal co-occurrence and co-segregation data, evolutionary conservation, and in-silico functional predictions.

Results: 74 (76.3%) were affected with cancer. 23 (23.7%) were pre-symptomatic. Majority 81 (83.5%) were female. The median age at onset of cancer in the affected group was 46 years [range: 4-82 years]. The commonest inherited cancers were breast cancer 31 (41.9%) and colorectal cancer 13 (17.6%). Germline variants were identified in 45 (60.8%) cancer affected patients and 9 (39.1%) pre-symptomatic individuals with family history of cancer. Variants were detected in the following 17 CPGs: BRCA2 - 13 (24.1%); BRCA1 - 7 (13.0%); MLH1 - 7 (13.0%); PMS2 - 5 (9.3%); MSH2 - 4 (7.4%); ATM - 4 (7.4%); CHEK2 - 2 (3.7%); APC - 2 (3.7%); MLH3 - 2 (3.7%); FANC1 - 1 (1.9%); BRIP1 - 1 (1.9%); BARD1 - 1 (1.9%); STK11 - (1.9%); MEN1 - (1.9%); CDKN2A - 1 (1.9%); PALB2 - 1 (1.9%); and ALK - 1 (1.9%). They consisted of: non-synonymous variants - 39 (72.2%); small deletions - 11 (20.4%); insertions - 2 (3.7%); synonymous variants - 1 (1.8%); and duplications - 1 (1.8%). They were clinically classified as: variants of uncertain significance (VUS) - 23 (42.6%); pathogenic - 19 (35.2%); and likely pathogenic - 12 (22.2%). 8 (14.8%) were novel variants which have not previously been reported in the scientific literature. They included: CDKN2A:c.377A>C:p.Gln126Pro; PALB2:c.2768T>G:p.Val923Gly; BRCA1:c.3392A>G:p.Asp1131Gly; APC:c.7781C>G:p.Ser2594Cys; MLH1:c.469dupT:p.Tyr157fs; PMS2:c.779C>T:p.Ser260Phe; PMS2:c.2212G>T:p.Val738Phe, and MEN1:c.1365+1G>C.

Conclusions: This study describes the pattern of germline genetic variants in 17 CPGs in a Sri Lankan cohort with inherited cancer syndromes. The findings of a significant number of VUS and novel variants highlights the importance of undertaking NGS-based testing in under-represented populations with inherited risk of cancer. Further investigations would be needed to delineate the functional significance of the novel and unclassified genetic variants identified in this study.

POS 09 - Tuesday 15:45 - 16:30

Whole Exome Sequencing Identifies Potentially Causative Variants in Brugada Syndrome-alike patients

Nutchavadee Vorasan[1,6], Kunaphas Kongkittimanon[7], Sakda Sathirareuangchai[2,6], Ekkapong Roothumnong[3,6],Wanna Thongnoppakhun[4,6], Rungroj Krittayaphong[3,6], Warangkna Boonyapisit[3,6], Manop Pithukpakorn[3,6], Satchana Pumprueg[3,6], Bhoom Suktitipat [5,6,7*]
[1] Research Division, [2] Department of Forensic Medicine, [3] Department of Medicine, [4] Department of Research and Development, [5] Department of Biochemistry, [6] Faculty of Medicine Siriraj Hospital, Mahidol University, Thailand, [7] Integrative Computational BioScience Center (ICBS), Mahidol University, Thailand

*Correspondence: bhoom.suk@mahidol.edu

Introduction
Brugada Syndrome (BrS) is a disease producing abnormal electrical activity in the heart, which increases the risk of sudden cardiac death among young healthy adults in several countries. The high incident rate of sudden cardiac death was found in Southeast Asia. We studied potentially common causative variants among patients with diagnostic electrocardiogram (EKG) abnormality similar to BrS using whole exome sequencing (WES).

Materials and methods
BrS-like patients were recruited from Siriraj Hospital. The blood samples were collected and WES performed using Agilent SureSelectXT v5 kit. List of related candidate genes were assembled from previous studies [1]. Combined variant call of BrS samples were performed using GATKv3-8 workflow (based on GRCh37 reference) and were annotated by Variant Effect Predictor (VEP) web interface. An in-house R script was used for variant prioritization. Variants functions with high and moderate impact were selected for further analysis. We excluded variants predicted by metaSVM to be tolerable or benign variants reported in ClinVar. Moreover, we filtered out the common variants with allele frequency >10% in public databases, such as 1000 Genome Project, Exome Aggregation Consortium (ExAc), ESP6500 Project, and Genome Aggregation Database to potentially pathogenic of BrS.

Results
A total of 15 patients with BrS were included in the study. Nineteen non-synonymous mutations variants of BrS were summarized from variants filtering by R script. From the WES data among patients, 80% (12/15 = 80%) presented the potential causative of BrS or SUDS. The majority of patients (6/15 = 40%) had a mutation in the TTN gene, followed by RBM20 (13%), and SCN5A (13%).

Conclusion
WES is a useful approach helping to improve the diagnostic rate and differentiate patients BrS from BrS-alike from EKG. With VEP and R script this process could be streamline to of improve the efficiency in clinical management among the patients or relatives of patients with BrS and SUDS.

Funding:
This work was supported by Mahidol University (Grant R015837002 to WB); Siriraj Core Research Facility (SiCRF) Grant to MP; Siriraj Chalermphrakiat Grant to BS, SS, WT, WB, MP and RK; the

Crown Property Bureau Foundation Grant to BS; Thanapat Fund (D003752) to MP, Faculty of Medicine Siriraj Hospital, Mahidol University, Grant Number (IO) R015734003 to BS.

References

1. Suktitipat B, Sathirareuangchai S, Roothumnong E, Thongnoppakhun W, Wangkiratikant P, Vorasan N, et al. Molecular investigation by whole exome sequencing revealed a high proportion of pathogenic variants among Thai victims of sudden unexpected death syndrome. PloS one. 2017;12(7):e0180056-e. Available from:https://doi.org/10.1371/journal.pone.0180056

POS 10 - Wednesday 15:30 - 16:15

A Case of a large inherited Benign Copy Number Variant (CNV)

Yong MH

Department of Pathology and Laboratory Medicine, KK Hospital, Singapore

Copy number variations (CNVs) have been identified with their important roles in the genomic diversity in human population and typically range in length from 1kb to 5Mb. The development of next-generation sequencing (NGS) technologies have redefined the size of SNVs to larger than 50 bp. However, even with the introduction of molecular methods into routine diagnostic services over the last two decades, the clinical interpretations of some CNVs continue to be a major challenge.

We report a case of a 15 Mb duplication (chr21:14693431-29688753 [hg19]) ascertained in a five-year old boy with global developmental delay and profound bilateral hearing impairment via Chromosome Microarray Analysis (CMA). This cytogenetically visible duplication includes three OMIM morbid genes, namely, LIPI, TMPRSS15 and APP. It was later found to be inherited from his clinically normal father.

A large 15 Mb duplication in a phenotypically normal individual is uncommon. Our study suggests that this duplication is likely to be a benign CNV.

MEETING ROOMS

Plenary Room - Iskandar Demonstration Theatre

Workshops - IT Cluster 1 & IT Cluster 2 (in Sir Christopher Edwards Building)

Coffee Breaks & Lunch - Fame Cafe