*Review*

# The Application of Whole-Exome Sequencing in Diagnosing Pediatric Rare Disease in Hong Kong

Christopher CY Mak, MBChB;^ Dingge Ying, PhD;^ Ka Chun Leung, BSc, MPhil; Pamela PW Lee, MBBS; Jing Yang, PhD; Wanling Yang, PhD; Brian HY Chung, MBBS, FCCMG;* Yu-Lung Lau, MD

*Department of Paediatrics & Adolescent Medicine, LKS Faculty of Medicine, University of Hong Kong, Hong Kong SAR, China*

**Whole-exome sequencing (WES) combines next generation sequencing (NGS) technology with capture methods to sequence all the coding regions of the genome. The application of WES has gained a lot of success worldwide in discovering new disease causing genes and in diagnosis. We reviewed some of the large collaborative efforts worldwide and summarized notable examples, observing an overall diagnostic yield of 16-30%. Locally in Hong Kong, there have been several applications of WES in research, as well as bioinformatics tools developed, and the field is continuing to grow. In our own department, we have applied this to pediatric rare diseases, by establishing our in-house research pipeline for WES, as well as utilizing core laboratory facilities in model animals and cell work for functional validation. We illustrate this approach using cases as examples. On the other hand, clinically, we are utilizing WES more as a diagnostic tool by analyzing selected pediatric cases via overseas laboratories. We see how this new tool is helping patients and families to obtain an answer for their condition, and subsequently helping them with their management and family planning. Finally, we discuss the challenges for WES in Hong Kong, and the future direction of the technology, with the potential to revolutionize clinical diagnosis and medical research.**
*[N A J Med Sci. 2014;7(4):139-147.* DOI: *10.7156/najms.2014.0704139]*

**Key Words:** *whole exome sequencing, Hong Kong, pediatric, rare disease, genetics, bioinformatics, counselling*

## INTRODUCTION

The human genome contains around 3 x $10^9$ bases, but only about 1-2% codes for proteins. Whole exome sequencing (WES), targets most of the coding and non-coding exons of the human genome, which is thought to harbor more than 85% of known disease causing mutations.[1] With the rapid decline in cost of DNA sequencing, and the development of high-throughput genomic technologies, it is now possible to sequence the whole exome or even genome of a human being, at a price that is more affordable, and hence more suitable for clinical use as well as research.[2,3] The technology has the potential to transform genetic diagnostic services for families with rare inherited diseases, and may also reduce the lengthy quest and high cost of many genetic tests.[4,5] Implementation of the WES service requires a combination of collaboration from clinical services, a reliable and methodical approach to obtain high quality sequencing data, and the right standards for the interpretation of the findings. The following review will focus on the technology of WES, its application worldwide, and how it has been utilized in the pathways of diagnosing pediatric rare diseases in Hong Kong.

Received: 09/12/2014; Revised: 09/28/2014; Accepted: 09/30/2014
***Corresponding Author:** Department of Paediatrics & Adolescent Medicine, Queen Mary Hospital, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong SAR, China.
Tel: 852-22554482. Fax: 852-28551523. (Email: bhychung@hku.hk)
^ authors have equal contribution

## NEXT-GENERATION SEQUENCING AND WHOLE-EXOME SEQUENCING

The automated Sanger sequencing, which had dominated the DNA sequencing industry for more than two decades, is considered as a 'first-generation' technology. Newer methods developed in more recent years are now referred to as 'next-generation' sequencing (NGS) or 'second-generation' sequencing. The main difference between Sanger sequencing and NGS is the advantage of 'massively parallel sequencing', which allows the sequencing of a large number of reads at the same time. When NGS technology is combined with 'exome capturing', WES of all the coding regions becomes possible.

In general, the process of WES starts with the fragmentation of DNA from an individual into small fragments (100-500 bp), which are then hybridized to prefabricated DNA probes on reactive surfaces designed for known coding regions. These sequences are next purified and amplified using polymerase chain reaction, completing the exome capture step.[6] Unique adaptors are added for the DNA fragments, and reactive surfaces that contain probes for these adaptors immobilize all the DNA fragments in the sequencing machine. The immobilized DNA fragments are then extended and amplified as a cluster. The sequencing and imaging cycle starts with adding all four nucleotides, each labeled with a different dye. The probes are then extended one of the four nucleotides and the remaining unincorporated nucleotides are

washed away. Imaging is then performed to determine the identity of the incorporated nucleotide. Imaging information from each cycle forms the sequencing information and quality of millions of reads, which are regarded as the raw yield of the WES technology. Illumina, Agilent and NimbleGen provide different capturing technologies with different total target size, functional regions and number of probes. The current sequencing platforms are mainly produced by Illumina.[7]

The main application of WES is to locate all types of functional variants from the sequenced reads, including insertions, deletions, missense variants, nonsense variants, splice variants and more recently, copy number variants.[1] The first proof of principle for the application of WES was performed by Ng et al in 2009, who initially tested the method by confirming a known gene of Freeman-Sheldon syndrome (FSS).[8] The same group then applied the technique to identify a novel gene for Miller syndrome (OMIM 263750) in 2010.[9] Since these successes, the use of WES technology to identifying disease causing mutations has grown exponentially.

**Table 1.** Examples of Large Whole-exome Sequencing Projects Worldwide.

| Study | Start Date | Brief Description | Institutions involved | Diseases studied | Results and Publications |
|---|---|---|---|---|---|
| Finding of Rare Disease Genes in Canada (FORGE) | April 2011 | Canada wide collaboration to study rare disease in children including more than 1,000 Canadian samples and 300 international samples. 264 from 371 submitted disorders were selected for WES. | 21 genetics centers and 3 science and technology innovation centers from Canada. | Conditions of pediatric onset, most likely monogenic and have a molecular etiology not understood. | Identified 67 genes not previously associated with human diseases, 41 validated. Success rate of 16%–25% for novel-gene discovery. Listed 20 gene discovery publications as of Jan 2014. |
| UK Deciphering Developmental Disorders (DDD) | April 2011 | UK wide project aiming at systemic phenotyping and detailed genomic analysis of 12,000 children with severe undiagnosed developmental disorders, using a combination of aCGH, SNP genotyping, and WES. | 24 Regional Genetics Services in UK and Ireland | Developmental Disorders | Progress from May 2014, nearly 9,000 children and their parents recruited (target 12,000), analysis of first 1,133 families completed, uncovered 12 new genes. Overall diagnostic rate of 30%. Listed 7 publications. |
| UK10K Project | March 2010 | Welcome Trust Sanger Institute project to understand the link between rare genetic variants and disease. Genome wide sequence of 4000 samples, and the exome sequence of 6000 samples selected for extreme phenotypes. | Led by Welcome Trust Sanger Institute involving various universities across UK as well as US | Contains a rare disease set which includes congenital heart disease, familial intellectual disability etc. | Listed 20 Publications on project website, identifying various novel disease causing genes. |
| NIH Undiagnosed Disease Program | May 2008 | NIH funded project, to study undiagnosed diseases. Up to 2014, 750 cases accepted, 40% pediatric. SNP array scans performed for approximately 1,600 individuals. WES analyses performed for approximately 900 individuals. Strength of detailed phenotyping and clinical evaluation. | 6 Key clinical sites from 2014 | Undiagnosed Disease | Out of 640 patients accepted in the first 4 years, diagnoses was provided for around 150 patients, translating to a diagnostic rate of 24%. Examples of disorders include congenital disorder of glycosylation IIb, and ACDCa. |
| NIH Mendelian exome program | March 2012 | Aim to discover new genes for Mendelian disorders by using NGS technologies. | 3 funded centers | Mendelian Disorders | 91 Publications listed on website, in high impact journals including Nature, JAMA, Cell etc. |
| The Congenital Heart Disease Genetic Network Study | December 2010 | A network study to investigate relationships between genetic factors, clinical features, and outcomes in Congenital Heart Disease. Proposed WES of trios, CNVs, candidate gene sequencing, tissue gene expression etc. | 6 main and 4 satellite sites at which subjects are recruited | Congenital Heart Disease | Up to June 2012, 3772 probands recruited. Proband median age is 5.5 years. One or both parents were enrolled for 72% of probands. Publications include a Nature paper identifying 10% de novo mutations in a large cohort of 362 CHD patients. |

[a] Arterial calcification due to deficiency of CD73 (ACDC)

## APPLICATION OF WES WORLDWIDE

From 2010, there have been numerous publications worldwide identifying causal mutations for diseases, as well as larger collaborative efforts to identify pathogenic variants. Examples of major international efforts are summarized in **Table 1**.[10-18] Most of these projects were started very soon after 2010, when the application of WES technology emerged. Each of these projects focuses on a specific area and uses a slightly different approach, but all involve the use of WES in a well-organized pipeline. The result is the identification of multiple novel disease-causing genes. The overall diagnostic rate using WES after using more conventional genetic investigations, ranges from 16% to 30%

(**Table 1**). There are also numerous publications resulting from these data sets, not just limited to the discovery of genes for diseases, but also other findings such as the impact of incidental findings from WES.[19]

As a result, there is a massively growing literature pool of WES findings. To illustrate this, at the time of writing this review in August 2014, a PubMed search limited to 'Title/Abstract' with the term 'whole-exome sequencing' revealed over 1,200 cumulative articles on the topic, while a similar search in May 2012 only found 102 publications.[20]

To look at an example of these international efforts in detail, the Canadian consortium, FORGE, performed WES analysis of 783 samples and identified 67 novel genes for 146 rare disorders in children over the first 2 years. The approach included grouping samples into 4 different categories; using very similar pipelines between centers for discovery; and subsequent validation by each team.[21] The data analysis was coordinated at The Hospital for Sick Children and the University of Toronto. Resources utilized included the Agilent SureSelect 50 Mb (V3) All exon Kit, and sequencing using Illumina HiSeq 2000. The informatics tools used in their pipelines are very much similar to those used worldwide, as well as our own pipeline in Hong Kong. Such tools include Burrows-Wheeler Aligner, Picard, GATK, SAMtools and ANNOVAR. Overall, there was a success rate of 16-25% in identifying a novel gene for the 264 disorders. Nevertheless, 118 out of 264 disorders remain unsolved by WES in this large effort, which is a reflection of the limitations and difficulties of exome sequencing.[21] As for examples of clinical application of the technology, WES had been successful in identifying de novo genetic changes in severe intellectual disability, claiming a diagnostic rate of 16%[22] - 55%.[23] More recently, in one of the largest clinical cohort published, Yang et al summarized the use of WES in 250 probands.[24] In this study, WES was ordered by physicians mainly from genetics, pediatric specialties. Among them, 80% were children with neurological phenotypes, and most patients referred were less than 18 years of age. WES was performed after a combination of genetic testing, aCGH and sequencing studies. Through a similar pipeline, and using an in-house annotation system, there was a molecular diagnosis obtained in approximately 25% of cases. Overall, this is thought to be a higher yield than other tests e.g. karyotype analysis (up to 15%) and aCGH analysis (up to 20%), and has implications on the strong potential of WES in the clinical setting to help physicians reach a diagnosis.[24]

The number of disease genes identified worldwide is growing as a result of these collaborations, and some of these projects are growing even further with more funds being injected into them.[25] With new technologies emerging, such as the detecting of copy number variations with exome sequencing data,[26,27] a greater proportion of these rare disease variants will be discovered, and is expect to translate into better understand of disease and improved clinical care.

**Table 2.** Reported application of WES in disease diagnosing in Hong Kong.

| Disease | Rare? | University | Children? | Samples | WES Platform | Mutation | Disease Model |
|---|---|---|---|---|---|---|---|
| Familial spastic paraplegia | Yes | HKU | No | 4 patients and 2 unaffected in family | Agilent SureSelect and Illumina GAIIx | Inherited novel heterozygous missense mutations in *PMCA4* | Autosomal Dominant |
| Retinitis Pigmentosa | Yes | CUHK | No | 2 patients and 1 unaffected in family | TIANGEN and Illumina HiSeq 2000 | Inherited novel heterozygous missense mutations in *RHO* | Autosomal Dominant |
| Dilated cardiomyopathy | No | HKU | No | 1 patient | Agilent SureSelect and Illumina GAIIx | Inherited novel heterozygous mutation in *DES* | Autosomal Dominant |
| Spinocerebellar ataxias | Yes | HKU | Yes | 2 patients in a family | NimbleGen and Illumina GAIIx | Inherited novel heterozygous mutations in *TGM6* | Autosomal Dominant |
| Neonatal-onset Crohn's disease | Yes | HKU | Yes | 1 trio with 1 patient and 2 healthy parents | Agilent SureSelect and Illumina GAIIx | Inherited compound heterozygous missense mutations in *IL-10RA* | Autosomal Recessive |
| Disseminated *Penicillium marneffei* | Yes | HKU | Yes | 1 patient | Agilent SureSelect and Illumina HiSeq 2000 | De novo heterozygous missense mutation in *STAT1* | Autosomal Recessive |

**Table 3.** WES related softwares developed in Hong Kong.

| Software | University | Application | Feature | Availabilty |
|---|---|---|---|---|
| BALSA | HKU | NGS raw reads to variants | fast, utilize the power of GPU | Free to download www.bio8.cs.hku.hk/dataset/BALSA/ |
| KGGSEQ | HKU | NGS data analysis and prioritization of NGS variants | efficient & comprehensive framework | Free to download and online statgenpro.psychiatry.hku.hk/limx/kggseq/ |
| Privar | HKU | Prioritization of NGS variants | a systematic prioritization pipeline | Free to download paed.hku.hk/uploadarea/yangwl/html/software.html |
| EFIN | HKU | Evaluation of variants | predicting the functional impact of non-synonymous SNV | Free online paed.hku.hk/efin/ |

## WHOLE-EXOME SEQUENCING RESEARCH IN HONG KONG

In Hong Kong, two of the largest universities both have their own capabilities of performing NGS for research. Most of this is performed in Centre for Genomic Sciences (CGS) in the University of Hong Kong (HKU) and Core Facilities Genome Sequencing Laboratory (CFGSL) in the Chinese University of Hong Kong (CUHK). CGS and CFGSL offer professional core services with advanced high-throughput technology platforms using Illumina GAIIx and Roche 454 GS FLX, and CGS also has High-Performance Computing Facility with disk storage of ~ 320 TB through 10 Gigabit network and 24 x 64 GB CPUs. There have been various research efforts in WES research in Hong Kong, including application in rare diseases (**Table 2**) and in development of bioinformatics tools (**Table 3**). We will discuss briefly 3 published examples of application in pediatric rare diseases as well as a few bioinformatics tools developed locally.

The first application of WES in pediatrics in Hong Kong was published by Mao and Yang et al in 2012.[28] WES was performed on a trio, where the child was affected by neonatal-onset Crohn's disease (CD). Novel compound heterozygous missense mutations were found in the affected child in the extracellular domain of *IL-10R1*. Functional analysis of the gene showed that these mutations abrogated IL-10R1 phosphorylation induced by IL-10, therefore leading to impaired STAT3 activation and suppression of inflammatory responses. After reconstitution with wild-type IL-10R1, the patient's cells showed fully restored IL-10R function including IL-10-induced STAT3 activation and expression of suppressor of cytokine signaling 3, confirming the causality of the *IL-10R1* mutations detected by WES. Interestingly, at around the same time, Glocker et al[29] also identified homozygous mutations affecting the same receptor by linkage studies using microsatellite markers followed by Sanger sequencing confirmation, and published their findings. These findings were published shortly before the group in Hong Kong who identified the mutations by WES approach.

The second report was on a Chinese family with spinocerebellar ataxias (SCA). One affected child and one affected adult in the family were subjected to WES.[30] Through bioinformatics analysis of the sequence variants in these two individuals, a pathogenic mutation in the *TGM6* gene (c.1528G>C) was identified, which showed perfect co-segregation with disease phenotype in all nine members of this family. *TGM6* has previously been reported to be a causative gene for SCA from a WES of two Chinese families, showing that mutations in *TGM6* are an important consideration in Chinese patients with SCA.[31]

Another reported WES application in pediatrics in Hong Kong was on three unrelated Chinese children affected by disseminated *Penicillium marneffei (PM).*[32] A novel heterozygous missense mutation in *STAT1* was detected in one of the patients by WES and then confirmed in the other two patients. When the patients were reviewed after this finding, the diagnosis of concurrent diagnosis of chronic mucocutaneous candidiasis (CMC) was identified. Parental testing showed that the mutations were all *de novo*. Missense mutations affecting the *STAT1* coiled-coil domain were demonstrated to be gain-of-function in nature. By flow cytometric analysis of intracellular phosphorylated STAT1, lymphocytes from all patients demonstrated a significantly higher percentage of pSTAT1+ cells and increased phosphorylation intensity in response to IFN-a and IFN-g stimulation. The key significance of this finding is the importance of *STAT1* mutations in disseminated *Penicillium marneffei*, which is endemic in Southeast Asia, expanding the known phenotype association of *STAT1* with CMC. The notion of doing WES in these cases of PM lead to this discovery, and not only challenges the 'one gene one disease' assumption, but also further suggests that one or several genes may be associated with mixed phenotypes.

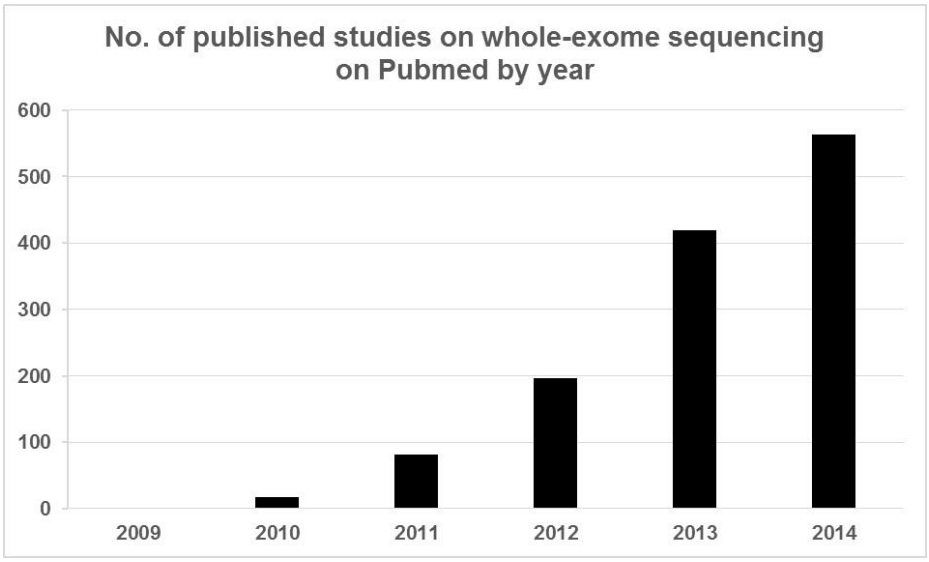


**Figure 1.** Number of published studies on "whole-exome sequencing" on PubMed by year. (Source: PubMed August 2014)

On a different note, various research teams in Hong Kong are focusing on bioinformatics development for NGS and WES. **Table 3** summarizes all the published WES tools developed in Hong Kong. BALSA, an integrated solution for the secondary analysis of WES/WGS data, performs much faster and accurate than previous tools by exploiting the computational power of GPU and integrating memory management.[33] BALSA also supports efficient identification of somatic SNVs and CNVs. Privar is designed as a systematic prioritization pipeline that takes the following factors into consideration: calling quality of the variants, their predicted functional impact, known connection of the gene to the disease and the number of mutations in a gene, and inference from linkage analysis.[34] EFIN is an online tool that can predict the functional impact of amino acid substitutions caused by non-synonymous mutations in the human genome.[35] It makes better use of sequence conservation information by grouping the homologous protein sequences into six blocks according to evolutionary distances to human and evaluating sequence conservation in each block independently. All these tools help the application of WES in rare disease to be more efficient and accurate.

Furthermore, NGS technology has also been applied in targeted sequencing with a novel approach by local research teams. The aim is to detect genetic variants on target regions of specific rare diseases, such as PID, in a large number of samples. Different sets of amplification primers on the target region with barcoding adapters can be designed, and multiple amplicons from different samples can be mixed together for further sequencing on a high throughput sequencing machine. After sequencing, the amplicon sequence can easily be distinguished through the pair-end barcode. This strategy can even allow the sequencing of thousands of samples from different individuals at the same time, saving the need to do repeated Sanger sequencing.

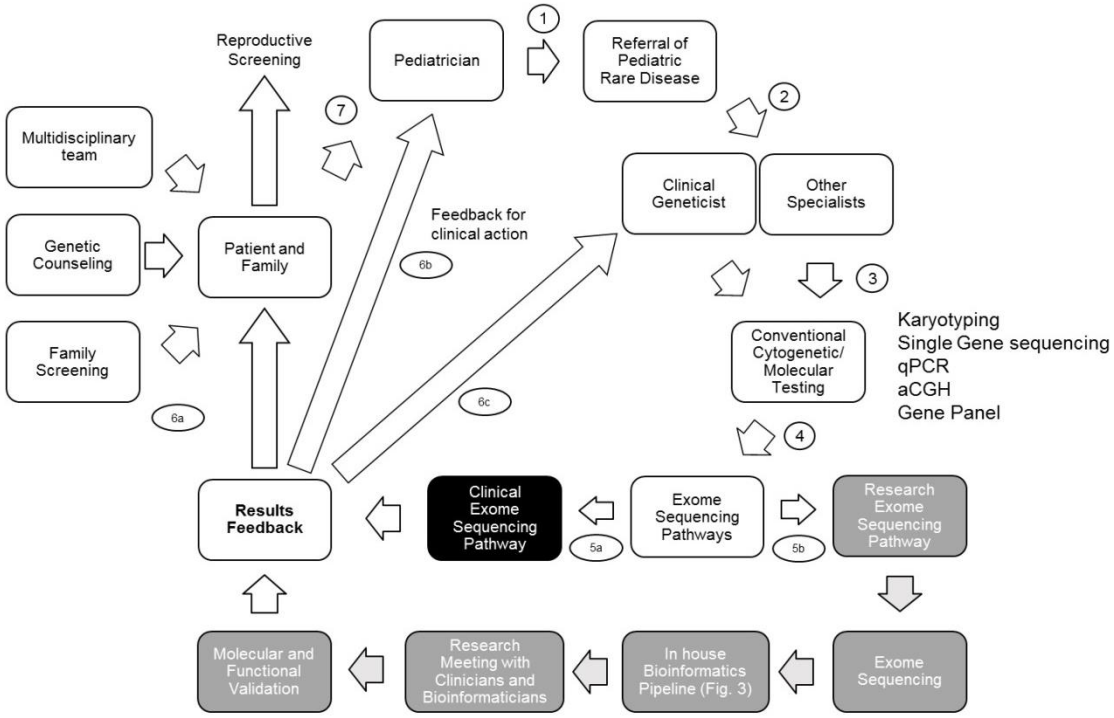


**Figure 2.** Pathways of using WES to diagnosing Pediatric Rare Disease in Hong Kong. Starting from step 1, the Pediatrician refers the case of rare disease to the geneticists and/or specialists for assessment, and conventional cytogenetic and molecular testing is first performed. When these tests are negative, either a clinical or research exome sequencing pathway is used to search for a molecular diagnosis. The results are then fed back to the patients and family via the geneticist and genetic counselor. The multidisciplinary team can provide further interventions based on these results, such as genetic counseling, family screening, reproductive screening and clinical action.

## STRATEGY OF DIAGNOSING PEDIATRIC RARE DISEASE

The Department of Pediatrics and Adolescent Medicine at the Queen Mary Hospital is affiliated with the University of Hong Kong. It is a major center offering tertiary-level care to children with various rare diseases, including congenital malformation, immunodeficiency, inborn errors of metabolism and complex neurological disorders, in Hong Kong. When a patient is referred to our clinical geneticist or other specialists e.g. immunologists for evaluation, detailed clinical assessment will determine the provisional differential diagnoses and formulate appropriate investigations, including various cytogenetic and molecular genetic testings. For patients with unexplained neurodevelopmental disorders or multiple congenital malformations, we adopted the international consensus and offer chromosomal microarray as the first-tier test.[36] Various NGS-based gene panels are also under development currently. WES will be considered if these 1st and 2nd tiered investigations cannot offer a definitive diagnosis. This can be provided via the clinical pathway or

in-house research pathway, with appropriate pre-test genetic counseling and informed consent from the parents/guardians (**Figure 2**). The results of one of these pathways are fed back to the referring pediatrician for necessary clinical actions including further investigations or disease/complication surveillance, and importantly to the families for genetic counselling, family screening, and reproductive options for future pregnancies.

## CLINICAL WHOLE-EXOME SEQUENCING PATHWAY FOR PEDIATRIC RARE DISEASE IN HONG KONG

The clinical WES pathway utilizes WES of overseas service laboratories to help reach a diagnosis clinically. Recently, Jamal et al. reviewed and compared services offered by 6 laboratories offering clinical WES service in the States.[37] Each trio test costs somewhere in the region of USD $4500-$9000, with a turnaround time of around 8-16 weeks. These overseas tests are not supported financially by the Department of Health/ Hospital Authority of Hong Kong and are usually paid out of pocket by parents. From our experience, medical insurance in this part of the world does not cover WES as a clinical test. Nevertheless, the WES results provided by these services can be very helpful, and have assisted clinicians to solve difficult diagnostic scenarios, and in particular when an urgent answer is required for making treatment decisions (Turn-around-time of 6-8 weeks can be offered on special request by some WES laboratories).

A good illustrative case was a baby with poor growth, hypotonia and dysmorphic features at birth. Initially physicians were unable to find the diagnosis despite various clinical tests, karyotyping and qPCR for 22q11.2 deletion. Whole-exome analysis found a missense mutation at the *ATRX* gene, suggesting the diagnosis of Alpha-thalassemia/mental retardation syndrome (ATRX), X-linked (MIM 301040). When reviewing the patient, the dysmorphic features were consistent with the published phenotype. The positive result not only helped identify what is perhaps the first case of ATRX in China, but also helped the geneticist formulate the need to perform further investigations such as checking for HbH inclusion bodies for alpha thalassemia, and ultrasound scan to exclude an enlarged spleen. Knowing the X-linked nature of the mutation and having description of the phenotype allows the appropriate counselling and family planning advice to be given. This was a very good example of how WES provided diagnosis of a rare disease, which would otherwise be difficult to ascertain on clinical grounds.

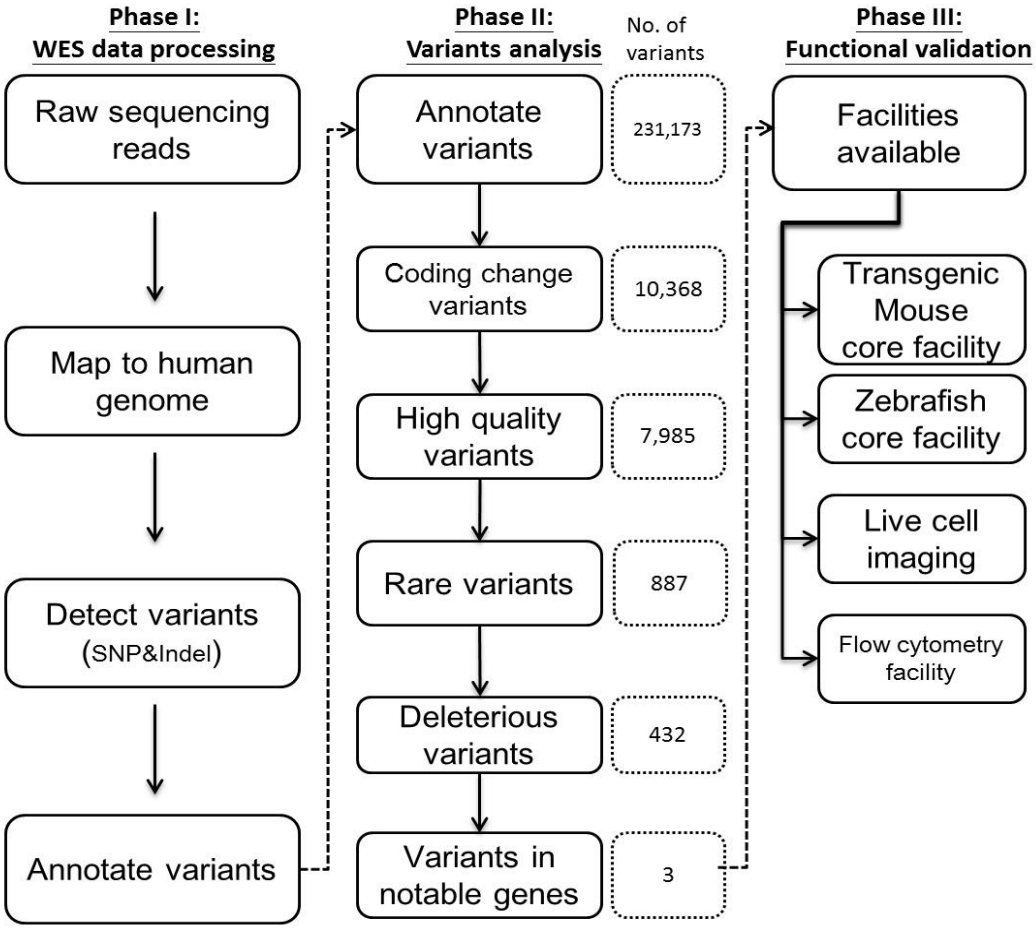


**Figure 3.** Pipeline and illustration of analyzing WES data. Phase I shows the standard pipeline for generation of annotated variants from raw sequencing reads. Most of the steps follow GATK best practice. Phase II shows the steps to figure out potential variants or genes that may responsible of the corresponding disease. Numbers to the right of the diagram indicates the number of variants remaining after each step for our illustrative case with PACS1 mutation. Phase III lists the facilities that can be applied for the functional validation of the mutations and genes.

## RESEARCH WHOLE-EXOME SEQUENCING PATHWAY

Alternatively, WES is offered at a research basis by various investigators within the department with special focus on immunodeficiency, congenital malformations and neurodevelopmental disorders. A standardized analytic pipeline has been established (**Figure 3**). To illustrate this, we will briefly describe how WES led to the diagnosis in a clinical case. The proband presented to us in infancy with global developmental delay, bilateral sensorineural hearing loss, facial dysmorphism, ectopic anus and duplex kidney. Clinical assessment by the clinical geneticist, neurologist and developmental pediatrician, together with chromosomal microarray, MRI brain and several other baseline metabolic investigations failed to reach a diagnosis. WES via the research pathway was offered with informed consent to the proband.

In brief, after receiving the raw sequencing reads from WES, they were mapped to the latest human genome using Burrows-Wheeler Aligner (BWA) together with quality scores of the bases.[38] GATK (The Genome Analysis Toolkit, http://www.broadinstitute.org/gatk/) was then used to realign the reads new common indels and to recalculate the sequencing qualities by normalization before calling the mutations, which includes single nucleotide variations and indels. ANNOVAR was used to annotate all the mutations, including the gene information, amino acid change, and various bioinformatic scores predicting conservation and pathogenicity of the coding change mutations, such as SIFT, PolyPhen2, etc.[39-41] A total of 231,173 variants were detected and annotated by WES. Only 10,368 variants were causing coding change of certain transcripts. By filtering out the mutations with low qualities (coverage < 10x, mapping quality < 30 and variant quality < 30), 7,985 variants remained. Variants with non-reference allele frequency larger than 1% either in 1000 genome project and NHLBI-ESP 6500 exome project or local population databases were filtered out, resulting in 887 rare variants, which can now be regarded as mutations. Around half (432) of the mutations were predicted as damaging by 4 out of the 6 pathogenicity prediction tools and were further screened by gene function and phenotype correlation.

Only 3 heterozygous mutations were identified to be related with global developmental delay/intellectual disability. Two of these mutations were validated to be *de novo* by Sangar sequencing on the child and parents. No mutations were identified by using a recessive model (**Figure 3**). With these 3 remaining mutations, a systematic literature review performed and as well as utilizing databases such as HGMD. Finally the *PACS1* mutation was found, to be relevant and was thought to correlate with the phenotype after discussion in a meeting with clinicians, geneticists and bioinformaticians. Interestingly, the mutation was only recently identified in patients with ID in 2012, and was identifed in our pipeline in 2013,[42] reflecting the importance of up-to-date literature and database review. In this case, as the mutation was identified in the literature corresponding to the phenotype, it was not necessary to further study novel genes via functional studies. Nevertheless, there are resources available as core facilities within our university for such work to be carried out, such as animal models e.g. mouse and zebrafish, as well as cell-based functional and imaging analysis.

## CHALLENGES IN APPLICATION OF WES IN RARE DISEASES AND THE FUTURE DIRECTION OF WES IN HONG KONG

There are many technical limitations in general for WES, as well as practical difficulties of applying the technology in Hong Kong for the pediatric population with rare disease. First of all, there are inherent challenges with the WES techniques. By using bioinformatics methods, a large amount of variants must be filtered out from the raw data. A normal individual has been estimated to have 50-100 mutations in the heterozygous state that can cause a recessive Mendelian disorder when being homozygous.[43] However, there is yet a consensus on the best filtering strategy. Potentially positive findings may be removed in the process as 'false negatives', which may be a result of low base coverage, poor capture efficiency of certain regions and difficulty in unambiguously aligning repetitive regions. There is also the difficulty in functional annotation of non-coding regions, hence current WES technology isn't capable of detecting changes surrounding exons.[44] Clinicians and researchers must be aware of the complexity of the methodology and potential limitations of WES, rather than viewing it as a simple laboratory test.

Another big problem is a lack of ethnic or region specific control set for determining variant frequency in Chinese patients. Large databases such as the 1000 genome project and NHLBI GO Exome Sequencing Project (ESP) only contains very few Chinese subjects, and specifically the lack of Southern Chinese, which makes up the majority of the Hong Kong population.[45] Recently, the Dutch whole-genome sequencing project constructed a haplotype map of more than 20 million variations on 250 families, which is the most robust kind of control data.[46] There is an urgent need to build a larger cohort of normal individuals that is closer to this in Hong Kong. Nevertheless, we are currently using data from many local rare disease studies as an internal database to identify frequency of variations. However, these are not normal subjects and the numbers are far from ideal.

Above all, the major bottleneck remains to be the difficulty in linking novel mutations with the disease phenotype under investigation. This involves the validation of the disease causing mutation using functional studies, such as animal experiments. This requires a thorough understanding of the pathophysiology of the disease and its underlying pathway, which can be challenging for each individual rare disease seen in the clinical genetics clinic. This is a tedious process which may take months and years to perform, and can consume a lot of research staff time and laboratory resources, causing an unrealistic turn-over-time for results be available for clinical use.

Last but not least, in Hong Kong there is still a lack of support and awareness of rare diseases in the local community and even among the medical professionals. There is an increasing demand for clinical geneticists, genetic counselors, clinical bioinformaticians as well as the necessary hardware including sequencers or facilities for quick functional validation of novel variants. The 2 medical schools in Hong Kong are actively providing post-graduate training for genetic counseling, medical genetics and bioinformatics. And hopefully with the grand opening of the Hong Kong Children's Hospital in 2018,[47] more of these personnel and facilities can be factored in and a hub for the care and management of patients with rare genetic disorders can be supported to allow provision of word-class genomic/genetic services to children in Hong Kong.

**CONFLICT OF INTEREST**

None.

**ACKNOWLEDGEMENTS**

The authors thank the SK Yee Medical Foundation and the SK Yee Medical Research Fund for supporting our research in this area.

**REFERENCES**

1. Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. What can exome sequencing do for you? J Med Genet. 2011;48(9):580-589.
2. Service RF. Gene sequencing. The race for the $1000 genome. Science. 2006;311(5767):1544-1546.
3. KA. W. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Available at: www.genome.gov/sequencingcosts.
4. Johansen Taber KA, Dickinson BD, Wilson M. The promise and challenges of next-generation genome sequencing for clinical care. JAMA Intern Med. 2014;174(2):275-280.
5. van El CG, Cornel MC, Borry P, et al. Whole-genome sequencing in health care: recommendations of the European Society of Human Genetics. Eur J Hum Genet. 2013;21(6):580-584.
6. Glass DA II, Nuara AA. Whole exome sequencing. Dermatitis. 2013;24(1):33-34.
7. Rabbani B, Tekin M, Mahdieh N. The promise of whole-exome sequencing in medical genetics. J Hum Genet. 2014;59(1):5-15.
8. Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009;461(7261):272-276.
9. Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet. 2010;42(1):30-35.
10. Bamshad MJ, Shendure JA, Valle D, et al. The Centers for Mendelian Genomics: a new large-scale initiative to identify the genes underlying rare Mendelian conditions. Am J Med Genet A. 2012;158A(7):1523-1525.
11. Pediatric Cardiac Genomics C, Gelb B, Brueckner M, et al. The Congenital Heart Disease Genetic Network Study: rationale, design, and early results. Circ Res. 2013;112(4):698-706.
12. Zaidi S, Choi M, Wakimoto H, et al. De novo mutations in histone-modifying genes in congenital heart disease. Nature. 2013;498(7453):220-223.
13. The Deciphering Developmental Disorders (DDD) study. http://www.ddduk.org/. Accessed 25/08/2014.
14. Firth HV, Wright CF, Study DDD. The Deciphering Developmental Disorders (DDD) study. Dev Med Child Neurol. 2011;53(8):702-703.
15. Mendelian Exome Sequencing Project (Mendelian Exome). https://www.nhlbi.nih.gov/research/resources/genetics-genomics/mendelian.htm. Accessed 25/08/2014.
16. Gahl WA, Markello TC, Toro C, et al. The National Institutes of Health Undiagnosed Diseases Program: insights into rare diseases. Genet Med. 2012;14(1):51-59.
17. Gahl WA, Tifft CJ. The NIH Undiagnosed Diseases Program: lessons learned. JAMA. 2011;305(18):1904-1905.
18. The UK10K project. http://www.uk10k.org/. Accessed 25/08/2014.
19. Ding LE, Burnett L, Chesher D. The impact of reporting incidental findings from exome and whole-genome sequencing: predicted frequencies based on modeling. Genet Med. 2014.
20. Rabbani B, Mahdieh N, Hosomichi K, Nakaoka H, Inoue I. Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. J Hum Genet. 2012;57(10):621-632.
21. Beaulieu CL, Majewski J, Schwartzentruber J, et al. FORGE Canada Consortium: outcomes of a 2-year national rare-disease gene-discovery project. Am J Hum Genet. 2014;94(6):809-817.
22. de Ligt J, Willemsen MH, van Bon BW, et al. Diagnostic exome sequencing in persons with severe intellectual disability. N Engl J Med. 2012;367(20):1921-1929.
23. Rauch A, Wieczorek D, Graf E, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet. 2012;380(9854):1674-1682.
24. Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med. 2013;369(16):1502-1511.
25. Kuehn BM. NIH's Undiagnosed Diseases Program expands: 6 new sites offer potential answers to more patients. JAMA. 2014;312(6):587.
26. Backenroth D, Homsy J, Murillo LR, et al. CANOES: detecting rare copy number variants from whole exome sequencing data. Nucleic Acids Res. 2014;42(12):e97.
27. Shi Y, Majewski J. FishingCNV: a graphical software package for detecting rare copy number variations in exome-sequencing data. Bioinformatics. 2013;29(11):1461-1462.
28. Mao H, Yang W, Lee PP, et al. Exome sequencing identifies novel compound heterozygous mutations of IL-10 receptor 1 in neonatal-onset Crohn's disease. Genes Immun. 2012;13(5):437-442.
29. Glocker EO, Kotlarz D, Boztug K, et al. Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. N Engl J Med. 2009;361(21):2033-2045.
30. Li M, Pang SY, Song Y, Kung MH, Ho SL, Sham PC. Whole exome sequencing identifies a novel mutation in the transglutaminase 6 gene for spinocerebellar ataxia in a Chinese family. Clin Genet. 2013;83(3):269-273.
31. Wang JL, Yang X, Xia K, et al. TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. Brain. 2010;133(Pt 12):3510-3518.
32. Lee PP, Mao H, Yang W, et al. Penicillium marneffei infection and impaired IFN-gamma immunity in humans with autosomal-dominant gain-of-phosphorylation STAT1 mutations. The Journal of allergy and clinical immunology. 2014;133(3):894-896 e895.
33. Luo R, Wong YL, Law WC, et al. BALSA: integrated secondary analysis for whole-genome and whole-exome sequencing, accelerated by GPU. PeerJ. 2014;2:e421.
34. Zhang L, Zhang J, Yang J, Ying D, Lau YL, Yang W. PriVar: a toolkit for prioritizing SNVs and indels from next-generation sequencing data. Bioinformatics. 2013;29(1):124-125.
35. Zeng S, Yang J, Chung BH, Lau YL, Yang W. EFIN: predicting the functional impact of nonsynonymous single nucleotide polymorphisms in human genome. BMC Genomics. 2014;15:455.
36. Miller DT, Adam MP, Aradhya S, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet. 2010;86(5):749-764.
37. Jamal SM, Yu JH, Chong JX, et al. Practices and policies of clinical exome sequencing providers: analysis and implications. Am J Med Genet A. 2013;161A(5):935-950.
38. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-1760.
39. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-249.
40. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073-1081.
41. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
42. Schuurs-Hoeijmakers JH, Oh EC, Vissers LE, et al. Recurrent de novo mutations in PACS1 cause defective cranial-neural-crest migration and define a recognizable intellectual-disability syndrome. Am J Hum Genet. 2012;91(6):1122-1127.
43. Abecasis GR, Altshuler D, Auton A, et al. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061-1073.
44. Kaname T, Yanagi K, Naritomi K. A commentary on the promise of

whole-exome sequencing in medical genetics. J Hum Genet. 2014;59(3):117-118.
45. Kuehn BM. 1000 Genomes Project promises closer look at variation in human genome. JAMA. 2008;300(23):2715.
46. Genome of the Netherlands C, Genome of the Netherlands C. Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat Genet. 2014;46(8):818-825.
47. Ground-breaking ceremony for Hong Kong Children's Hospital. http://www.info.gov.hk/gia/general/201402/25/P201402250508.htm. Accessed 29/08/2014.