HYWEL WILLIAMS
SENIOR RESEARCH ASSOCIATE
UCL GREAT ORMOND STREET INSTITUTE OF CHILD HEALTH

IT IS ESSENTIAL TO WORK SEAMLESSLY WITH THE CLINICAL EXPERTS WHO CARE FOR AND TREAT PATIENTS WHEN PERFORMING NGS ANALYSES IN THE CLINICAL ENVIRONMENT.

THE NEED FOR SPEED

THE RAPID PAEDIATRIC SEQUENCING (RAPS) PROJECT, UCL/GOSH IS HELPING TO DELIVER RAPID GENETIC DIAGNOSES TO CRITICALLY ILL CHILDREN ON INTENSIVE CARE UNITS IN THE UK.

Babies and young children with undiagnosed rare genetic disorders being treated on Paediatric Intensive Care Units (PICU) represent a clinical cohort that potentially has the most to gain from a rapid genetic diagnosis. In these individuals there are high rates of mortality and clinical decision making is complicated by rapid disease progression and undifferentiated phenotypic presentation. Obtaining a timely genetic diagnosis is the crucial first step in treating individuals with appropriate interventions based on empirical evidence. However, with over 8,000 known single gene disorders it is often incredibly difficult to know which individual genes are candidates and need to be tested. In this scenario we require the ability to sequence the entire human genome, rapidly, to allow an unbiased assessment of all the potential causative variants.

"WE BELIEVE THE RAPS FRAMEWORK THAT WE HAVE DEVELOPED IS ROBUST AND READY FOR ADOPTION BY OTHER GENETIC DIAGNOSTIC LABORATORIES, NOT JUST IN THE UK BUT ACROSS THE GLOBE."

Over the past decade, the application of next generation sequencing (NGS) has had a profound impact on genomic research. However, arguably the most significant application of NGS, to date, has been its use in human genomic health research - healthomics.

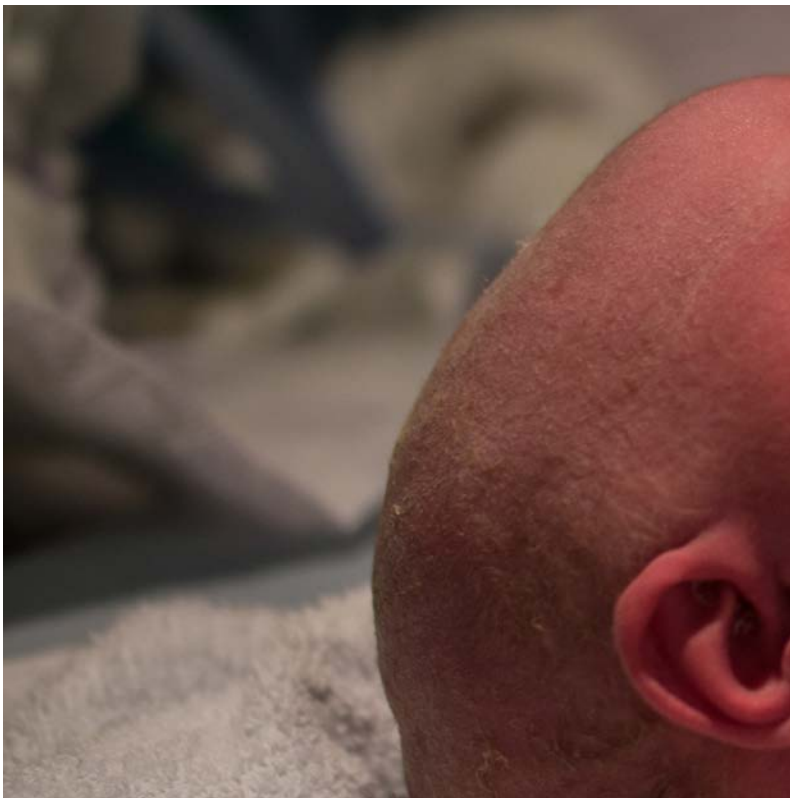
In this context, NGS has proved to be such a robust and accurate technique that it is now routinely used in genetic diagnostic laboratories across the globe. In the vast majority of cases this involves sequencing a specific set of genes, known as a gene panel, that are known to harbour mutations that can cause the disorder displayed by the individual being treated. In some rare cases the test may involve sequencing all the genes in the genome (whole exome sequencing – WES), or even the entire genome (whole genome sequencing – WGS), but these are very much the exception.

For the infants on PICU, we want to be able to rapidly perform NGS under the strict guidelines adhered to by diagnostic genetic laboratories. However, current routine genetic tests can take weeks to months to return, a timescale incompatible with our needs on PICU. These tests also only look at a specific set of genes which is not applicable in our situation, where the phenotype is often undifferentiated or masked by secondary clinical symptoms. We therefore need to overcome the constraints on time by developing a framework to allow rapid preparation of samples, sequencing, processing and interpretation of data, and we need to do this on samples that have undergone WGS, in order to provide us with an unbiased view of all variants in the genome.

When considering NGS, it is useful to split the workflow into two main components, both of which need to be optimised. The first relates to the laboratory side and consists of the physical technology of the sequencing machinery, the chemical reagents for sample preparation and sequencing, in addition to the laboratory protocols used to process the DNA to make it ready for sequencing. The second component relates to the computing infrastructure required to house and process the vast quantities of data generated, as well as the software programs written to analyse and interpret the data. Therefore, to run a NGS pipeline from start to finish 'in house' requires a multidisciplinary team of scientists and technologists. Furthermore, when performing NGS analyses in the clinical environment it is essential to work seamlessly with the clinical experts who care for and treat the patients, to further ensure the results are clear and compatible with the standard diagnostic reports they are used to working with.

This is the position we were in when we first devised our Rapid Paediatric Sequencing (RaPS) project. Previous work by Stephen Kingsmore's group[1,2] had shown the feasibility of performing rapid WGS in the PICU environment, but the protocols and equipment used in those studies were not compatible with the 'real life' working conditions in a standard genetic diagnostic laboratory. For example, they processed their samples in an around-the-clock 24 hour timeframe, and used a sequencer that had been modified by the manufacturer to reduce the time taken to sequence the DNA. →

"IN THE FUTURE, AS THE COST OF WGS CONTINUES TO FALL, IT IS LIKELY THAT ALL INDIVIDUALS WITH A SUSPECTED GENETIC DISORDER WILL UNDERGO WGS. TO THIS END, THE GENOMICS ENGLAND, 100,000 GENOMES PROJECT IS BEGINNING TO TRANSFORM THE GENOMIC CAPABILITIES OF THE NHS, AND IN THE FUTURE WGS WILL UNDOUBTEDLY BECOME A ROUTINE DIAGNOSTIC TEST. THIS VIEW IS REFLECTED IN THE RECENT COMMITMENT OF THE UK GOVERNMENT TO RECOGNISE TRANSFORMATIVE POWER OF NGS IN ITS LIFE SCIENCES SECTOR DEAL AND IN THE CHIEF MEDICAL OFFICER'S GENERATION GENOME ANNUAL REPORT."



Nonetheless, such work demonstrated the clinical utility of a rapid sequencing approach, and so to further that work we sought to develop a fully comprehensive framework that details the steps necessary to run RaPS under the strict conditions adhered to by National Health Service (NHS) hospitals in the UK. We systematically optimised our framework from the very beginning, when patients are recruited and consented, through to the preparation of the DNA for sequencing, the sequencing of samples, as well as the processing of sequence data, the interpretation of the results and the generation of a diagnostic report to disseminate the most likely causative genes back to the referring clinical team. We ensure this data is freely available so that any diagnostic group wishing to replicate our work has a set of protocols to follow.

Setting up our protocol was an iterative process that would not have been possible without the buy-in from the multiple partners who comprised the RaPS team. This really is the first and most essential part of the protocol. The next vital part of this project was the funding, which was considerable. We are fortunate enough that our funding is derived from a National Institute Health Research (NIHR), Biomedical Research Centre (BRC) grant awarded jointly to UCL Great Ormond Street Institute of Child Health (GOS ICH) and Great Ormond Street Hospital (GOSH), that is focused on the translation of academic research into the clinical realm to improve patient benefit. Without the translational focus of our BRC funding it would have been very difficult to complete a project like this within the UK.

Once we had our multidisciplinary team together, we then divided the protocol into four major stages and worked hard to optimise each individual stage as well as the transition phase between each one. It is the transition between the stages that represents a transfer point between the specialities, and thus is a point where errors can arise.

The first stage in the final version of the framework is patient recruitment, which involves the close interaction of clinical genetics personnel and PICU clinicians. We have developed strict inclusion/exclusion criteria to select the most appropriate individuals, and it is at this point that the individual being treated and their parents (all analyses comprise parent-proband trios) are consented and a sample of blood is taken.

The next stage involves the transition of the blood sample to the genetic diagnostic laboratory where the DNA is rapidly extracted and passed on to the laboratory team who prepare the sample for sequencing. Following a final quality check, the samples are run on two sequencing machines. The individual being studied is run on their own to generate an average of 30-times coverage across the genome, and their parents are combined together and run at the same time to give 15-times genome coverage each. The time to sequence the samples is around 27 hours.

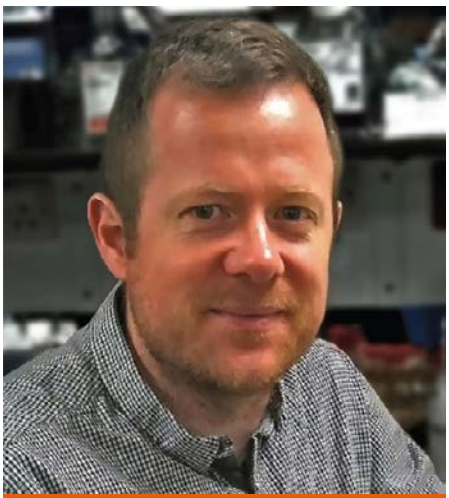
Upon completion of the sequencing there is a large amount of data that needs to be processed before we can begin to look for causative variants. The first step in this process involves mapping the sequencing reads back to the human genome reference. Once we know where all of the reads map to, we can begin identifying the regions of the genome where there are differences between the individual that we are studying and the human genome reference. Typically, it can take up to 48 hours to process a single WGS sample using open source software (BWA-mem/GATK), even utilising the computing power available to us at UCL. This could add up to six days to our protocol and is therefore not compatible with our rapid timeframe. To overcome this problem we contacted the company, GENALICE who provide a hardware/software solution for the processing of genomic data. We incorporated one of their servers into our cluster which allowed us to drastically reduce the sample processing time, down to an average of 20 minutes per sample,

meaning we could process a full trio in just one hour. Work performed by us 'in house' and by others[3] has shown that this time reduction does not come at the price of accuracy or sensitivity.

The next stage in our framework involves the analysis of the WGS variant data. A fundamental concept to our approach is that we are studying very rare and severe genetic disorders. We know the human genomic reference will not contain such damaging variants, and thus any difference between the sequence of the individual we are studying and the human reference is a potential causative variant. This process also drastically reduces the size of the data set we need to analyse, as instead of looking at all 3.5 billion bases of DNA in the human genome we will typically expect to see just 3.5 million bases that differ between the human genome reference and our individual. Furthermore, by using a filtering strategy to focus on just the regions of DNA that code for proteins, we can reduce this number down to around 50,000 variants per individual. More specific filters can then be applied that use information contained within publically available databases, to ensure that we only look at very rare variants that are predicted to have a damaging effect on the gene in which they reside. This brings the number of variants we need to look at in a typical individual to less than 50.

Once we have a list of potentially damaging variants we then introduce clinical information about the individual being studied through the use of Human Phenotype Ontology terms. This information is used in conjunction with publically available databases, such as Genomics England PanelApp to derive a list of candidate genes to prioritise our analyses. If no potentially damaging variants are identified in this list of candidate genes, we then expand the list to include all genes known to cause genetic disorders and finally, if no variants are identified in this analysis we later expand our list of genes to include every gene in the genome. By structuring our analysis into three tiers we can prioritise our analyses on the most likely candidate genes first which drastically decreases the time taken to reach a diagnosis.

The evaluation of variants in candidate genes requires a close working relationship between our genomic analysts, clinical geneticists and clinical scientists. Any likely causative variants are discussed in multidisciplinary team meetings to arrive at a consensus diagnosis that can be fed back to the referring clinical team to help inform their clinical management. →



Hywel began his research career at the MRC Centre for Neuropsychiatric Genetics and Genomics at Cardiff University, working on the genetics of schizophrenia. In 2013, he moved to UCL and joined the GOSgene group and soon began leading it. During his four years doing so, he set up 66 projects and sequenced over 500 exomes, as well as over 250 whole genomes. In 2016, Hywel began working fulltime on setting up the RaPS project and bringing together a multidisciplinary team.

Figure 1. Major transitional stages of RaPS framework.

If the identified variant affects a gene that may allow for an immediate change in clinical management we will issue a preliminary report to the referring clinical team, to help in their clinical management. Finally, for each individual we issue a diagnostic report detailing the clinical presentation, the sequencing metrics, analysis pipeline, genes studied and the results of the study, irrespective of whether a genetic diagnosis is reached or not.

We have now used this framework for the analysis of 24 individuals from PICU who have a range of phenotypes covering the spectrum of clinical specialities. We have been able to make a clinical diagnosis in ten of these individuals (42%), and in three of the individuals the identification of a causative variant had an immediate impact on the individual's clinical management. The shortest time taken to reach a diagnosis in this study was just five days.

We believe the RaPS framework that we have developed is robust and ready for adoption by other genetic diagnostic laboratories, not just in the UK but across the globe. The work we present is the result of an iterative process, and so the time taken to reach a diagnosis has varied from over six weeks, to as little as five days, but as our practices have matured we are now regularly able to obtain a result within a two week timeframe. In some cases this may be a genetic diagnosis, but in other cases we may identify a novel gene that requires further functional validation in the research environment, and in a number of instances we are unable to identify any causative variants.

The barriers to obtaining a rapid diagnosis were initially centred on getting the different specialities together and optimising the transition steps between them. Gaining access to sequencing machines for a research project, when collaborating with a clinical diagnostic laboratory also raised obvious limitations as we had to fit in around the essential diagnostic sequencing runs.

Our RaPS framework demonstrates the feasibility of performing a rapid genetic diagnosis under the strict parameters adhered to by UK NHS genetic diagnostic laboratories. Although the cost of a RaPS test (~£5,600/trio) looks relatively high, it must be viewed in the context of a bed on PICU which is estimated to cost £4,500 a day. Therefore, if a rapid diagnosis brings about improved prognosis and allows an individual to move to a less high dependency environment, it begins to pay for itself. A full health economics study would be beneficial to extrapolate the health benefits of a RaPS test but until then it is important to carefully select those who will benefit most. Given the costs involved in managing critically ill individuals, a rapid genetic diagnosis in this group may currently be the most cost effective for NHS and healthcare providers.

In the future, as the cost of WGS continues to fall, it is likely that all individuals with a suspected genetic disorder will undergo WGS. To this end, the Genomics England, 100,000 Genomes Project is beginning to transform the genomic capabilities of the NHS, and in the future WGS will undoubtedly become a routine diagnostic test. This view is reflected in the recent commitment of the UK government to recognise transformative power of NGS in its Life Sciences Sector Deal and in the Chief Medical Officer's Generation Genome Annual Report.

Our paper describing the full RaPS framework in detail is currently under review. ■

References:

1. Miller NA, Farrow EG, Gibson M, Willig LK, Twist G, Yoo B, et al. A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. Genome medicine. 2015;7(1):100. Epub 2015/10/01.
2. Saunders CJ, Miller NA, Soden SE, Dinwiddie DL, Noll A, Alnadi NA, et al. Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. Science translational medicine. 2012;4(154):154ra35. Epub 2012/10/05.
3. Pluss M, Kopps AM, Keller I, Meienberg J, Caspar SM, Dubacher N, et al. Need for speed in accurate whole-genome data analysis: GENALICE MAP challenges BWA/GATK more than PEMapper/PECaller and Isaac. Proc Natl Acad Sci U S A. 2017. Epub 2017/09/17.