Interview

Personal account of the discovery of a new disease using next-generation sequencing

Gholson J Lyon speaks to Natalie Harrison, Commissioning Editor

Gholson J Lyon

Utah Foundation for Biomedical Research, Salt Lake City, UT, USA
and
Center for Applied Genomics, Children's Hospital of Philadelphia, PA, USA
and
NYU Child Study Center, NY, USA
GholsonJLyon@gmail.com

Gholson J Lyon, MD, PhD, is a research scientist at the Center for Applied Genomics at The Children's Hospital of Philadelphia and at the Utah Foundation for Biomedical Research (UFBR). He is also an adjunct assistant professor of Psychiatry at New York University School of Medicine. He is a board-certified child, adolescent and adult psychiatrist. He earned an MPhil in Genetics, working with Nobel laureate Martin Evans in Cambridge, UK, then received a PhD and MD through the Cornell/Rockefeller/Sloan-Kettering Tri-Institutional MD-PhD Program. He has published peer-reviewed papers in biochemistry, genetics, pharmacology and molecular biology. In addition to his research interest in neuropsychiatric illnesses, Dr Lyon is focusing on the discovery of rare previously undiscovered Mendelian diseases. Recently Dr Lyon and a team of international researchers used a new disease-variant finding tool (variant annotation, analysis and search tool [VAAST]) to identify the cause of an extremely rare X-linked genetic disorder that is lethal in infancy. The disease-causing mutation is in the *NAA10* gene. Affecting only males, the mutation causes an aged appearance, facial abnormalities, developmental delay and cardiac arrhythmias, among other conditions. The team studied a family in Utah with a history of several boys with these symptoms who died in infancy, and analyzed DNA from five boys in the family. They are tentatively calling the disease Ogden syndrome, reflecting the family's city of residence. This study is one of the first times in which next-generation sequencing has been used to help identify and confirm the genetic etiology of a previously unknown and unreported disease. Dr Lyon answers some questions below regarding this work.

■ Can you describe your recent work?

This is a personal account of the recent discovery of a new disease that had not been previously described (at least to my knowledge) [1]. There were many colleagues involved in this work, so this is simply one version of the discovery, as seen through my eyes. This project was quite multidisciplinary, drawing upon the expertise of clinical geneticists (Alan Rope, John Carey, John Opitz, Lynne Bird, Cathy Stevens and Sarah South), pathologists (Theodore Pysher and Steven Chin), bioinformaticians (Kai Wang, Evan Johnson, Chad Huff, Barry Moore, Jinchuan Xing and Mark Yandell) and biologists (Rune Evjenth, Johan Lillehaug and Thomas Arnesen), just to mention a few of the contributions.

My colleagues at the University of Utah (UT, USA) and Omicia, Inc. (CA, USA), a privately held company developing tools to interpret personal genome sequences, announced in *Genome Research* on 23 June 2011 a new software tool called variant annotation, analysis and search tool (VAAST), a probabilistic disease-causing mutation finder for individual human genomes [2]. A data interpretation bottleneck has limited the utility of personal genome information for medical diagnosis and preventive care. VAAST is a new algorithm to assist in overcoming this bottleneck. VAAST is the product of a collaboration between Mark Yandell, PhD, associate professor of human genetics at the University of Utah School of Medicine, and colleagues, and the Omicia scientific team under the leadership of Martin Reese, PhD, the company's CEO and Chief Scientific Officer. In their paper, Yandell and colleagues demonstrate that VAAST provides a highly accurate, statistically robust means of rapidly searching personal genomes for genes with disease-causing mutations. The authors demonstrate that as few as three genomes from unrelated children, or those of the parents

part of Future Medicine fsg

and their two children, are sufficient to identify disease-causing mutations.

In a separate paper that we arranged for publication on the same day in the *American Journal of Human Genetics*, we reported the use of VAAST and other methods to identify the mutation responsible for a newly discovered childhood disease. This new illness is characterized by aged appearance, craniofacial abnormalities, cardiac arrhythmias and other symptoms. We used X-chromosome exon capture, next-generation sequencing and the VAAST tool to identify the disease-causing mutation in *NAA10* that has resulted in this fatal disease in children of two unrelated families. We are now applying VAAST to many other unknown conditions, including rare Mendelian disorders and other common disorders such as attention deficit hyperactivity disorder (ADHD) and autism.

■ Why is this work important & what is the significance of it?

Harnessing the new generation of rapid, highly accurate gene-sequencing techniques, our research team has identified the disease-causing mutation in a newly characterized rare genetic disease by analyzing DNA from just a few individuals. The power and speed of the innovative bioinformatics tool (VAAST) marks a step toward personalized genomics—discovering causative mutations in individual patients. Our research is proof-of-principle that VAAST can identify disease-causing mutations with greater accuracy, using DNA from far fewer individuals, more rapidly, than was previously possible. Although several of the existing software tools for analysis of personal genome sequences have been demonstrated to be sufficiently powered to identify mutations underlying previously known disorders, the current report is one of the first times a personal genome analysis tool has identified a previously unknown syndrome.

Sydney Brenner famously stated the following (paraphrase): "What I'm advocating is: go the other way. Let's go from bedside to bench. We don't have to look for a model organism anymore, because we are the model organisms". I agree with Dr Brenner 100%. Instead of compartmentalizing research and medicine, the two should be integrated so that physicians who are most familiar with human 'phenotypes' can inform the other arms of science. The advent of next-generation sequencing has made it increasingly cheaper, faster and more feasible to sequence entire genomes or at least parts of the genome. This is leading to many new discoveries in genetic medicine, all of which is aided enormously by the initial sequencing of the human genome that was mostly completed approximately 10 years ago. The advent of exon capture and sequencing has made it possible to capture certain exons for sequencing and/or all of the known exons (referred to as the exome) for sequencing [3]. This enables the sequencing of unusual clinical presentations which may have not been amenable to standard linkage analysis in the past, particularly if the clinical cases stem from just one family. Examples include the 2010 publications on Miller syndrome [4,5], and similar studies aimed at identifying the unknown genetic defects responsible for some early childhood diseases [6].

One way to make headway in the genetics of many complex disorders is to search for large pedigrees living in the same geographic location, where one can study the penetrance and segregation of variants in a similar environmental background, free of population stratification concerns, particularly given the possible penetrance of only approximately 40–60% for mutations in some disorders [7]. With the above in mind, 2 years ago, I moved to the state of Utah, as I realized that studying genetic diseases in pedigrees will once again be quite fruitful going forward, with all the new technologies. While I started some of this work at the University of Utah, I have completed the work at the nonprofit Utah Foundation for Biomedical Research, which was founded in 2010 to study many more pedigrees in Utah going forward [101].

While we were preparing and submitting the manuscript, a second research group at the National Human Genome Research Institute notified us that they too had identified the same *NAA10* mutation in a second family with three boys who had similar symptoms to those found in the Utah family. Further analysis demonstrated that the two families were unrelated – indicating that the disorder is a syndrome and not an isolated condition found only in one family. Although the detailed biological mechanisms remain to be investigated, the mutation in this disease alters an enzyme

involved in a process called N-terminal acetylation, in which one end of a protein is modified by the addition of a chemical called an acetyl group. N-terminal acetylation occurs in approximately 80% of human proteins [8]; however, abnormalities in this specific protein modification have not previously been demonstrated to give rise to a human disorder. In this case, disrupting N-terminal acetylation results in symptoms ultimately causing death in infancy.

■ What challenges did you overcome?

To my knowledge, we were the first group to perform exon capture and sequencing in the core facility at the University of Utah, which was both challenging and exciting. We were aided enormously in this by Brian Dalley, who is an author on the paper.

Collecting pedigrees in Utah is certainly made easier by the fact that the families tend to be larger and many people tend to remain in Utah. However, one challenge is that there is unfortunately no state-wide or even university-wide biobank, and each individual investigator must collect and store their own samples. In this instance, I drove to the family's home in Ogden to meet the family and draw the blood on eight or so of the family members, and I obtained the rest of the DNA samples later on. Another challenge is that despite the fact that our medical system generates enormous amounts of clinical information on each patient, much of this information (rightfully so) is protected by privacy concerns. For one of the more recently deceased boys, a single-sided copy of the medical record is about one foot high! In the end, the easiest way for me to obtain the medical record was to get the family to obtain their own record, and the family graciously gave a copy of the record to me. In addition, several of the clinicians, including Alan Rope and Lynne Bird, summarized much of the clinical phenotype for the paper, which was quite helpful.

Another challenge was that after we had completed the sequencing of five samples from the family, I decided that I wanted to prove further that the mutation was the cause of the disease. In order to do this, I wanted to test the DNA from the other three already deceased children in the family. However, no one had collected their DNA for research. Luckily, at my request, Theodore Pysher, a pathologist and author on the paper, was able, with substantial persistence, to find two paraffin-embedded tissue blocks from the autopsies of two of the boys, saved in one case for approximately 30 years. Once again, owing to privacy concerns, the family had to sign written authorization to release these blocks to me. Another author on the paper, Jeffrey Swensen, was able to isolate usable DNA from these blocks to check whether the mutation was present or not. It was present, thus providing further proof that this is the causative mutation.

■ What are your next steps going to be?

I am collaborating actively with Thomas Arnesen in Norway on trying to understand how the defect in acetylation by the enzyme can lead to the pathophysiology of the disease, including the aged appearance, large hearts and eventual arrhythmias leading to death. I am also beginning to screen for drugs which might be able to restore activity of the enzyme, both to learn about how the enzyme functions and also perhaps to help any future boys born with this illness. However, it has taken over 20 years since the discovery of the genetic basis of cystic fibrosis for any drugs to be developed, and these drugs are still not US FDA approved to my knowledge [9]. So, it is a very long process from gene discovery to any drug development. More immediately, we have almost finished developing a Clinical Laboratory Improvement Amendments (CLIA)-certified laboratory test to screen for mutations in this gene, so we can tell the women in this and other families if they are carriers or not. Preimplantation genetic diagnosis and implantation could be used to allow female carriers to have normal children free of this disease, and it has been quite gratifying to discuss these possible options with the family in Utah.

I am also expanding my studies of pedigrees in Utah, and I work actively at the Utah Foundation for Biomedical Research (UFBR), which was founded last year by Reid Robison and Scott Duvall, partly to enable collection of pedigrees for research. We are planning to perform whole-genome sequencing of unusual cases going forward, as the costs of sequencing continue to drop dramatically.

■ What are the wider implications of the VAAST algorithm?

In retrospect, the VAAST algorithm could have identified the causative mutation using data from just two individuals – an affected boy in one family and a mother (who was a carrier and not affected) in the unrelated family. This demonstrates that VAAST can potentially identify disease-causing mutations based on DNA from only two unrelated individuals, with the caveat that we had restricted our search space to the X-chromosome under the assumption that this was likely to be an X-linked disease. Based on this proof of principle, we hope that VAAST will accelerate the discovery of disease-causing mutations in both common, complex disorders such as ADHD and autism, and in rare Mendelian disorders. This will likely be true as more and more 'normal' genomes are sequenced and added to the background genome file utilized by VAAST.

■ What are the regulatory considerations for this type of work?

This research had to be approved by an institutional review board as it involved human subjects. As I was employed at the time at the University of Utah, I was added to the Institutional review board protocol led by David Viskochil and approved for this use at the University of Utah.

This was considered as 'research' during the discovery of the mutation. However, there is substantial debate in the medical genetics and ethics communities concerning whether research results can be returned to participants or not. There really is no major or 'correct' consensus on this. Some people argue that all research results must be confirmed in a CLIA-certified laboratory prior to giving any such results back to patients. The Centers for Medicare and Medicaid Services (CMS) regulates all laboratory testing – except research – performed on humans in the USA through the CLIA. In total, the CLIA covers approximately 200,000 laboratory entities. The Division of Laboratory Services, within the Survey and Certification Group, under the Center for Medicaid and State Operations has the responsibility of implementing the CLIA Program. The objective of the CLIA program is to ensure quality laboratory testing. Given this, we have had to develop a CLIA-certified lab test at ARUP laboratories, in order to deliver any results back to the family. Obviously, the women in the family who already had affected children knew that they were carriers of the mutation; however, there are women in the family that have not had children (yet) and they would like to know whether they are carriers of the mutation or not. We are working actively to get the test developed, so that we can return the results to these women in the family (along with making the test available to any other families, including Family 2 reported in our paper). To summarize the procedure that we have had to undergo; any clinical test at ARUP first has to go through a validation process by the R&D group; known samples are run through the test in a specific way, then the results are documented in a formal validation packet that is approved by management. All results are reviewed by medical directors. After the validation, accounting has to decide how the test will be billed, genetic counselors have to write the information that will appear on the clinical reports and the test has to be formally transferred into the clinical laboratory (which involves putting paperwork in place, getting reagents in stock and training employees). In addition, the test usually has to be integrated into the computer system and ARUP test directory, although some of this can be bypassed if one requires that the test be specially ordered. Jeffrey Swensen and colleagues at ARUP have accomplished this for the specific mutation in *NAA10*, so we are now ready to begin performing the testing.

■ What future technology development would you like to see that could improve this disease-variant finding tool?

Mark Yandell's group is working to incorporate into VAAST the ability to prioritize small deletions and duplications of DNA in the genome. Right now, VAAST mainly handles single nucleotide variants, which are single base pair changes in the 6 billion bp known to exist in a diploid human genome. It really is like finding a needle in a haystack, and VAAST was developed to help find these needles.

I expect that the price of sequencing a whole genome will be US $1000 or less in the next 2–5 years. Once this occurs,

it will become cost effective to sequence entire genomes, rather than the targeted capture and sequencing that we performed for this study. I am particularly impressed by the whole genomes currently being sequenced and provided by the company, Complete Genomics (CA, USA), although the sequencing by Ion Torrent (USA) and Illumina (CA, USA) is also quite promising.

■ Do you have any concluding remarks?

We are entering an exciting period in which we will discover and explain many idiopathic diseases, and I expect that we will soon discover many more mutations involved in autism, schizophrenia, Tourette's syndrome, bipolar disorder, ADHD and obsessive-compulsive disorder. It is too early to conclude what portion of neuropsychiatric disorders will end up being readily explained by single changes in the genome, and whether they will be point mutations, deletions, duplications or other forms of variation consistent with an oligogenic explanation. However, recent research has confirmed that 5% of autism can be explained by various large copy number variants [10,11], so it is likely that some additional portion of the heritability will be explained by smaller mutations just not currently detected. However, proving causality will be the major issue, so having access to research subjects and derived tissues will be critically important, hence the pressing need for clinician scientists to advance these efforts.

Financial & competing interests disclosure

The author has no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Bibliography

1 Rope AF, Wang K, Evjenth R *et al.* Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to N-terminal acetyltransferase deficiency. *Am. J. Hum. Genet.* 89, 28–43 (2011).

2 Yandell M, Huff C, Hu H *et al.* A probabilistic disease-gene finder for personal genomes. *Genome. Res.* 21(9), 1529–1542 (2011).

3 Ng SB, Turner EH, Robertson PD *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461, 272–276 (2009).

4 Ng SB, Buckingham KJ, Lee C *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet*. 42, 30–35 (2010).

5 Roach JC, Glusman G, Smit AF *et al.* Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 328, 636–639 (2010).

6 Worthey EA, Mayer AN, Syverson GD *et al.* Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet. Med.* 13, 255–262 (2011).

7 Mitchell KJ, Porteous DJ. Rethinking the genetic architecture of schizophrenia. *Psychol. Med.* 41, 19–32 (2011).

8 Arnesen T. Towards a functional understanding of protein N-terminal acetylation. *PLoS Biol.* 9, e1001074 (2011).

9 Accurso FJ, Rowe SM, Clancy JP *et al.* Effect of VX-770 in persons with cystic fibrosis and the G551D-*CFTR* mutation. *N. Engl. J. Med.* 363, 1991–2003 (2010).

10 Sanders SJ, Ercan-Sencicek AG, Hus V *et al.* Multiple recurrent *de novo* CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 70, 863–885 (2011).

11 Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D. Rare *de novo* variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron* 70, 898–907 (2011).

■ Website

101 Utah Foundation for Biomedical Research www.utahresearch.org