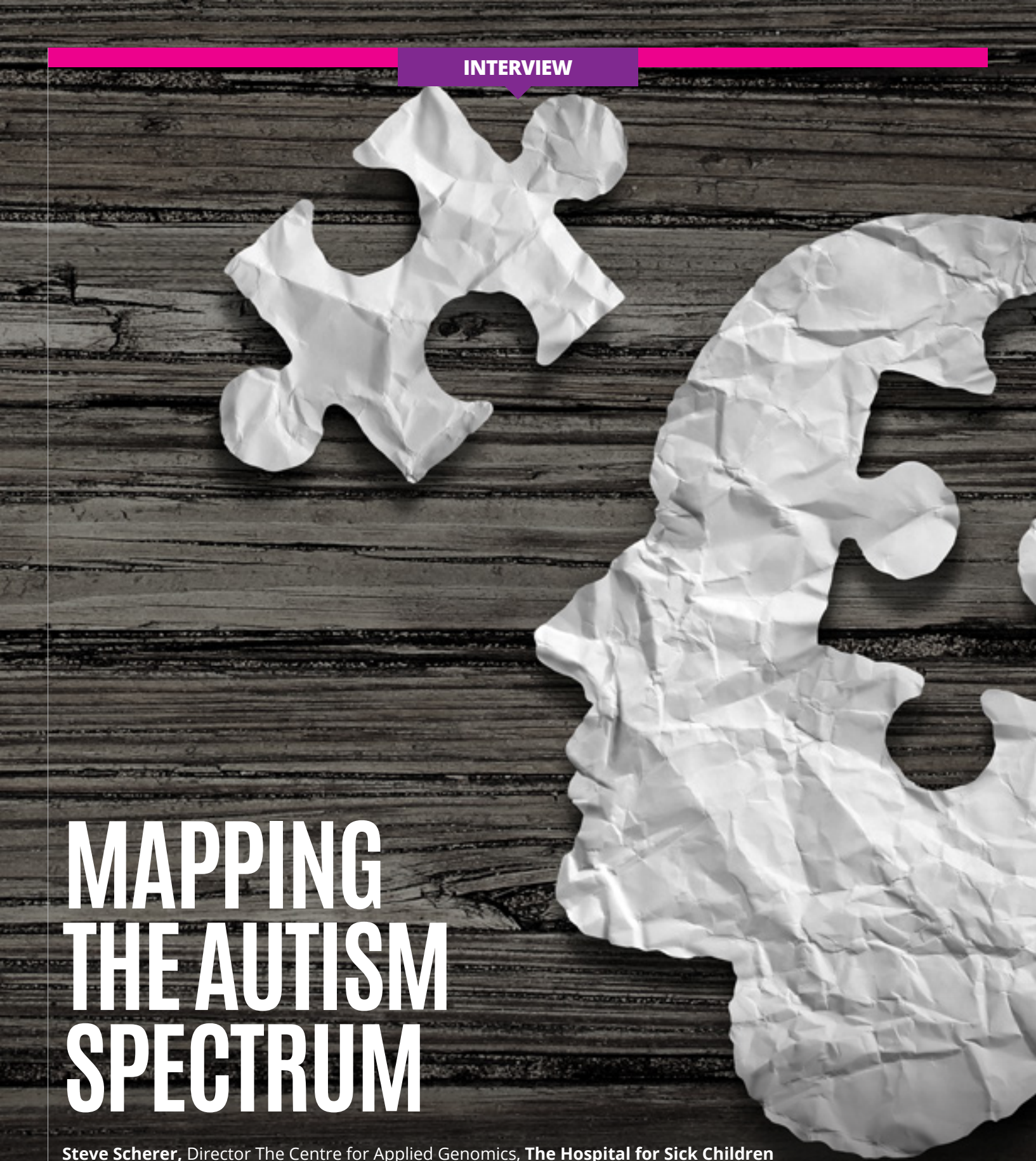


MAPPING THE AUTISM SPECTRUM

Steve Scherer, Director The Centre for Applied Genomics, **The Hospital for Sick Children**

A HIGHLY HETEROGENEOUS CONDITION, AUTISM PRESENTS A UNIQUE CHALLENGE TO MEDICAL GENETICISTS. HEADING THE CHARGE TO UNDERSTAND THE GENETICS OF AUTISM, AND MAKING IT EASIER TO GET PUBLISHED, IS DR STEVE SCHERER.

Awareness of autism as a genetic, or partially genetic, condition has been on the rise over the past decade. According to Steve Scherer, Director of the Center for Applied Genomics at the Hospital for Sick Children (SickKids) in Toronto, he was in the right place at the right time (and speaking to the right people!) to be able to start piecing together some of the first genomic mapping for the condition. Most recently his group succeeded in defining the copy number variations and other genetic features underlying autism, and are now able to use that information to provide much-needed assistance and genetic counselling to affected families.

FLG: How did you become involved in autism research?

SS: I worked in the early days of mapping chromosomes. Our group in Canada contributed to the mapping of chromosome seven as part of the Genome Project. Back in the 90's, when I started my own laboratory here at SickKids in Toronto, a very interesting paper came out of Tony Monaco's group, who was in Oxford at the time. The paper showed the first linkage in autism to chromosome seven, which was the chromosome we were the experts on. Interestingly, and this is the beauty of the story, the same day that paper came out I got a fax from a family in California that had a son that was autistic. He had a chromosome translocation that intersected chromosome seven in the same region the Oxford group had published their linkage to.

I knew what autism was, but I didn't now a lot about it. From our paediatric hospital, I found out that we were seeing hundreds of kids with the condition. As a result of this serendipitous communication, and being at the right place at the right time, we started the autism programme here. We moved very quickly. We assembled a lot of genomic data from our existing work on chromosome mapping, and my group was already working on new technology, so again it was somewhat of a perfect storm for us.

We were the first in Canada to get the one megabase CGH (comparative genomic hybridization) microarray. Our very first samples were actually samples with autism. When we first described the phenomena of copy number variation, we were running samples from autistic individuals and their parental controls. It was a mixture of foresight, but also building on being in the right place at the right time.

FLG: What makes autism such a challenge to understand?

SS: Autism is exceptionally clinically heterogeneous and, as we now know, it's exceptionally genetically heterogeneous. People hear the term 'autism', or 'autisms', or 'autism spectrum disorders', and they often think of it as a single condition. Psychiatrists and physicians were trained for fifty years to think of it as a behavioural condition, so people have been thinking about it as a single entity for a long time now. However, there are over a hundred different disorders, with different genetic names like Rett syndrome, or fragile X syndrome, that can have autism as a primary component of their clinical diagnosis. That in itself indicates that autism is heterogeneous. But the name comes down to where the child first showed up. If they are labelled with autism, or if they come through a clinical genetics route their condition may get the name of the microdeletion syndrome.

From the research side, everyone was focusing on the so-called idiopathic autism - the autism that was unexplained genetically. Now it seems that we've come full circle and acknowledging that there is a big mix. About 60-80% of what is reported as autism is yet to be defined. The field is now coming to the point where we don't just focus on the idiopathic autism that has very, very stringent features of the behavioural classification. We're starting to study anybody who has an autism-like clinical presentation. It's been amazing. Around ten years ago we didn't have any of the genes identified that accounted for these idiopathic cases, and now we do. This is because of the microarrays and then the sequencing. The whole field is now completely empowered with the progress from the last decade, and it's starting to chip away at the heterogeneity. That's what our sequencing project, the MSSNG Project (with Autism Speaks and Google), is doing. We're using whole genome sequencing to subgroup the autisms into their genetic contributors.

FLG: Your approach was a little different to what many were doing at the time. What lead to you taking that different kind of approach to what other people were doing?

SS: The beauty of a genomic approach is that it is hypothesis free, and you get a lot of data. To be able to process these large amounts of data you tend to have to use algorithms to smooth it over, and cherry pick the 'low-hanging fruit'. But when we're looking at undiscovered phenomena in genetics, often the answers are in the data that is much harder to interpret.

With copy number variation, we were comparing genome sequence assemblies before anyone else because we were one of only a few groups that had access to the public data and the private data through Celera. Craig Venter, Mark Adams, and Richard Mural of Celera were very helpful during this time. We published a paper in *Science* in 2003 and another in *Nature Genetics* in 2006 where we reported aberrations in genome sequences at the copy number and structural gene level. Well, if you remember back around 2000 there were a lot of papers being published criticising Celera versus the public draft sequences, and the public draft versus Celera's work. Well it turns out that many of the differences people were quick to point out, we later reported as being copy number variables. So rather than mistakes, or errors in quality, what researchers were actually getting angry about was natural biological variation.

It's insight we had based on our decade-long scrutiny of chromosome seven. A lot of people were seeing what we now understand to be copy number variation, but they were throwing the data away thinking it was just noise. We eventually recognised it for what it was, and it was just the philosophy of looking harder at 'failed' experiments and explaining the data. It's something I still teach my students today – if you're doing cutting edge science most of your experiments should be failing, but you need to interpret what that failing means. Often it's those 'failures' that lead to the big discoveries that no one else has seen because they've just thrown that data away.

FLG: It must have been an exciting thing to realise what you'd found and knowing that you had the data to back it up. →

"WE HOPE TO GET THOSE PAPERS THAT EVERY ONCE IN A WHILE MIGHT TURN OUT TO BE A GEM IN THE FUTURE."

SS: It was really challenging. The major message from the genome project was this 99.9% identity. We knew that if you looked under the microscope for cytogenetic changes, about 0.6% of the population would have big chromosome changes to the normal population, and that the genome data was accurate for the single nucleotide changes. But if you actually plot the variants between the two extremes, systems equilibrium would predict that you should see an equal distribution of classes of size variants across the genome. The reason we didn't see that is we just didn't have the technologies to resolve that. That meant the "dimensional" differences in copy number variation were never catalogued. Now when you look back at some of those early genomes, like the Venter de novo whole genome assembly published in 2007, which Craig and Sam Levy included us in, with today's technology you do actually see that equal of distribution of different-sized variants across different sizes. So once we saw that first microarray data, it was obvious to me that most of it would be real data. It just fit, because it's something that should have been there. I think most people just didn't think it through.

We ended up co-publishing with Charles Lee, a Canadian in the United States at the time, who was seeing the same thing using the same microarray data as us. Everyone else was telling us, 'Your data's crap because we would have seen this before.' There was a lot of push back from the community. It was so hard for us to get our findings published. It took forever to get through review and I don't think Charles or I could have done it alone. In retrospect, my greatest pride was the persistence we had in convincing the community. The paper was published just before the 2004 ASHG meeting coincidentally in Toronto, and it was all everyone was talking about there. The cytogeneticists, and more traditional geneticists, completely got it and believed it. The sequencing crowd didn't want to believe it because they didn't find it. It was at that meeting that those guys were convinced, not by us but by the cytogeneticists. But there was a year and half, almost two year period, where we had the data and it was very hard to get people to believe it.

FLG: Would you classify that as a strength or a weakness of the system? It held you up considerably, but did also force you to present a very compelling argument.

SS: I was funded by Canadian agencies. They were already bought into it, and I had enough of a reputation that I was very well funded. I got two big grants written in 2004 based on the preliminary data ahead of the publication of that first paper. Charles Lee, on the other hand, was a new investigator and NIH funded and got killed with this. I don't think he even received NIH funding for human CNV work for a couple of years. It seemed that people there were less likely to believe it because they had other vested interests.

I was fortunate, and it essentially gave me two years of lead-time on the rest of the field. But it would have been better for the field if there hadn't been so much resistance. It is frustrating that the same people who really discredited our ideas have come back saying they knew it all along. Charles and I had a few people come up to us at that 2004 meeting saying 'I memorised your paper. We need to have more of this; it's explaining a lot of our unexplained phenomena.' That was a great compliment and really kept us going.

FLG: From your standpoint and the research that you do, what are the new technologies that are really driving a greater understanding of autism?

SS: Whole Genome Sequencing is just spectacular, it's like being back in the early 2000's. It gives us a good viewpoint at a decent

Steve Scherer
Director, The Centre for Applied Genomics
The Hospital for Sick Children

Known for contributions to discovering the phenomena of global copy number variation (CNVs) of DNA and genes as the most abundant type of genetic variation in the human genome, Dr. Scherer leads one of Canada's busiest laboratories. His group has discovered numerous disease susceptibility genes and most recently has defined CNV and other genetic factors underlying autism. He collaborated with Craig Venter's team to decode human chromosome 7 and to generate the first genome sequence of an individual.

cost across the whole genome. However, much of the community is focussing on calling single nucleotide variants, and small insertions or deletions. That's what the established pipelines do, and they do that very well. But for autism we know that the larger changes are critical. We've shown in our newest paper in press that 50% of interpretable mutations in autism are actually CNV's or large indels. You have to pay attention to them, but the high throughput centres are now using standardised pipelines that do not do well at calling these things.

Our group is doing a lot of what we did in the microarray days; establishing approaches that use whole genome sequence data to call structural variants and CNV's. We've published on some of this already. The cancer world is already there, and it's something that is being rediscovered by the big sequencing centres. We're going to see a lot of papers over the next few years looking beyond single nucleotide and small sequence level variants. I think a lot of answers are going to come from looking deeper into these structural variants, some being somatic in origin.

It's tougher science, as you might imagine. The false discovery rate is higher, but we're finding it's definitely worth it to go back and use some monkey grease or look at the data with your eyes and develop some new approaches and the clinical yield goes up. That's exciting, maybe because I feel like I can still help.

FLG: Many of the same networks in the brain seem to be affected in several cases of autism, which potentially means that there are druggable targets. How do you guys collaborate with drug developers to try work towards some effective therapies?

SS: We take a multifaceted approach; there are some very rare forms of autism that are arising due to mutations in single genes that elucidate either new pathways or typically common pathways. However, these common pathways represent large networks of proteins. This presents a lot of targets in these networks, but you need to understand the mechanism acting within each protein.

There are some very rare cases where some drugs are used in other disorders that overlap with autism, where kids are being tested based on a mutation they carry. Typically we develop induced pluripotent stem cell (IPS) lines from these individuals, and differentiating them into skin cells, blood cells, and also neuronal cells. We test them to see if the drug will influence their neuronal phenotype that has been established using more traditional methods. The community is starting to work together in this area to develop a biobank of these IPS differentiated lines, to make higher throughput screening possible. Drug companies are very interested in the new pathways that are being published on by our group, and by others. Here in Toronto we're starting to do clinical trials just enrolling kids with defined mutations, and using drugs that will target their specific pathways. This is a change from the more traditional retrospective sequencing we've typically seen in clinical trials.

So we're working either directly, or indirectly via publishing, with companies interested in the molecular targets we identify. That's a big part of the MSSNG Sequence Autism Project – to get all that data out there into the public domain so any autism researcher, including drug companies, can use it to push the envelope faster. We try to interact with anybody who wants to move that needle forward. Our area of expertise is to subgroup based on the genotype, doing functional experiments to move the clinical resources into a form in which they can be tested on, and turning it over to the experts in pathway modification.

FLG: What's exciting you most from your research at the moment?

SS: Being able to explain to some of the families why autism comes about in their kids. We have a genetic counsellor working part of our project returning two reports back per week to families. It's amazing being able to give them answers; giving them hope, or giving them information on what they should or shouldn't be doing with their children.

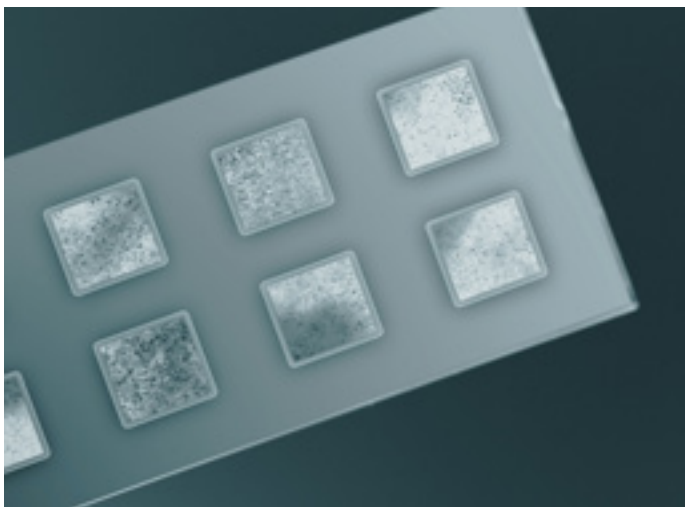
Right now there is no medicine that treats the core features of autism, but we are helping to provide new clues that can get us there. My personal focus is on pushing that envelope further. We have the practical "genomic" aspects set up like an assembly line now, so I can spend time thinking about the intellectual aspects of really refining these drug targets and clinical trial design. I've had some incredible meetings with clinical trial experts from pharma, and in many cases the questions they have asked me have enlightened my own thought process on functional experiments we should be doing.

That interaction is really key for people like myself. It helps us design our functional experiments with clinical trials in mind. We're just entering that phase now in 2017, so I think it's something you'll see more of over the next couple of years that will lead to successes.

FLG: Speaking of successes, you started Genomic Medicine as one of the Nature Partner Journals. How did that come about?

SS: I've been in the field for about 25 years. I've been exceptionally well funded, and the Canadian system has afforded me a lot of freedom to pursue research without feeling the pressure to follow a particular agenda. This isn't something I see in many other countries. →

"THE PAPER SHOWED THE FIRST LINKAGE IN AUTISM TO CHROMOSOME SEVEN, WHICH WAS THE CHROMOSOME WE WERE THE EXPERTS ON. INTERESTINGLY, AND THIS IS THE BEAUTY OF THE STORY, THE SAME DAY THAT PAPER CAME OUT I GOT A FAX FROM A FAMILY IN CALIFORNIA THAT HAD A SON THAT WAS AUTISTIC."



Particularly in non-traditional science countries, it is often very hard for people to get their work published. As the impact of genetics is going to be regional, that's a problem. That's exactly the type of information diagnostic companies and pharma companies need.

Let's take Toronto as an example. It's a very multicultural city, more so than even London or New York now, with around tens of thousands of new immigrants coming each year. They all bring in their own genetic variants, which present a challenge to us. If they are rare variants that have not been previously reported, we don't know they are benign or involved in the disease that brings them to the hospital.

These outlier populations often provide the key clues that help develop new medicines, so we really thought it was time to set up a journal with a global view working in genomic medicine. That's something we reflect in our editorial board, with good representation from across the world. They are typically people like myself who are working in countries that have provided good funding and the freedom to work outside the influence of mainstream thought.

It continues to develop, and we are trying to solicit papers from sources that haven't always had the opportunity to publish in, what we hope will become, a high impact journal. Of course we do entertain any paper. We try to focus on the quality of each paper and the impact it might have in the field of genomic medicine.

We did think long and hard about whether or not there was a need for another journal in this area. Ideally you want to see all the existing journals continue to get better and better, but I didn't see that happening. The field has just gone through the roof, so the idea was to give a shot at trying to have more of a global impact. That's critical because genetics will always play out based on the jurisdiction of where you're studying it. Geneticists all have their own style.

FLG: What does your day-to-day role look like as Editor in Chief of the journal?

SS: I spend about an hour across the day deciding what's going out to review, and which editor is going to handle it. My amazing programme manager here at the McLaughlin Centre, Hin Lee, helps out and interfaces with the journal.

We rely a lot on our associate editors. I often get asked 'How the hell did you get those people?' It's a really amazing group, and we've managed to keep them together because they all share the same philosophy. While we rely on them, we do also try to minimise that work and make it as easy as possible for them.

We've published around 34 papers in our first year, and we'll be on PubMed in the next month or so. All of the papers are good, but there are some outstanding ones in there. According to NPJ (Nature Partner Journals) we are their top NPJ, so it's going well!

FLG: Are you surprised at how well it's gone so far?

SS: I'm surprised we published so many papers. They put a lot of resources into this, so we get a tonne of help. The interaction with the office is great and the associate editors have all been really responsive. I only expected us to publish around 20 papers in the first year. Near the end of last year we got some big papers from big groups, and we're also getting a lot of papers transferred in from other Nature journals like *Nature Genetics, Nature communications,* and *Nature Medicine.* My only concern is how much work I'm giving to the associate editors. But it's been a real success so far. It's been fun, and forces me to read a lot more.

FLG: Who are you really looking to have submit their papers to you?

SS: Ideally for me, I'd love to have the person, whoever they may be around the world, who's earlier in their career that has a great hunch or early preliminary data that they need to get published fast and have an impact. I want them to feel that they can actually benefit by going through the process with us. We hope to get those papers that every once in a while might turn out to be a gem in the future. We can make them shiny faster than other journals.

There are a lot of great local studies coming from different areas in the world where there's a unique family or a unique case report that may open up a whole new drug target pathway or a new diagnostic realm. We're already getting some of those papers, particularly from Asia. We think this is our sweet spot going forward, with the emphasis being on making this a truly international journal.

I had discussions with the editors from the big Nature journals, all of the big names, at the very beginning. They were all very supportive, and I told them that I do expect that if we have a really good paper that comes through *NPJ Genomic Medicine,* that I think should be moved up the chain of impacts I will try to do that. I'm more interested in giving a paper its best landing spot. So if we get a great paper that we think should be in *Nature Genetics,* or wherever, I fire an email off to that editor saying they should look at this paper. I did it a few times in 2016 actually. Some people are surprised, but it is nothing special, just the Canadian thing to do. I think it's a great dynamic and another advantage of submitting to *NPJ Genomic Medicine.* ■

"WHOLE GENOME SEQUENCING IS JUST SPECTACULAR, IT'S LIKE BEING BACK IN THE EARLY 2000'S. IT GIVES US A GOOD VIEWPOINT AT A DECENT COST ACROSS THE WHOLE GENOME. "