Harnessing Gene Discovery and Cellular Models to Understand the Genetics Underlying Neurodevelopmental Disorders

Dr. Lachlan Jolly and his team at the University of Adelaide are applying functional genomic methods to identify the role of *UPF3B* and *HCFC1* in intellectual disability.

Introduction

Living in Australia, Lachlan Jolly, PhD, takes advantage of everything the outdoors has to offer. When Dr. Jolly is not cycling, surfing, hiking, or playing basketball, he's involved with equally expansive research programs to identify and study genes responsible for neurological disorders. To discover novel gene variants and describe their role in intellectual disability, Dr. Jolly integrates cell-based models with a multitude of genetic and functional genomic technologies.

As an Australian Research Council Fellow at the University of Adelaide, Dr. Jolly has a leading role in the Neurogenetics Research Program. While other groups identify gene variants underlying neurodevelopmental disorders, Dr. Jolly directs research to discover the effects of genetic change. "We want to know what functional effect gene variants have in the development of the brain," Dr. Jolly said. "We know that they cause human disorders, but we want to know how they affect the way the cells behave and communicate." The group uses next-generation sequencing (NGS) to identify novel mutations, and cell culture models of embryonic neural development to gain insight into novel genetic networks important for brain development and function. Some of his most recent work has led to the discovery that 2 genes (*UPF3B* and *HCFC1*) on the X chromosome play a role in intellectual disability. At a break in his busy schedule, Dr. Jolly spoke with iCommunity about his gene discovery research.

Q: What sparked your research studies that led to the identification of the *UPF3B* and *HCFC1* gene variants?

Lachlan Jolly (LJ): In 2009, we contributed to a published study of over 200 families with intellectual disability that had a predominance of male members affected, indicating that the X chromosome was likely involved. We performed exome sequencing of the entire coding region on the X chromosomes of this cohort to discover causative gene variants. This large genomic study implicated many genes that I'm now studying, including *UPF3B* and *HCFC1*.

We found several families in the cohort that had complete loss-of-function mutations in *UPF3B*. Based on genetics, we knew immediately that *UPF3B* was critically involved in brain development. We went straight into deciphering how mutations in this gene could lead to the abnormal brain development seen in affected individuals.[1,2]

The *HCFC1* study was initiated when we recognized that there was a large multigenerational family within the cohort that had many affected males. We were certain of X chromosome involvement, but we were unable to find any causative variants in the coding regions that could explain the phenotype. A more extensive search with additional genomics assays suggested involvement of a noncoding variant in the *HCFC1* gene.[3]

Q: What can we learn about brain development from using different types of cell lines?

LJ: We often initiate functional interrogations of new genes and genetic variants with in vitro cell lines because they allow us to screen variants quickly, and decide whether more sophisticated models are warranted. With *in vitro* cell lines, we can access the material easily, and manipulate it genetically to suit our model and obtain a quick readout.

We find that *ex vivo* cultures, derived from the developing brain or embryonic stem cells that have the potential to differentiate into brain cells, can be used as surrogates for brain development. As opposed to static cell models that have no or limited potential for differentiation or modeling aspects of brain development, we choose cell culture models that recapitulate events of brain development. This gives us a correlation between what we see *in vitro* and what might be happening in patients' brains during development. With this information, we can move on to more sophisticated animal models where we can study behavioral and physiological systems, and brain networks.

Lachlan Jolly, Ph.D. is an Australian Research Council Fellow at the University of Adelaide, where he focuses on understanding the genetic networks involved in brain development and function.

Q: What pathway does *UPF3B* impact?
LJ: *UPF3B* is known to participate in the nonsense-mediated mRNA decay (NMD) pathway. NMD was initially discovered as an mRNA surveillance mechanism where it identifies and degrades transcripts with premature termination codons. As such, it is involved in almost one-third of all genetic disorders, the same number of genetic disorders caused by premature termination codons. However, the NMD pathway is now also recognized as a major gene regulatory pathway, with many endogenous mRNA transcripts harboring features that are regulated via NMD. Knocking out the core NMD components, UPF1 or UPF2, results in deregulation of 5–10% of the normal transcriptome. These transcripts are important to development, as embryonic lethal phenotypes are observed in mouse, zebrafish, and fly models. In contrast, the *UPF3B* mutations in humans are not lethal, but instead give rise to intellectual disability. Given its role in NMD, we hypothesize that the NMD pathway might be compromised in affected individuals, and that their transcriptomes might be altered.

> "We are discovering genes that have never been previously associated with genetic disease, and neurodevelopmental brain disorders in particular."

Q: What is the impact of *UPF3B* mutation on brain development and intellectual disability?
LJ: Intellectual disability resulting from the *UPF3B* mutation ranges from mild to severe, and can include additional behavioral problems such as autism, attention deficit hyperactivity disorder (ADHD), and schizophrenia. We've identified a phenotypic spectrum in the affected individuals we've studied. Phenotypic variability existed even within a single family that featured two brothers with the same *UPF3B* mutation, with one brother having severe intellectual disability and the other only mildly affected.

We have discovered a mechanism that potentially modifies *UPF3B* mutation-caused disease outcome. The *UPF3B* gene has a paralog, *UPF3A*. In the absence of *UPF3B* function, *UPF3A* can partially function redundantly, stepping in and taking the place of *UPF3B*. We found that in the mildly affected brother, *UPF3A* expression was elevated compared to his severely affected sibling, a trend we see in other individuals with *UPF3B* mutations. We are now testing in a systematic fashion if *UPF3A* function might be a modifier of the disease outcome. Through our studies, we have generated evidence that shows that loss of *UPF3B* alters the transcriptome in a way the causes changes in how the primordial cells of the developing brain behave. We are now testing if *UPF3A* function might rescue these defects.

Q: What led you to look in noncoding regions to find *HCFC1*?
LJ: We assessed samples from a large multigenerational family with many affected male individuals, screening every coding region on the X chromosome to find a variant. However, we didn't find any coding variants.

We then performed linkage analysis to identify the part of the chromosome that was shared in affected individuals and not shared in unaffected males. That enabled us to focus on a small region of the X chromosome that contained only 108 genes. We sequenced all the coding and regulatory sequences around those genes with the Illumina Genome Analyzer™ System, obtaining extremely high coverage, averaging more than 100 reads per base pair. When we again discovered there was no variant in that region, we looked further outside of the coding region. We found an interesting noncoding variant that affected a highly conserved transcription factor binding site in an intergenic region. The noncoding mutation results in a loss of binding of a transcription factor called YY1, which normally represses *HCFC1* expression.

Q: How did you discover the role of *HCFC1* in normal brain development?
LJ: As *HCFC1* is itself a transcriptional coregulator, we looked to see if in patients we could identify transcriptomic changes that might further implicate its involvement. We used patient-derived blood cell lines, called lymphoblastoid cell lines, to study *HCFC1*. While these cell lines are of blood origin, their transcriptional signature overlaps with those of neurons and brain supporting cells, enabling them to act as surrogates for what might be going on inside the brain. Using the lymphoblastoid cell lines, we found *HCFC1* was overexpressed, and that this was associated with deregulation of about 200 other genes. Running the genes through gene ontology analysis further supported the involvement of *HCFC1*, identifying processes that *HCFC1* was already known to control, such as mitochondrial and chromatin regulation. More importantly, the gene ontology analysis linked *HCFC1* with neuronal development and differentiation. From there, we decided to model the effect of *HCFC1* overexpression to discover its role in brain development.

> "Illumina technologies are absolutely critical for us to understand the genetics behind undiagnosed cases of neurodevelopmental disorders."

Q: How did you model *HCFC1*?
LJ: The gene ontology pathway indicated that *HCFC1* was involved with forebrain development. We isolated neurons and neural stem cells from embryonic mouse forebrain and cultured them *in vitro*. This gave us an opportunity to manipulate them and try to create a model of the disease. We overexpressed *HCFC1*, as is the case in patient-derived cell lines, and monitored how this affected behavior of the neural cells. *HCFC1* had a potent effect on normal cell behavior. The neurons displayed reduced axonal

and dendrite growth, while overexpression in neural stem cells induced them to exit the cell cycle and differentiate into astrocytes. This provided supporting evidence that the noncoding mutation upstream of *HCFC1* that resulted in its overexpression would likely have an effect on the way that human brains develop *in vivo*.

Q: How does overexpression of *HCFC1* lead to intellectual disability?

LJ: That is difficult to answer currently. We don't have enough information to link what we're seeing *in vitro* to what we can derive from the structure of the brain. The patients haven't had detailed MRI analysis to look at their brain structure. I can hypothesize that when you have alterations in the way a neural stem cell grows, the phenotypic correlate would be microcephaly if you're exiting the cell cycle, or macrocephaly if the neural stem cell is endowed with extra proliferative capacity. It could be that there are malformed parts of the brain we're not aware of, which would be consistent with what we're seeing *in vitro*. Other factors with more subtle effects might be contributing, such as defective neuronal cell connectivity or communication. More sophisticated models such as genetically modified mice are required to better understand the role of *HCFC1* in brain disorders.

> "Our functional studies help indicate which cell types or aspects of brain development are affected, and could lead to the discovery of new therapeutic targets and regimes."

Q: How do you decide which technologies to use in your research?

LJ: If we find a gene that we think is involved, we're guided by what is known about that gene. *UPF3B* and *HCFC1* are both genes that regulate gene expression. *UPF3B* is part of an mRNA decay mechanism and *HCFC1* is a transcriptional coregulator. Knowing the transcriptome might be affected, we used RNA-Seq to analyze how much influence these particular mutations have on the transcriptome of patient-derived cell lines. If this information implicates any aspects of brain development, we could look for alterations using *in vitro* neural cell model systems. In the case of these 2 genes, this approach was very successful.

Q: How do Illumina sequencing systems and arrays enable your studies?

LJ: We perform exome sequencing on the HiSeq™ 2500 System for gene discovery. We use the HiSeq X™ Ten System for whole-genome sequencing and we use Illumina arrays to perform linkage and loci mapping. Illumina technologies are absolutely critical for us to understand the genetics behind undiagnosed cases of neurodevelopmental disorders. They enable us to discover new genes, which we can then analyze further by conducting functional studies in cell-based models.

For the *HCFC1* project, we performed transcriptome profiling with RNA-Seq on the HiSeq 2500 System, and with the HumanHT-12 v4 Expression BeadChip. We chose this array because it offers good concordance with qPCR and RNA-Seq data.

For the *UPF3B* project, we used RNA-Seq and HumanOmni BeadChips to analyze copy number variants within patient cell lines. Studying copy number variants is important because each individual has their own personal genomic changes. Whether a gene expression change can be correlated to a copy number gain or loss was something we were interested in studying. Specifically, we wanted to understand if the transcriptome changes in individuals with *UPF3B* mutations were due to defective NMD, or due to a personal copy number variant.

Q: How could your findings with *UPF3B* and *HCFC1* improve human health?

LJ: We are discovering genes that have never been previously associated with genetic disease, and neurodevelopmental brain disorders in particular. When we publish the phenotypes and genotypes of what we discover in various public databases, this contributes to our collective understanding of the potential causes of intellectual disability that affects both research efforts and clinical diagnosis. We hope that by highlighting these cases further, it will lead to diagnostic tests and solutions for others affected by changes in these genes.

In terms of contributing to the development of new therapies, we are further away. Our functional studies help indicate which cell types or aspects of brain development are affected, and could lead to the discovery of new therapeutic targets and regimes. For example, we have on-going interrogations of *UPF3A* and its potential role in modifying the outcome of the *UPF3B* mutations. However, the development and scientific testing of therapeutic strategies is a process that often takes many years or decades. We're just at the beginning of the process in studying these newly discovered genes.

> "There will be many variants that end up with a question mark next to them. The use of NGS in more research-based environments will be required to aid the process of assigning function."

Q: How do you see NGS impacting your research in the future?

LJ: NGS has emerged from a primarily research-based tool to a prevalent method in clinical diagnostics. As such, many new genetic variants are being identified, but with this comes a mounting challenge of assigning function to variants of unknown significance. We can compare large data sets and cohorts together, and this provides some utility in assigning significance, such as how rare it might be, or if it has been previously found in patients with similar phenotypes. However, there will be many variants that end up with a question mark next to them. The use of

NGS in more research-based environments will be required to aid the process of assigning function, as we have done in our studies. For example, interrogation of the transcriptome in our UPF3B and HCFC1 patient cell lines, for which NGS is well-suited, provided evidence of their involvement, and guided our neural cell-based studies to help identify the developmental, cellular, and molecular mechanisms. NGS technologies will undoubtedly play a significant role, supporting an ever building need for high-throughput, and/or high-depth analysis of samples to answer clinical diagnostic and more research-focused questions.

Learn more about the products and systems mentioned in this article:

HiSeq 2500 Systems, www.illumina.com/systems/hiseq_2500_1500.html

HiSeq X Ten Systems, www.illumina.com/systems/hiseq-x-sequencing-system.html

HumanHT-12 v4 Expression BeadChips, www.illumina.com/products/humanht_12_expression_beadchip_kits_v4.html

Omni Whole-Genome DNA Analysis BeadChips, www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_omni_whole-genome_beadchips.pdf

References

1. Nguyen LS, Jolly L, Shoubridge C, et al. Transcriptome profiling of *UPF3B*/NMD-deficient lymphoblastoid cells from patients with various forms of intellectual disability. *Molecular Psychiatry.* 2012;17:1103–1115.
2. Jolly LA, Homan CC, Jacob R, et al. The *UPF3B* gene, implicated in intellectual disability, autism, ADHD, and childhood onset schizophrenia regulates neural progenitor cell behavior and neuronal outgrowth. *Hum. Mol. Genet*. 2013;22:4673–4687.
3. Huang L, Jolly LA, Willis-Owen S, et al. (2012). A Noncoding, Regulatory Mutation Implicates *HCFC1* in Nonsyndromic Intellectual Disability. *Am. J. Hum. Genet*. 2012;91:694–702.

Illumina, Inc. • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com