Finding the Genetic Voice Underlying Language and Reading Disorders

Yale researchers use Infinium™ arrays and the HiSeq™ System to uncover genetic variants associated with developmental language and reading disorders.

Introduction

Language forms the thread that holds humans together in families and communities. It lets us express our needs, summon help, and share our innermost wants and desires. Most of us find speech a simple and straightforward way to communicate. However, for the estimated ~10 percent of individuals with a language disorder, including developmental language disorder (difficulty mastering the foundations of language, such as grammar) or specific reading disorder (trouble mastering reading, despite normal intelligence), it's anything but simple.[1] Language disorders include difficulties with speaking or writing language; understanding meaning, grammar, or syntax; or reading and spelling. Because many of these disorders cluster in families, scientists have long believed that genetic factors play an important role in the development of language difficulties. With a few important exceptions, such as the discovery that mutations in the *FOXP2* gene cause verbal dyspraxia (difficulty producing clear speech) in some families, the genetics of language disorders remains unclear.

Yale University geneticist Elena Grigorenko, PhD became interested in neurology and the genetics underlying child development as an undergraduate and graduate student at Moscow State University, Russia. She has spent most of her 25-year career studying the genetic factors associated with child development, language disorders, and other disabilities. After receiving her PhD in psychology in 1990 (in what was then the Soviet Union), she moved to Yale where she earned a PhD in developmental psychology and genetics, and later formed a behavioral and molecular genetics research lab (Elena Grigorenko lab or EGLAB). She is now transitioning to a new position at the University of Houston and Baylor College of Medicine.

"It wasn't possible to conduct the kind of human behavior research that I wanted to in Russia," Dr. Grigorenko said. "When I came to the US and began working at Yale, it seemed like a great opportunity to develop this line of research. My lab focuses on several developmental disorders, including language disorders and learning disabilities (eg, reading disability or dyslexia), as well as conduct behavior (a pattern of disruptive and/or violent behavior in children and adolescents)."

Her career has enabled her to watch and participate in the transition from Sanger sequencing to arrays to next-generation sequencing (NGS). Dr. Grigorenko, Dr. Sergey Kornilov (a postgraduate associate), Maria Lee Eastman (a research associate), and their lab colleagues use Illumina Infinium arrays and the HiSeq System to identify genes that are involved in normal child brain development, and mutations that affect a child's ability to comprehend and/or produce normal language.

In the beginning, genetic studies using RFLP and STRP

Genetic studies have revealed that developmental speech and language disorders are strongly heritable. Researchers now believe that these disorders are the result of many genes working in concert. Although some children outgrow their language difficulties, others do not. It is this subset of children that are most likely to demonstrate academic and psychiatric impairments. Identifying the specific genetic pathways that contribute to language disorders could help scientists understand how the brain develops and processes language.

When Dr. Grigorenko formed the EGLAB at Yale, she relied on Sanger sequencing and genotyping using restriction fragment length polymorphisms (RFLPs) and short tandem repeat polymorphisms (STRPs) to understand the complicated genetics of language and developmental disorders. DNA microarray technology was then introduced, helping scientists to associate single nucleotide polymorphisms (SNPs) to specific traits. Dr. Grigorenko was one of the first scientists to apply for an NIH grant to study language disorders based on the use of microarrays.

Dr. Elena Grigorenko was the Emily Fraser Beede Professor of Developmental Disabilities, Child Study, Epidemiology and Public Health, and Psychology at Yale University. She is now the Hugh Roy and Lillie Cranz Cullen Distinguished Professor of Psychology at the University of Houston.

For Research Use Only. Not for use in diagnostic procedures.

1070-2015-007-A | 1

Infinium microarrays burst onto the scene

Dr. Grigorenko had been outsourcing her array studies to the Keck Core lab, now known as the Yale Center for Genome Analysis. As time progressed, she began performing more specialized types of genotyping, gene expression, and methylation studies.

"We talked with four or five principal investigators (PIs) in other labs who were performing genotyping, gene expression, and epigenetic work on campus," Dr. Grigorenko added. "We realized that there was a lot of local expertise at Yale that we could rely on as we came up to speed using Infinium arrays, and that we could call or drop by their labs with questions."

"Infinium BeadChips are inexpensive, reliable, and robust," said Ms. Eastman. "We have processed more than 5000 samples using Infinium arrays and are still sorting through the data. These arrays are easy-to-use and informative."

Over the last 10 years, the EGLAB has used various Infinium genome-wide association study (GWAS) arrays, including the HumanCoreExome, and custom Infinium BeadChips. They've also used the HumanMethylation450 BeadChip for epigenetic studies.

> ## "Infinium BeadChips are inexpensive, reliable, and robust."

The move to NGS

Not long after Ms. Eastman joined the EGLAB, NGS became more affordable and Dr. Grigorenko decided they were ready to bring NGS-powered studies in house. The lab already had familiarity with Illumina products and they knew many other researchers who used HiSeq Systems, making it an easy decision. The EGLAB began a small pilot project with the HiSeq System to study the transcriptome of 3 different areas of the neocortex at various ages.[2] They also began performing ChIP-Seq epigenetic studies on the HiSeq System.

"HiSeq System data quality is top notch," Dr. Grigorenko said. "The surprise for us was that sequencing generated new analytic and data management challenges that we were not ready for initially. We worked with the NGS experienced colleagues and with Illumina to determine how to process the data efficiently."

> ## "HiSeq System data quality is top notch."

Identifying genetic variants with the HiSeq System

The EGLAB team has been steadily identifying genes that affect language and reading abilities using the HiSeq System. As with other language disorders, researchers knew that genetics played a significant role in dyslexia, although little work had been done linking reading ability to specific SNPs. In a 2013 study, Dr. Grigorenko and colleagues from Yale and Haskins Laboratories found that individuals carrying a substitution mutation in the catechol-O-methyltransferase (COMT) gene, where the ancestral valine was replaced by methionine

at rs4680, demonstrated better performance on reading-related tests.[3] They also discovered neural activation patterns that were linked with better reading skill. Because other studies have linked this polymorphism to broad effects on overall cognition, Dr. Grigorenko believes that it likely modulates reading ability via the functioning of the frontal lobe. These types of polymorphisms, combined with developmental changes in gene expression, appear to play a significant role in brain development and the development of language disorders.[4]

More recently, Dr. Grigorenko and members of her lab have begun performing whole-exome sequencing to understand language problems and conduct disorder in children.[5] They began this work when researchers at the Yale Center for Mendelian Genomics contacted them, asking whether their studies had uncovered samples with interesting or unique pedigrees in which specific disorders segregated in a Mendelian fashion. Dr. Grigorenko surveyed her collection, identifying several Mendelian segregating segments and sequencing a subportion of those samples. Dr. Grigorenko found the data eye-opening.

"The language reflection provided interesting findings, pointing to a developmental, highly significant pathway that seems to be derailed in the various stages of brain maturation," Dr. Grigorenko said. "After we gathered leads from sequencing data, we looked at the brain expression data that were publically available. We found cross-references that make our story much more compelling and interesting. We are excited about this pattern of results because it maps well with previous findings that were sparsely distributed through the literature."[5]

> ## "After we began using NGS in our studies, we transitioned to using arrays for validating sequencing results."

Integrating array and NGS data

Dr. Grigorenko and colleagues in the EGLAB and other Yale labs are also beginning a new study that integrates various sequencing techniques and analyses. In collaboration with the laboratory of Dr. Matt State, the lab published a case study in 2012 of a young boy who was diagnosed with a developmental language disorder.[6] Sequencing his DNA revealed a balanced t(10;15)(q24.1;q21.1) translocation. The breakpoint on 15q21.1 interrupted the gene coding for SEMA6D, which belongs to a family of proteins called semaphorins that are involved in axon pathfinding in the developing brain. Because this gene has been linked to a language disorder in one child, Dr. Grigorenko and her colleagues hypothesize that polymorphisms in this gene might affect language disorders in other children. As a result, they have begun a targeted sequencing study that sequences the SEMA6D gene in a larger group of children with language disorders to see if they can find any links.

"The *SEMA6D* gene is huge (55,735 bp), so we're moving forward, albeit slowly," Dr. Grigorenko said.

Dr. Grigorenko has found that integrating array and NGS data in her studies has been extremely helpful. In one study, their microarray results did not map directly on the lead provided by NGS.[5] Yet, upon closer inspection, they were highly complementary and converged on a particular pathway.

"When you contextualize it within the literature, it all makes much more sense from our point of view that a single finding of a single gene provides an endpoint for everything to converge on," Dr. Grigorenko said. "We're quite excited about this finding."

Ms. Eastman finds that both types of genetic analysis data are crucial to making discoveries. "We like the arrays because they offer higher throughput than running SNP genotyping analysis by hand," Ms. Eastman said. "After we began using NGS in our studies, we transitioned to using arrays for validating sequencing results."

Moving forward

Dr. Grigorenko and her colleagues have become fascinated with using expression quantitative trait loci (eQTL) mapping to detect polymorphisms. eQTLs are regions of the genome containing DNA sequence variants that influence the expression level of one or more genes. "Developmental disorders are complex, multifaceted, and the phenotypes change developmentally," Dr. Grigorenko said. "We're beginning studies using eQTL mapping with NGS data to detect genetic variants for these complex phenotypes. It's very exciting."

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

Infinium HumanCoreExome BeadChip,
www.illumina.com/products/humancore_exome_beadchip_kits.html

Infinium HumanMethylation450 BeadChip,
www.illumina.com/products/methylation_450_beadchip_kits.html

Infinium iSelect™ Custom Genotyping BeadChips,
www.illumina.com/products/infinium_iselect_custom_genotyping_beadchips.html

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

References

1. Spoken Language Disorders. American Speech-Language-Hearing Association. www.asha.org/PRPSpecificTopic.aspx?folderid=8589935327&section=Incidence_and_Prevalence. Accessed February 22, 2016.

2. Naumova OY, Palejev D, Viasova NV, et al. Age-related changes of gene expression in the neocortex: preliminary data on RNA-Seq of the transcriptome in three functionally distinct cortical areas. *Dev Psychopathol*. 2012; 24(4): 1427–1442.

3. Landi N, Frost SJ, Menci WE, et al. The COMT Val/Met polymorphism is associated with reading related skills and consistent patterns of functional neural activation. *Dev Sci*. 2013; 16(1): 13–23.

4. Naumova OY, Lee M, Rychkov SY, et al. Gene Expression in the Human Brain: The Current State of Study of Specificity and Spatio-Temporal Dynamics. *Child Dev*. 2013; 84(1): 76–88.

5. Kornilov S, Rakhlin N, Koposov R et al. Genome-wide association and exome sequencing study of language disorder in an isolated population. *Pediatrics*. 2016; pii:peds.2015–2469. [Epub ahead of print].

6. Ercan-Sencicek AG, Davis Wright NR, Sanders SJ, et al. A balanced t(10;15) translocation in a male patient with developmental language disorder. *Eur J Med Genet*. 2012; 55(2): 128–131.