

Using Large Sample Sizes and Deep Sequencing to Find the Root Cause for Psychiatric Disorders

Dr. James Knowles trusts the HiSeq® 2000 system and TruSeq® Sample Preparation Kits for cost-effective, large-scale collaborative projects.

Introduction

As a Professor of Psychiatry at the Keck School of Medicine at the University of Southern California (USC), James A. Knowles, M.D., Ph.D., is interested in discovering the genetic factors that have an etiological role in the development of psychiatric illness. The genetic component is confined to the three billion base pairs of the human genome, whereas the environmental component is much broader, including both prenatal and lifelong interactions. Dr. Knowles explains, “The most tractable way to approach the problem is to look in the genome to see if we can find the genetic variations that predispose individuals to psychiatric illnesses. Once we discover and understand the involvement of genetics, we can parse out the effect of the environment.”

A Strategy to Determine the Location of Gene Targets

“In GWAS studies of equal size, more genome-wide significant loci are discovered for the metabolic or autoimmune disorders than the psychiatric disorders,” says Dr. Knowles. “It may be that the risk loci for the psychiatric disorders are spread more broadly across the genome, as the vast majority of genes are expressed in the brain during sometime during the lifespan.” This means a larger number of targets could be mutated, causing cognitive and psychiatric symptoms. If this is the case, the risk for any given loci may be smaller and hence more samples are needed to filter out the noise. “But that’s hand waving until we actually have the data and can definitively show that.” For this, Dr. Knowles put his trust in proven Illumina next-generation sequencing technology.

Trusting Illumina for Large-Scale Projects

For such large-scale studies, Knowles needs to tally up rare variants across different transcription units throughout the genome to reach statistical significance. That requires large sample sizes, combined with deep sequencing of all samples. “When HiSeq came out, it was very clear it was going to provide the most amount of information for the least cost,” explains Dr. Knowles. His lab has been using Illumina DNA sequencers since the original Genome Analyzer™ sequencer was introduced and found the HiSeq 2000 to be a natural progression, since they already had the processes in place to make libraries and analyze data. Dr. Knowles explains, “Switching to something like SOLiD, which has a very different laboratory and informatics workflow, seemed like a lot of work for an unclear gain. HiSeq was a winner from that point of view. With HiSeq we’re getting Q30s of 90% in every lane for both reads at 100 bp. The data quality has just gotten increasingly better.”



James Knowles, M.D., Ph.D. Professor of Psychiatry at the Keck School of Medicine at USC, uses the Illumina HiSeq sequencing system to study psychiatric disorders.

“When HiSeq came out it was very clear it was going to provide the most amount of information for the least cost.”

Deploying HiSeq for a Variety of Applications

The Knowles lab is using the HiSeq system in a number of ways to try to understand the basis of psychiatric disease. When performing DNA sequencing, they are predominantly doing whole-genome sequencing at higher levels of coverage. “If it’s a rare inbred family, or a large family with multiple affected individuals, we will use the HiSeq system to sequence at higher depth, 30x or 40x, to ensure we discover all variants within a particular individual, or small set of individuals, from that family,” Dr. Knowles describes. In other rare singleton cases, such as individuals who had onset of schizophrenia before the age of 18, they’ll also sequence at high coverage, 30x or 40x, “because we don’t want to miss any potentially causative variants”.

The Knowles lab is also pursuing a sequencing strategy similar to that of the 1000 Genomes Project. “We’re sequencing 700 individuals with schizophrenia and 700 controls at lower coverage and with the current output we are able to generate 12–15x coverage in a single lane,” he notes. They’ll also do a case control analysis of the whole-genome data and use it to impute the less frequent polymorphisms into larger sets of samples that have been genotyped with GWAS chips. “This HiSeq data will hopefully provide us with substantially more discovery of all the sequence variants in the disease population and might ultimately lead to a fuller understanding of the disorder. Particularly if what is causing the psychiatric disorders is a number of rare variants that are only seen in a few patients.”

As part of a collaboration with Yale University and the Allen Brain Institute, the Knowles lab is using RNA-Seq on HiSeq to determine the transcriptional activity associated with the developing brain. “We’re sequencing up to 66 human brains from 11 stages of development and doing RNA-Seq on about 1,000 samples.”

The Knowles lab also uses targeted sequencing. They’ve done some work with pulmonary arterial hypertension, a Mendelian disorder where three disease genes are already known. Here they’re using the Illumina TruSeq Exome Enrichment Kit on HiSeq to find other disease genes for pulmonary arterial hypertension.

“Unimaginable” Genome-Scale Sequencing as Reality

“We think we need to have the complete genomic sequence of each individual to really get the complete architecture of a disease like schizophrenia,” asserts Dr. Knowles. He hopes to understand the top 100 or 1,000 loci for each of the different complex disorders and build models that predict risk and outcome and guide treatment. “We’re sequencing genomes on a scale that was just inconceivable to me. We’ll do 115 low-coverage genomes a month. The throughput on HiSeq enables experiments that were previously unimaginable and hopefully will lead to finding causal mutations for complex disorders.”

Learn more about the portfolio of HiSeq systems at www.illumina.com/hiseq

Summary

Overview

Dr. James Knowles at USC is interested in discovering genetic factors that predispose individuals to psychiatric disorders, to ultimately predict risk and outcome and guide treatment choices.

Challenge

SNPs and loci for psychiatric disorders might be spread across the genome, making them difficult to capture in GWAS studies.

Solution

Dr. Knowles uses the HiSeq 2000 system to sequence large numbers of samples at deep coverage to reach statistical significance.

Benefits

Illumina technology makes previously unimaginable large-scale sequencing studies possible through fast, accurate, and cost-effective platforms.

“The throughput on HiSeq enables experiments that were previously unimaginable and hopefully will lead to finding causal mutations for complex disorders.”

Illumina • 1.800.809.4566 toll-free (U.S.) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

FOR RESEARCH USE ONLY

© 2012 Illumina, Inc. All rights reserved.
Illumina, illuminaDx, BaseSpace, BeadArray, BeadXpress, cBot, CSPro, DASL, DesignStudio, Eco, GAllx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, NuPCR, SeqMonitor, Solexa, TruSeq, VeraCode, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks or registered trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners.
Pub. No. 770-2011-033 Current as of 26 July 2012

illumina[®]