Screening and characterization of causative structural variants for bipolar disorder in a significantly linked chromosomal region on Xq24-q27 in an extended pedigree from a genetic isolate

Jenny M Ekholm1, Aaron Wenger1, Steve Kujawa1, Anand Sethuraman1, Petri Auvinen2, Lars Paulin1, Olli-Pekka Smolander2, Pia Laine1, Denise Raterman1, Todd A. Richmond1, George Mayhew3, Jennifer Wendt1, Daniel L. Burgess1, Timo Partonen1, Johana Liuhanen1,6, Tiina Paunio1,4
1PacBio, 1305 O’Brien Drive, Menlo Park, CA 94025, 2Institute of Biotechnology, University of Helsinki, Helsinki, Finland, 3Roche NimbleGen, Inc, Madison, WI 53719, USA, 4National Institute for Health and Welfare (THL), Department of Health, Helsinki, Finland, 5Department of Psychiatry, University of Helsinki

Introduction

Bipolar disorder (BPD) is a phenotypically and a genetically complex and debilitating neurological disorder that affects 1% of the worldwide population. There is compelling evidence from family, twin and adoption studies supporting the involvement of a genetic predisposition in BPD with estimated heritability up to ~ 80%. The risk in first-degree relatives is ten times higher than in the general population. Linkage and association studies have implicated multiple putative chromosomal loci for BPD susceptibility, however no disease genes have been identified to date.

Here, our aim is to characterize the ~12 Mb significantly linked genomic region on chromosome Xq24-q27 in an extended family from genetically isolated population using long-read Single Molecule, Real-Time (SMRT) Sequencing. It has been demonstrated in several studies that PacBio long reads discover many structural variants (SVs) missed by short-read sequencing.

The selected family segregates BD in at least 4 generations with 16 out of 61 affected individuals. Thus, this family portrays a highly elevated reoccurrence risk compared to the general population.

Methods

For sequencing, we have selected 16 key individuals from the X-chromosomally linked family who either carry the disease haplotype, are non-carriers of the disease haplotype, or who married into the family and therefore serve as controls. We designed a Nimblegen capture array that enriches for 5-9 kb fragment spanning the entire 12 Mb region. These were sequenced using long-read SMRT Sequencing and screened for potential causative variants.

Results

The linkage (lod score=3.54) was originally discovered as part of the genome-wide survey using microsatellite markers. The region was then further narrowed down to a ~12 Mb critically linked region. It is expected the genetic complexity would be reduced in isolated populations, even in genetically complex disorders such as BPD, as in the case of this extended family. The lack of linkage evidence to other genomic regions aside from Xq24-q27 supports this.

Table 1 / Figure 1. SV discovery comparisons in NA12878. The 10-fold PacBio call set recovers (A) 88% of true deletions, and (B) 81% of true insertions. The 10-fold PacBio set also includes thousands of novel variants, most of which are directly confirmed by a FALCON-Unzip de novo assembly from 60-fold PacBio RS II coverage.

Table 2. All 16 samples and # Events

For Research Use Only. Not for use in diagnostics procedures. © Copyright 2017 by Pacific Biosciences of California, Inc. All rights reserved. Pacific Biosciences, the Pacific Biosciences logo, PacBio, SMRT, SMRTbell, Iso-Seq, and Sequel are trademarks of Pacific Biosciences. BluePippin and SageELF are trademarks of Sage Science. NGS go and NGSengine are trademarks of GenDx. FEMTO Pulse and Fragment Analyzer are trademarks of Advanced Analytical Technologies. All other trademarks are the sole property of their respective owners.

References

7) Rosas AO, et al. (2016) "Structural variants can be more informative for disease diagnosis, prognosis and translation than current SNP mapping and exome sequencing. Expert Opin Drug Metab Toxicol. 12(2),135-47.