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 Ekholm[1], Aaron Wenger[1], Steve Kujawa[1], Anand Sethuraman[1], Petri Auvinen[2], Lars Paulin[2], Olli-Pekka Smolander[2], Pia Laine[2], Denise Raterman[3], Todd A. Richmond[3], George Mayhew[3], Jennifer Wendt[3], Daniel L. Burgess[3], Timo Partonen[4], Johana Liuhanen[4,5], Tiina Paunio[4,5]

[1]PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025, [2]Institute of Biotechnology, University of Helsinki, Helsinki, Finland, [3]Roche NimbleGen, Inc, Madison, WI 53719, USA, [4]National Institute for Health and Welfare (THL), Department of Health, Helsinki, Finland, [5]Department 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.

Set	Platform	Deletions	Insertions
1000 Genomes[1]	Illumina	1,910	1,090
Genome in a Bottle[2]	Multiple	2,668	n/a
10-fold PacBio[3]	**Sequel System**	**8,209**	**11,350**

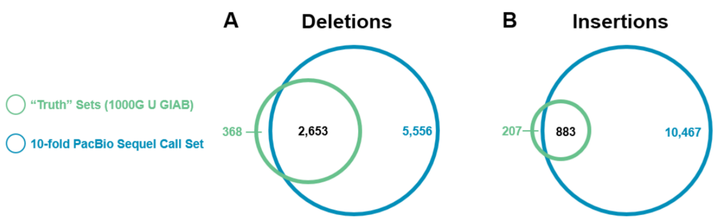

Table 1 / Figure 1. SV discovery comparisons in NA12878. The 10-fold PacBio call set recovers (A) 88% of true deletions, and (B) 81% of true insertions. The 10-fold PacBio set also includes thousands of novel variants, most of which are directly confirmed by a FALCON-Unzip *de novo* assembly from 60-fold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

Original Linkage Finding in Extended Pedigree

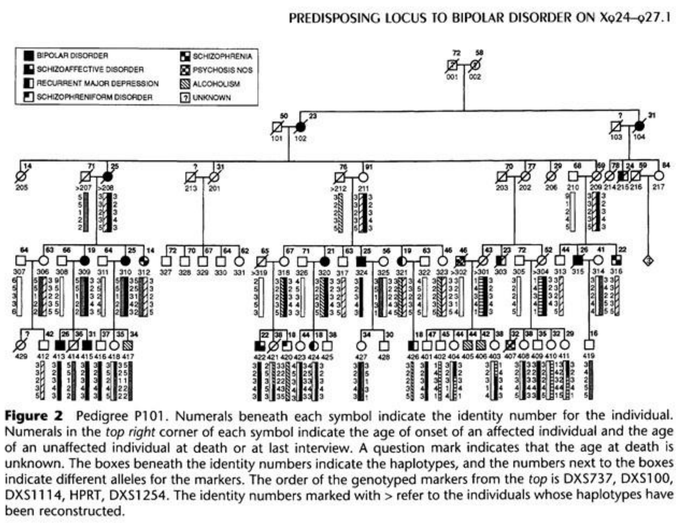

Figure 2 Pedigree P101. Numerals beneath each symbol indicate the identity number for the individual. Numerals in the *top right* corner of each symbol indicate the age of onset of an affected individual and the age of an unaffected individual at death or at last interview. A question mark indicates that the age at death is unknown. The boxes beneath the identity numbers indicate the haplotypes, and the numbers next to the boxes indicate different alleles for the markers. The order of the genotyped markers from the *top* is DXS737, DXS100, DXS1114, HPRT, DXS1254. The identity numbers marked with > refer to the individuals whose haplotypes have been reconstructed.

Figure 2. Extended pedigree linked to chromosome Xq24-q27.1[4]

Original Linkage Finding in Extended Pedigree

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.

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 fragments spanning the entire 12 Mb region. These were sequenced using long-read SMRT Sequencing and screened for potential causative variants.

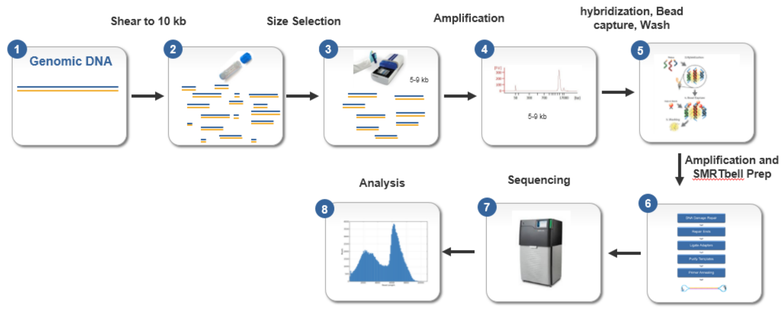

Figure 3. The targeted sequencing workflow. The region of interest was enriched using NimbleGen custom designed probe set and sequenced using SMRT Sequencing.

PacBio long reads were mapped using NGM-LR (https://github.com/philres/ngmlr), a tool that has been designed specifically for structural variant discovery in PacBio reads, and pbsv chain, that connects primary and supplementary local alignments from one read into a merged alignment. Structural variants were called using PBHoney (https://sourceforge.net/projects/pb-jelly/) and visualized in IGV v.3 with optimized settings for PacBio long reads.

Results

All 16 samples	# Events
Deletions >100 bp	923
Insertions >100 bp	3961

Table 2. Aside from SNPs and smaller variants, altogether 923 deletion and 3961 insertions larger than 100 bp were discovered in the 16 samples.

All deletion and insertion calls were merged across individuals to see if any variant was present in many affected (n= 7) individuals, but few unaffected (n= 9) individuals. One variant was called in 6 affected but 0 unaffected individuals.

Results

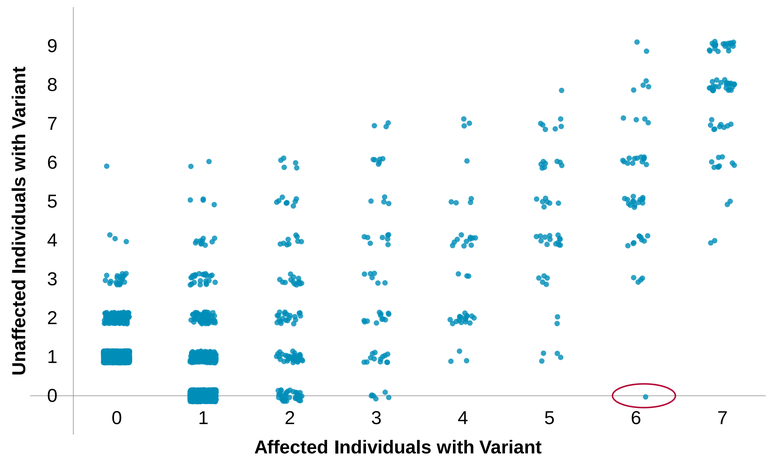

Figure 4. Variant calls for affected vs. unaffected. The dot plot shows the affected individuals on the X-axis and unaffected on Y-axis. The circled variant is seen in 6 affected and 0 unaffected individuals.

The insertion is located in intron 1 of the Fibroblast Growth Factor 13 (*FGF13*) gene which is a microtubule-stabilizing protein regulating neuronal polarization and migration. Secretory fibroblast growth factors and their receptors are known for their regulatory function in the early stages of neural development[4].

The loss of *FGF13* impairs neuronal polarization and increases the branching of axons and leading processes. Genetic deletion of *FGF13* in mice results in neuronal migration defects in both the neocortex and the hippocampus. Mice lacking *FGF13* gene show defects in cortical development, learning, and memory[4].

The *FGF13* gene was also one in 13 genes that was differently expressed in brain tissue of both schizophrenia and bipolar disorder patients[5].

Conclusion

PacBio SMRT Sequencing identified a candidate variant that is present in 6/7 affected individuals and 0/9 unaffected individuals in the pedigree. The candidate is an insertion in an intron of *FGF13*. Such regions have been shown to regulate different elements of mRNA transcription and processing. In fact, intron splicing regulatory elements are frequently located in introns and mutations in cis-regulatory elements can cause dysregulation of splicing and alternative splicing[6]. However, further studies are needed to investigate the function of this insertion in larger study cohort as well using functional studies.

References

1) Sudmant PH, et al. (2015). An integrated map of structural variation in 2,504 human genomes. *Nature*, 526(7571), 75-81.
2) Parikh H, et al. (2016). svclassify: a method to establish benchmark structural variant calls. *BMC Genomics*, 17:64.
3) Wenger A, et at (Feb, 2017) Structural variant detection with low-coverage PacBio sequencing, *Poster #715 presented at Advances in Genome Biology and Technology*, Hollywood, FL.
4) Pekkarinen P, et al. Evidence of a predisposing locus to bipolar disorder on Xq24-q27.1 in an extended Finnish pedigree. Genome Research. 1995 5(2),105-115.
5) Wu QF, et al. (2012) Fibroblast growth factor 13 is a microtubule-stabilizing protein regulating neuronal polarization and migration. *Cell.* 149(7), 1549-1564.
6) Zhao Z, et al. (2015) Transcriptome sequencing and genome-wide association analyses reveal lysosomal function and actin cytoskeleton remodeling in schizophrenia and bipolar disorder. *Mol Psychiatry.* 20(5), 563-72.
7) Roses AD. et al . (2016). Structural variants can be more informative for disease diagnostics, prognostics and translation than current SNP mapping and exon sequencing. *Expert Opin Drug Metab Toxicol.* 12(2),135-47.