Integration of Data Sources and Omics to Identify Susceptibility Variants for Bipolar Disorder

K Truvé[1], D Vizlin-Hodzic[2], TZ Parris[3], S Illes[3,4,5], S Salmela[3], H Ågren[6], K Funa[3,7]

1 Bioinformatics Core Facility, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden. 2 Sahlgrenska University Hospital, Clinical Pathology and Genetics, Gothenburg. 3 Department of Oncology, Institute of Clinical Sciences, Sahlgrenska Cancer Center, Sahlgrenska Academy at University of Gothenburg. 4 Institute of Neuroscience and Physiology, Department of Physiology, Sahlgrenska Academy at University of Gothenburg. 5 Cellectricon AB, Mölndal, Sweden. 6 Institute of Neuroscience and Physiology, Section of Psychiatry and Neurochemistry, Sahlgrenska Academy at University of Gothenburg. 7 Sahlgrenska Cancer Center at the Sahlgrenska Academy, Institute of Biomedicine, University of Gothenburg, Gothenburg, Sweden.

Aim

The focus of our current research is to identify variants and genes involved in development of bipolar disorder (BD). Ultimately, knowledge gained could be used to develop diagnostic tests and improve current treatments.

Background

Bipolar disorder (BD) is a severe chronic psychiatric disorder affecting >1% of the population worldwide [1]. The disease is characterized by recurrent episodes of mania and depression. About 15% of patients with bipolar disorder are expected to die from suicide [2]. Thus, early detection, diagnosis and initiation of correct treatment are critical. In an earlier study aiming to identify shared mis-regulated genes or pathways, we combined iPCS technology and neural differentiation with RNA-seq to investigate differences in the global transcriptome of neural stem cells (NSC) between BD patients and healthy controls. We found *NLRP2* to be the most significant differentially expressed gene, demonstrating a clear difference in expression for all cases and controls (Fig 1).

Study design

We performed Whole Genome Sequencing (WGS) in the same sample as we previously performed RNA-seq. Variants were filtered as illustrated in Fig 2, and candidate variants were genotyped in a larger cohort. Variants in candidate genes were further investigated in a dbGAP data set (n=66) with WGS data from six large families affected with DB.

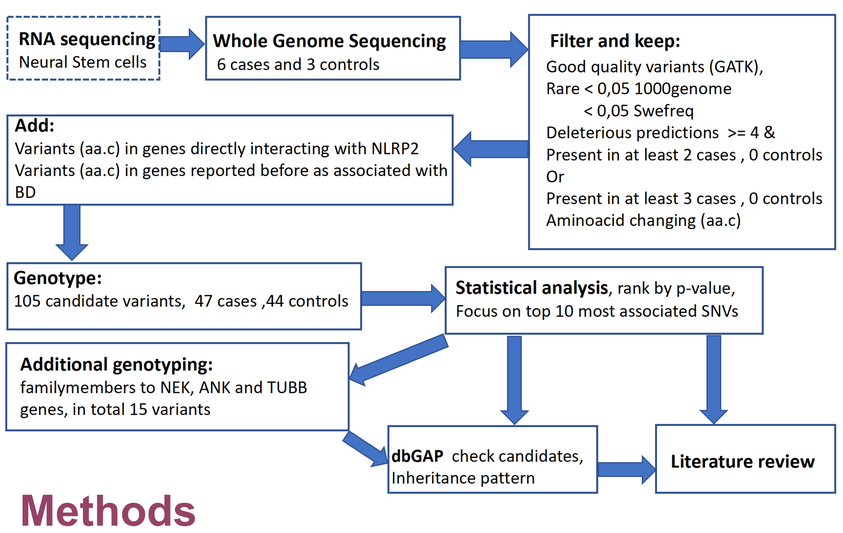


Fig 2. Overview of workflow showing steps taken in the analysis, including different types of data resources and filtering of candidate variants.

Methods

Library type for WGS was Illumina TruSeq PCR-free with a read length of 350 bp. Samples were sequenced on HiSeqX. Data was aligned to the reference (human hg19) using a Burrows Wheeler Aligner (BWA). Samtools was used to sort index and assess mapping statistics. The Genome Analysis Toolkit GATK best practice was used for realignment and recalibration. SNVs and Indels were called using the GATK tool HaplotypeCaller. Called variants that passed the GATK quality filtering was further filtered against 1000 Genomes and SweGen in order to remove variants with an alternative allele frequency >0.05. Remaining variants were annotated with the knownGene database using the Annovar tool.

References

1 Merikangas KR et al. Arch Gen Psychiatry. 2011;68(3):241-51.

2 Miller C et al. Curr Psychiatry Rep. 2014;16(11):499.

3 Vizlin-Hodzic D et al. Translational psychiatry. 2017;7(1):e1010.

4 Chang Jet al. J Cell Sci. 2009;122(Pt 13):2274-82.

5 Cohen S et al. Biochim Biophys Acta. 2013;1833(5):1104-13.

6 Schmid-Burgk JL etal.J Biol Chem. 2016;291(1):103-9.

Annovar was further used to annotate exonic variants with functional predictions from the following tools: SIFT, PolyPhen2-HDIV, LRT, MutationTaster, MetaLR, FATHMM and Radial-SVM. Genotyping of candidates were performed using iPLEX Sequenom MassARRAY platform. The tool PLINK was used for Chi-square calculations to test for association between SNV allele and phenotype.

Results

RNA-seq

Our previous results from RNA-seq showed the *NLRP2* gene to be the most differentially expressed gene (Fig 1).

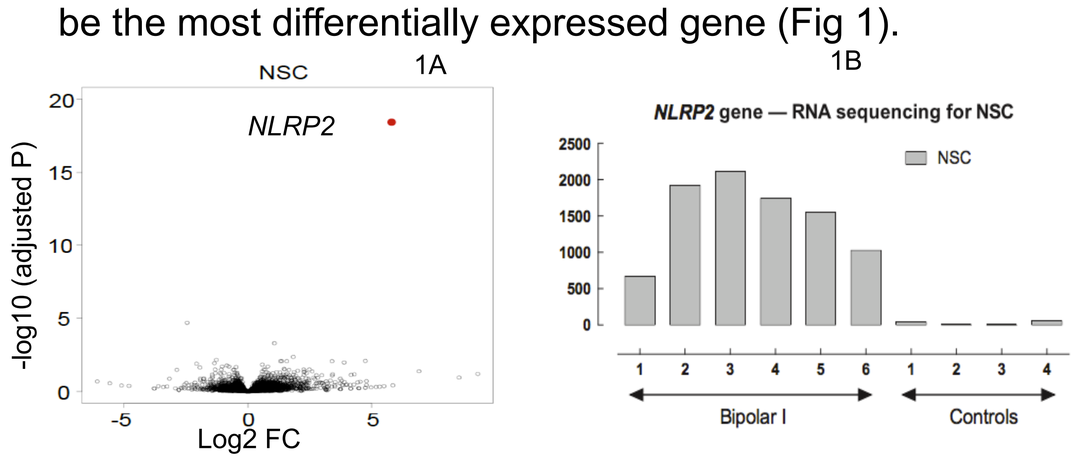
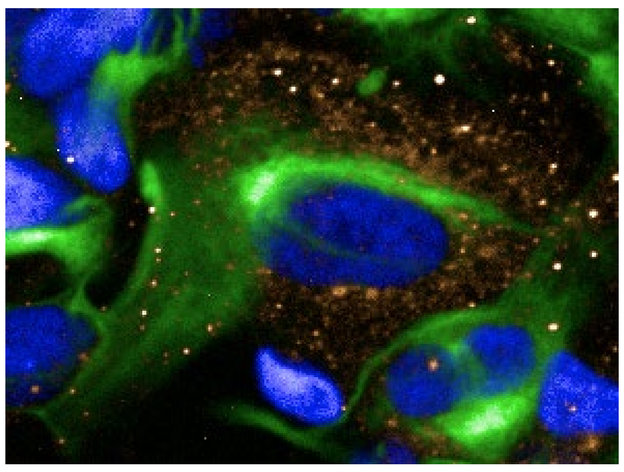


Fig 1. RNA-seq analyses identified BD differentially expressed genes in Neural Stem Cells (NSC). 1A) Volcano plots of log10 adjusted P-values versus the log2 fold change between six BD patients and four controls. 1B) Altered *NLRP2* gene expression in six BD patients and four controls. Y-axis shows normalised gene-counts.

WGS, Genotyping and DbGAP

In total, 120 candidate variants were selected after WGS to be genotyped in a larger cohort. Among the most associated variants, four genes were the most promising as being involved in disease development: *ANK3* that has been proposed in several studies, and the novel findings *NEK3, NEK7* and *TUBB1*. In the dbGAP dataset, a few additional novel variants were identified in *ANK3* and the same variant in *NEK7,* as putative susceptibility variants. NEK3 and NEK7 have been shown to affect microtubule acetylation and microtubule dynamic instability [4,5]. Interestingly, ANK3 has also been shown to increase microtubule dynamics, suggesting dysfunctional microtubule as involved in BD development. Furthermore NEK7 has previously been shown to activate the NLRP3 (relative NLRP2) inflammasome [6]. A possibility of a similar activation of NLRP2 will be explored.

NLRP2 Protein Expression in NSC



It was confirmed with immunocytochemistry that the protein NLRP2 is expressed in NSC (Fig 3). We plan to investigate in more detail how the expression of NLRP2 can be influenced by certain genetic variants.

Fig 3. NLRP2 protein expression in NSC NESTIN / NLRP2 / DAPI

This project was financially supported by the Swedish Foundation for Strategic Research.

Funding support for the dbGAP dataset was provided by the National Institute of Mental Health (NIMH) through grant R01MH094483 and R01MH078151. These datasets were provided by Dr. John Kelsoe from University of California, San Diego.

Correspondence: Katarina Truvé, Bioinformatics Core Facility, Sahlgrenska Academy at University of Gothenburg, email: katarina.truve@gu.se

The Sahlgrenska Academy

UNIVERSITY OF GOTHENBURG