Coding and non-coding RNA dysregulation in bipolar disorder

Jurjen J. Luykx*[1,2,3#], F. Giuliani*[1], G. Giuliani[1], J.H. Veldink[1,4]

[1] Department of Translational Neuroscience, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands
[2] Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht (UMCU), Utrecht University, Utrecht, the Netherlands
[3] Department of Psychiatry, ZNA Hospitals, Antwerp, Belgium
[4] Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht (UMCU), Utrecht University, Utrecht, the Netherlands

* These authors equally contributed to this work.

To whom correspondence should be addressed at:
Jurjen J Luykx | Brain Center Rudolf Magnus | University Medical Center Utrecht, Universiteitsweg 100, office 4.127 (Stratenum), HP 4.205 | 3584 CG Utrecht | The Netherlands | +31 (0)88 75 68638 | j.luykx@umcutrecht.nl

Supplemental Methods

RNA-Seq library preparation and sequencing

Stranded, paired-end sequencing libraries were prepared by the ServiceXS sequencing facility (BaseClear, Leiden, The Netherlands) using the NEBNext Ultra Directional RNA Library Prep kit (Illumina), after which samples were subjected to quality control, cluster generation and sequencing on the Illumina HiSeq 2500 platform. Stranded libraries were chosen as strand orientation information offers a better resolution for regions with overlapping transcription from opposite directions.

Read mapping

Following trimming of low quality bases and adapter sequences using FASTQ-MCF version 0.0.13, processed reads were mapped to the hg19/GCh37 reference human genome (iGenomes) with TopHat2 (version 2.0.13)(1), using the 'fr-firststrand' option for strand orientation to generate BAM-formatted genomic coordinates. Mapped counts were measured for each gene using the python script htseq-count (2). For lncRNA alignment, reads were aligned using Tophat2 aligner against the NONCODEv4 (3) reference database that contains 54,073 human annotated lncRNA sequences.

WGCNA analysis

Weighted gene co-expression network analysis (WGCNA) was constructed from variance-normalized RNA-seq count data using R and the WGCNA package (4). Following the removal of genes with low variance to remove noise, a dissimilarity matrix that measures the level of concordance between BPD and control groups across the samples was generated based on a matrix of pairwise Pearson's correlation coefficients. Then the dissimilarity matrix was transformed into an adjacency matrix by raising the correlation matrix to a power β, chosen based on a fit to scale-free topology and a threshold of 0.9. Modules were finally defined as branches of the resulting clustering tree. To cut the branches, we used hybrid dynamic tree-cutting because it produces robustly defined modules. To obtain moderately large and distinct modules, we set the minimum module size to 50 genes and the minimum height for merging modules at 0.1.

Validation and replication of transcriptome changes by RT-PCR and RT-qPCR

500 ng of total RNA reaction was reverse transcribed using the SuperScript III First-Strand Synthesis System (ThermoFisher Scientific) with random hexamers as described by the manufacturer. Real-time RT–PCR (RT-qPCR) for quantification of mRNA amount was conducted

using iQ™ SYBR® Green Supermix (Bio-rad) for all samples, in triplicate, with a QuantStudio™ 6 Flex Real-Time PCR System (ThermoFisher Scientific). Real-time reaction was carried out as follows: pre-denaturing at 95°C for 10 min, followed by 40 cycles at 95°C for 15 s and 60°C for 1 min. Melting curves were generated for the final PCR products by decreasing the temperature to 60°C for 1 min followed by an increase in temperature to 95°C. Real-time PCR MyIQ software (Bio-rad) was used to determine the amplification cycle in which product accumulation was above the threshold cycle values (CT). Relative quantification was determined using the $2^{-\Delta CT}$ method (5) and normalized to the endogenous control GAPDH. Unpaired Welch t-test was used for multiple comparisons with a 95% confidence cutoff. We considered the difference between the two groups to be significant when $p < 0.05$ (*$p < 0.05$, **$p < 0.01$, ***$p < 0.005$, ****$p < 0.001$). The primer sequences used for RT–qPCR reactions were the following: ABCA1_forward TCCTGATCTCTGTTCGGCTG, ABCA1_reverse GGGTAACGGAAACAGGGGTT, BCL6B_forward CAGAGCACACAAGGCAGTTC, BCL6B_reverse TGAAGTCCAATAGAGGGGCGA, CD93_forward ATGCGGCAGACAGTTACTCC, CD93_reverse AAAGCTCTGAGGATGGTGGC, PODXL_forward TACCCTGCCAGAGACCATGA, PODXL_reverse TCCTGTGAGGTTCAGGACGA, SOCS3_forward GTCACCCACAGCAAGTTTCC, SOCS3_reverse GGTCACTGCGCTCCAGTAG, CREG1_forward TCACAACACAAGTCTCACGA, CREG1_reverse CCTCTACTGGCTTCTTCACT, LARP7_forward TTGCCAGAGCATTGAGAAGT, LARP7_reverse TGCGTTCATCCTCATCCTTT, GAPDH_forward TGGAAGGACTCATGACCACA, GAPDH_reverse GGGATGATCTTCTGGAGAGC.

References

1. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14:R36.
2. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31:166-169.
3. Zhao Y, Yuan J, Chen R. NONCODEv4: Annotation of Noncoding RNAs with Emphasis on Long Noncoding RNAs. Methods Mol Biol. 2016;1402:243-254.
4. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
5. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25:402-408.