**Supplemental Information:**

**The role of microRNA expression in cortical development during conversion to psychosis**

Amanda B. Zheutlin, M.S., Clark D. Jeffries, Ph.D., Diana O. Perkins, M.D., Yoonho Chung, M.S., Adam M. Chekroud, M.S., Jean Addington, Ph.D., Carrie E. Bearden, Ph.D., Kristin S. Cadenhead, M.D., Barbara A. Cornblatt, Ph.D., Daniel H. Mathalon, Ph.D., M.D., Thomas H. McGlashan, M.D., Larry J. Seidman, Ph.D., Elaine F., Walker, Ph.D., Scott W. Woods, M.D., Ming Tsuang, M.D., Ph.D., & Tyrone D. Cannon, Ph.D.

Table S1

Symptom, Medication, and Substance Use Information

Supplementary Materials and Methods

Subject Recruitment

RNA Sequencing

Supplementary Results

miRNA Classifier Related to Other Cortical Regions of Interest

Pathway Analysis Using Whole Genome Reference Set

Figure S1

**Table S1**

*Symptom, Medication, and Substance Use Information*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | CHR, Non-Converter (N = 34) | CHR, Converter  (N = 13) | Control  (N = 27) | *p*:  CHR Contrast |
| SOPS Scores |  |  |  |  |
| Total | 36.8 (7.6) | 42.5 (11.3) | 4.9 (5.5) | .164 |
| Positive | 12.4 (4.4) | 14.0 (3.2) | 1.4 (1.9) | .245 |
| Negative | 11.7 (6.0) | 12.5 (6.0) | 1.4 (1.9) | .682 |
| Disorganized | 5.0 (2.7) | 6.4 (3.4) | 0.8 (1.1) | .155 |
| General | 7.6 (4.4) | 9.5 (4.4) | 1.3 (1.8) | .191 |
| Medications |  |  |  |  |
| Antipsychotic | 10 (29%) | 3 (23%) | 1 (4%) | .944 |
| Antidepressant | 12 (35%) | 7 (54%) | 1 (4%) | .408 |
| Stimulant | 7 (21%) | 3 (23%) | 0 | 1 |
| Mood stabilizer | 1 (3%) | 1 (8%) | 0 | 1 |
| Benzodiazepine | 3 (9%) | 2 (15%) | 0 | .902 |
| NSAID | 1 (3%) | 1 (8%) | 1 (4%) | 1 |
| Antihistamine | 3 (9%) | 0 | 3 (11%) | .660 |
| Sedative | 2 (6%) | 2 (15%) | 2 (7%) | .646 |
| Other | 4 (12%) | 3 (23%) | 9 (33%) | .606 |
| No medications | 21 (62%) | 8 (62%) | 14 (52%) | 1 |

Group means and standard deviations were reported for symptom sub-scales. Number and percent of subjects who reported taking medications in each category are listed. One-way analysis of variance tests (SOPS) or chi-square tests (medications) comparing only the CHR groups were run. CHR = clinical high risk; SOPS = Structured Interview for Prodromal Syndromes.

**Supplementary Materials and Methods**

*Subject Recruitment*

The CHR cases met Structured Interview for Prodromal Syndromes/Scale of Prodromal Symptoms criteria for a psychosis risk syndrome(McGlashan *et al*, 2010), excluding individuals who had ever met DSM-IV criteria for a psychotic disorder. Control participants were excluded if they met criteria for a psychotic disorder, had a first-degree relative with a current or past psychotic disorder, or met prodromal criteria. General exclusions included substance dependence, neurologic disorder, or IQ estimates < 70. All participants provided written informed consent; a parent or guardian consented for subjects under 18 years of age. The study was approved by the Institutional Review Board at each site. Excluded subjects were not more likely to be CHR cases (*N=*23) than expected by chance given the overall sample group distribution, *χ2*=2.31, *p=*.129, though numerically outnumbered excluded healthy control subjects (*N=*3).

*RNA Sequencing*

Briefly, leukocytes were isolated immediately after phlebotomy, RNA was extracted using a modified LeukoLOCK procedure (Life Technologies, Foster City, CA), and small RNA libraries were prepared with Illumina TruSeq kits (San Diego, CA) following the manufacturer’s protocol. Each of 2,588 mature miRNA sequences from miRBase v21 was sought as an exact sequence match within each read. From an initial set of 101 samples, we excluded six (three unaffected, three CHR) with low abundance reads, leaving 27 unaffected and 68 CHR subjects (*N* = 95). For all analyses, we included the 136 miRNAs that were robustly expressed, defined as 10,000 or more total reads in the 95 subjects.

RNA sequencing can be used to quantify the absolute number of any transcription sequence, including those employed here: canonical mature miRNA transcripts. However, raw values can vary considerably between samples (i.e., differences in library size), and thus it is best practice to normalize these values before comparisons across samples (Dillies *et al*, 2013). There are a variety of methods used towards this end, most of which aim to normalize library size via one of several simple adjustments, including scaling each gene in each sample based on the dataset mean, upper quartile, or median for that gene, or gene-specific mean / median ratio-based adjustments. Details of these methods are available elsewhere and have been compared empirically (Dillies *et al*, 2013). In general, the ratio-based methods, which adjust read counts for genes according to the magnitude of their differential expression, appear to perform the best (Dillies *et al*, 2013).

Normalization for our dataset was developed and described by Jeffries et al. (Jeffries *et al*, n.d.), summarized as follows. For our dataset, first, we noted miR-485-5p accounted for 62% of the grand total of 3.63E8 miRNA reads. Since we sought informative sets of miRNAs, the overwhelming miR-485-5p levels were discarded, leaving 136 robustly expressed miRNAs for further analyses. We divided read counts for each sample by the average of read counts for the top 30 miRNAs, forming quotients. This flattened the miRNA numbers and made the ratios maximum:minimum counts among the top 30 miRNAs less extreme. For each miRNA, we then used the average and standard deviation over all unaffected controls of those quotients to convert all quotients to z-scores; final values were in a 4X range.

Technical duplicates of one non-converter sample were assayed twice. The original tube of blood was divided into two equal aliquots and each portion was subsequently treated as an independent sample without awareness of the assaying laboratory. After normalization as described above, correlation of the duplicates over all 136 miRNAs was 0.61; this was at the 98th percentile of correlations of all 4465 possible pairs of 95 samples. Correlations >0.61 in 136-dimensional space would be very unlikely among random, normally distributed vectors; p-value would be 1.60E-15. Over non-converters only, this correlation was at 97th percentile. Given this correlation between duplicate samples, we ascertained the normalization process employed here was successful.

**Supplementary Results**

*miRNA Classifier Related to Other Cortical Regions of Interest*

We previously showed that clinical high risk subjects who converted to psychosis showed a steeper annualized rate of reduction in cortical thickness in the right superior frontal, middle frontal, and medial orbitofrontal regions relative to those who did not convert, accounting for baseline cortical thickness, gender, site, and multiple comparisons using a false-discovery rate correction (Cannon *et al*, 2015). Our primary analyses focused on the region of interest (ROI) with the most pronounced effect, but we opted to test the relationship between the miRNA classifier developed for this ROI with the other two significant ROIs identified previously, as well. This can help establish the specificity of this classifier to some degree, although rates of cortical thinning across these three regions were highly correlated (*r*’s between .47 and .74). In any case, we entered the miRNA classifier for superior frontal cortical thickness as a predictor, along with age, sex, and diagnostic group (converters vs. non-converters and controls), into linear regressions of the other two ROIs.

The classifier developed for the superior frontal cortical ROI was significantly related to cortical thinning the middle frontal region of the cortex, *t*=2.72, *p=*.008, *R2*=.097. It was not significantly related to cortical thinning in the medial orbitofrontal cortex, *t=*1.57, *p=*.121. Spatially, the superior frontal cortex is closer to the middle frontal than the medial orbitofrontal cortex, which may imply a higher degree of similarity with respect to biological substrates; these two phenotypes were also more strongly correlated within individuals (*r=*.71) than the superior frontal and orbitofrontal cortex rates were (*r*=.47).

*Pathway Analysis Using Whole Genome Reference Set*

Ingenuity Pathway Analysis was employed as described in Methods: Functional Annotation using two different reference sets, 1) genes expressed in immune tissue, 2) all genes. The results from the first are reported in the main text; the results from the second are extremely similar, but include many more significant pathways due to the increased denominator. The full list includes in order of significance: protein kinase A signaling, molecular mechanisms of cancer, factors promoting cardiogenesis in vertebrates, HGF signaling, ERK5 signaling, axonal guidance signaling, regulation of epithelial-mesenchymal transition pathway, PPARa/RXRa activation, germ cell-sertoli cell junction signaling, p70S6K signaling, WNT/β-catenin signaling, GNRH signaling, HIPPO signaling, reelin signaling in neurons, ErbB signaling, SAPK/JNK signaling, dopamine-DARPP32 feedback in cAMP signaling, mouse embryonic stem cell pluripotency, neuregulin signaling, role of tissue factor in cancer, NGF signaling, role of NFAT in cardiac hypertrophy, B cell receptor signaling, macropinocytosis signaling, synaptic long term potentiation, HER-2 signaling in breast cancer, epithelial adherens junction signaling, FGF signaling, ErbB4 signaling, ERK/MAPK signaling, superpathway of inositol phosphate compounds, prolactin signaling, gap junction signaling, cholecystokinin/gastrin-mediated signaling, ephrin receptor signaling, IGF-1 signaling, cardiac hypertrophy signaling, production of nitric oxide and reactive oxygen species in macrophages, neuropathic pain signaling in dorsal horn neurons, growth hormone signaling, tec kinase signaling, pyridoxal 5’-phosphate salvage pathway, NF-kB activation by viruses, FAK signaling, break cancer regulation by stathmin1, paxillin signaling, ILK signaling, UVB-induced MAPK signaling, virus entry via endocytic pathways, UVA-induced MAPK signaling, thrombin signaling, P13K/AKT signaling, CREB signaling in neurons, 3-phosphoinositide biosynthesis, PAK signaling, P2Y purigenic receptor signaling pathway, myc medicated apoptosis signaling, neurotrophin/TRK signaling, actin cytoskeleton signaling, aldosterone signaling in epithelial cells, Huntington’s Disease signaling, 14-3-3-mediated signaling, melatonin signaling, renin-anglotensin signaling, glioma signaling, thrombopoletin signaling, integrin signaling, human embryonic stem cell pluripotency, PTEN signaling, rac signaling, FLT3 signaling in hematopoietic progenitor cells, type II diabetes mellitus signaling, Wnt/Ca+ pathway, melanocyte development and pigmentation signaling, xenobiotic metabolism signaling, phagosome formation, AMPK signaling, CNTF signaling, IL-3 signaling, G beta gamma signaling, clathrin-mediated endocytosis signaling, corticotropin releasing hormone signaling, cardiac β-adrenergic signaling, CXCR4 signaling, ephrin A signaling, circadian rhythm signaling, p53 signaling, VEGF signaling, insulin receptor signaling, erythropoletin signaling, non-small cell lunch cancer signaling, hypoxia signaling in the cardiovascular system, synaptic long term depression, calcium signaling, endothelin-1 signaling, sertoli cell-sertoli cell junction signaling, RAR activation, RAN signaling, salvage pathways on pyrimidine ribonucleotides, UVC-induced MAPK signaling, STAT3 pathway, signaling by rho family GTPases, regulation of cellular mechanics by calpain protease, CCR3 signaling in eosinophils, prostate cancer signaling, CDK5 signaling, PI3K signaling in B lymphocytes, leukocyte extravasation signaling, cell cytle regulation by BTG family proteins, glycerol-3-phosphate shuttle, eNOS signaling, role of NANOG in mammalian embryonic stem cell pluripotency, PKC8 signaling in T lymphocytes, LPS-stimulated MAPK signaling, colorectal cancer metastasis signaling, role of osteoblasts, osteoclasts and chondrocytes in rheumatoid arthritis, phospholipase C signaling, VEGF family ligand-receptor interactions, mTOR signaling, D-myo-inositol-5-phosphate metabolism, agrin interactions at neuromuscular junction, rhoGDI signaling, fcyRIIB signaling in B lymphocytes, sphingosine-1-phosphate signaling, G protein signaling mediated by tubby, GPCR-mediated nutrient sensing in enteroendocrine cells, regulation of IL-2 expression in activated and anergic T lymphocytes, inositol pyrophosphates biosynthesis, UDP-N-acetyl-D-glucosamine biosynthesis II, erbB2-erbB3 signaling, leptin signaling in obesity, BMP signaling pathway, and mechanisms of viral exit in host cells.

*bootstraps.pdf*

*Figure S1. Bootstrapping approach to miRNA classifier validation.* miRNAs chosen more frequently in subsets of the data are more likely to generalize to an external sample, as they are less dependent on the exact structure of the training sample. As such, we quantified frequency for each miRNA across 2000 random subsets containing 80% of subjects (Figure 1B). However, as this logic also holds true for bootstrapped samples, we aimed to verify these findings by additionally quantifying the frequency for each miRNA across 2000 bootstrapped samples (i.e., samples of the same sample size as the true data that were randomly selected with replacement). We found highly substantively similar results leading to the same conclusion (i.e., exclusion of the same one miRNA from the original classifier) using this approach. The 25 most frequently selected miRNAs identified across analyses were displayed here and colored according to weather or not they were selected in the model using all data.

**References**

Cannon TD, Chung Y, He G, Sun D, Jacobson A, Erp TGM Van, *et al* (2015). Progressive reduction in cortical thickness as psychosis develops: A multisite longitudinal neuroimaging study of youth at elevated clinical risk. *Biol Psychiatry* **77**: 147–157.

Dillies MA, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, *et al* (2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform* **14**: 671–683.

Jeffries CD, Perkins DO, Chandler SD, Stark T, Addington J, Bearden CE, *et al* Insights into Psychosis Risk from Leukocyte microRNA Expression. *Transl Psychiatry* .

McGlashan TH, Walsh BC, Woods SW (Oxford University Press: Oxford, 2010). *The Psychosis-Risk Syndrome: Handbook for Diagnosis and Follow-up*. .