Supplemental Methods & Materials

TOP/NORMENT Sample
The TOP study is part of a large ongoing study on psychiatric disorders – Norwegian Centre for Mental Disorders Research (NORMENT), previously Thematically Organized Psychosis (TOP). It consists of Northern European Caucasians, mainly Norwegians, which have previously been demonstrated to be genetically homogenous (1; 2). To be included in the study, patients had to fulfill the criteria for a DSM-IV diagnosis of a psychotic spectrum disorder, be 18 years or older at inclusion and be willing and able to provide written informed consent. In the healthy control group, subjects were excluded if they or their close relatives had a lifetime history of a severe psychiatric disorder (schizophrenia, bipolar disorder, and major depressive disorder).

Diagnostic evaluation was performed by trained psychologists and psychiatrists, all of whom participated regularly in diagnostic meetings supervised by professors in psychiatry. Reliability measures of the diagnostic assessment in the study were performed, and the overall agreement for the DSM-IV diagnostic categories tested was 82% and the overall kappa 0.77 (95% CI: 0.60–0.94) (3).

The patients were divided into three broad spectrums according to DSM-IV diagnoses: bipolar spectrum disorders (subdivided into BD1, BD2 and BDNOS), schizophrenia spectrum disorders (subdivided into SCZ, SCZA, SCZF) and mixed (consisting of psychosis not otherwise specified and major depressive disorder with psychosis).

The healthy control subjects came from the same catchment area as the patient group and were selected randomly from the national statistics records (www.ssb.no), and they all underwent an initial interview where demographic and clinical information was obtained. A history of a medical condition potentially interfering with brain function (hypothyroidism, uncontrolled hypertension, and diabetes) or an illicit drug abuse/addiction diagnosis was also exclusion criteria. Clinical assessment of the patients and healthy controls participating in this study is described in detail in previous reports (4).
TOP/NORMENT Sample - Exome Chip Genotyping and Quality Control
A total of 1261 TOP samples were genotyped on the Illumina HumanExome Beadchips v1.0 at the University of Miami, Hussman Institute for Human Genomics (HIHG), USA. Following sample QC (Quantitation and Qualitation Control), the samples were arranged in 96 well plates for preparation for genotyping with one HIHG control sample per plate to evaluate run-to-run genotype concordance and experimental reproducibility. Genotype reproducibility for all HIHG controls was calculated to 99.99-100%. In total 1371 samples, divided on 15 HIHG controls, 95 Hapmap controls and 1261 samples from TOP were genotyped on the Illumina HumanExome Beadchip v1.0. The University of Miami did not filter any SNP based on technical quality control metrics.

Genotype calling was performed using Illumina's Genome Studio V2011.1, genotyping module v1.9.4. The Cluster file provided by Illumina (HumanExome-12v1.egt) for the HumanExome-12v1 Beadchip was used to generate genotype calls and analyze data in Illumina's GenomeStudio software. 247,870 SNPs were present on the chip and 101,784 SNPs were not monomorphic among the 1261 individuals.

As a first step, individual QC was performed separately on the 1261 samples. Samples with disconcordant gender, excess heterozygosity estimated as more than 5 standard deviations from the mean, intentional duplicates, mixed-up samples (pihat >.55, <.99), one of two good intentional duplicates were removed. IBS was calculated in PLINK and non-European clusters were removed. The Individual missingness filter was set at .98, leaving 1132 individuals with 86,389 non-monomorphic SNPs. Before analysis, a further eight relatives were removed leaving 1124 individuals (794 cases, 330 controls) and 83,129 non-monomorphic SNPs.

Cases can be subdivided into bipolar disorder spectrum (n = 402), schizophrenia spectrum disorder (n = 293), mixed (n = 99) which includes “psychosis not otherwise specified” (n = 73) and major depressive disorder (n = 26).

Swedish Replication Sample - Exome Chip Genotyping and Quality Control
The primary replication sample was predominantly Swedish and exclusively Northern European. Cases from Sweden were collected through two recruitment streams: the Stanley study and the St. Göran Bipolar Project. Most cases were recruited from the Stanley study that identified cases through the Swedish Quality Register for BD (BipoläR), which has been described in detail previously (5; 6). In brief, BipoläR contains individualized data on diagnoses (i.e., BD type 1, type 2, not otherwise specified, or schizoaffective disorder bipolar type according to the DSM-IV-TR), medical intervention, and outcomes. It also captures basic clinical epidemiological data as well as longitudinal data on the natural history and clinical course of the disease. Genotyping has been completed for 2001 BD patients.
Hughes et al.

(passing quality control) enrolled in the Stanley study. The other recruitment stream for cases was from the St. Göran Bipolar Project (n = 285), which provides assessment, treatment, and follow-up of BD patients within the Northern Stockholm Mental Health Service and the Affective Unit in Mölndal. The work-up procedures have been previously described in detail (7–9). Both projects were approved by the Regional Ethical Review Board in Stockholm (Sweden) and all participants provided written informed consent. Control subjects (n = 5359) were randomly selected from Swedish population registers, ascertained on a national basis. The exclusion criterion was any hospitalization for SCZ or BD. DNA collection procedures have been previously described (10). A small number of controls (n = 55) were from St. Göran Bipolar Project as described previously (7).

Cases and controls were genotyped across four waves on the Illumina HumanExome Beadchip at the Broad Institute of MIT and Harvard. Variant calling was performed using Beadstudio. The genotyping waves were first merged to create a single dataset. Samples and SNPs with >5% missingness were excluded. Samples with ambiguous genetic sex, outlier levels of heterozygosity, population outliers (non-Northern European), and one of each pair of duplicate or highly related individuals (pi-hat > .2) were removed. Markers with no variation (monomorphs), those out of Hardy-Weinberg equilibrium (p < 1 x 10^-6), or manifesting differential missingness rates between cases and controls (p < 1 x 10^-6) were removed.

Following QC steps, 2286 BD cases and 5414 controls remained. Approximately half of the cases had available subtype information, and there are 552 BD type 1 and 605 BD type 2 cases.

Norwegian Replication Sample – TaqMan Genotyping and Association Analysis

An additional 259 TOP/NORMENT samples, also of Norwegian ethnicity, were used as an independent replication sample and were genotyped for rs41283526 (CTR: 39, BD: 102, SCZ: 118) using a TaqMan® SNP Genotyping Assay, with assay ID C__86344586_10 (Life Technologies Corporation, Carlsbad, CA, USA). Briefly, 30 ng of genomic DNA from each sample was amplified in 5 µl total volume in a single 384 well plate, using TaqMan® Genotyping Master Mix (Life Technologies Corporation, Carlsbad, CA, USA) and assay mix according to the manufacturer’s protocol on the QuantStudio 12K Flex Real-time PCR system (Life Technologies Corporation, Carlsbad, CA, USA).

We used this small sample as a second replication test which we performed in PLINK. In this sample, rs41283526 occurs with an allele frequency of 2.6% in controls (which is similar to the frequency in the Norwegian exome chip dataset), but it is not seen in any of the 118 SCZ or 102 BD cases. Thus, for all diagnostic groups, we obtained an odds ratio of zero that is significant for all diagnostic groups (see Table S1).
Table S1. Allelic association tests for Norwegian replication sample performed in PLINK

<table>
<thead>
<tr>
<th>Norwegian TaqMan Replication Sample</th>
<th>CTR</th>
<th>BD</th>
<th>SCZ</th>
<th>BD &amp; SCZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of samples</td>
<td>39</td>
<td>102</td>
<td>118</td>
<td>220</td>
</tr>
<tr>
<td>Allele frequency</td>
<td>0.026</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( p )-value</td>
<td>-</td>
<td>2.2E-02(^b)</td>
<td>1.4E-02(^b)</td>
<td>7.6E-04(^a)</td>
</tr>
<tr>
<td>OR</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\(^a\) significant at 1%.
\(^b\) significant at 5% (standard cut-off for replication of a single SNP).

RNA Extraction and cDNA Generation

Blood samples for gene expression analysis were collected using Tempus Blood RNA Tubes (Life Technologies Corporation, Carlsbad, CA, USA), and stored at -80°C until analysis. Total RNA was extracted either automated with the ABI PRISM 6100 Nucleic Acid PrepStation (Life Technologies Corporation, Carlsbad, CA, USA) and the Tempus 12-port RNA Isolation Kit or manually with the Tempus Spin RNA Isolation Kit (Life Technologies Corporation, Carlsbad, CA, USA) according to the manufacturer’s protocol. RNA quantity was determined using a NanoDrop 8000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA).

cDNA was generated from 1 µg of total RNA, using the High-Capacity cDNA Reverse Transcription Kit (Life Technologies Corporation, Carlsbad, CA, USA) according to the manufacturer’s protocol.

Quantitative RT-qPCR of Whole Blood cDNA

We designed two sets of PCR primers, one pair spanning the junction between the “little” exon and the 3’ constitutional exon which is potentially affected by the splice site (5’-TGCACATGGTACAAATTTCCCA-3’ and 5’-TGCAGGCAGGGAATCATCAC-3’, see Figure 3) and one pair further upstream spanning a junction common to all three major RefSeq isoforms (5’-GCTTGATCGAGCTGCCCA-3’ and 5’-GCTGGAGGAGAAGCTGGACGC-3’).

Quantitative RT-PCR was performed on the ABI PRISM 7900HT Real-Time PCR System (Life Technologies Corporation, Carlsbad, CA, USA), using a SYBR green assay. Each sample was assayed in triplicate with two different ANK3 “little” exon specific primer sets, and normalized to the gene expression of a primer pair upstream in the ANK3 transcripts, and to the expression of the glucuronidase, beta gene (\(GUSB\)). Briefly, 10 ng of cDNA was used as input in each PCR reaction, which consisted of 1X Power SYBR Green Master Mix (Life Technologies Corporation, Carlsbad, CA, USA), 1 µM of the ANK3 “little” exon specific primers, or .3 µM of the upstream primers, or 1 µM \(GUSB\) primers. After an initial heating step at 95°C for 10 min to activate the polymerase, 40 PCR cycles were performed. Each cycle consisted of a denaturation step at 95°C for 15 s, followed by a
combined annealing and extension step at 60°C for 1 min. A melt curve analysis was performed after each PCR, to assess the specificity of the assay. Using qBasePLUS software (v2.6), the delta-delta-Cq model was used to determine relative target gene expression (11).

**Fragment Analysis of Brain and Blood cDNA**

We designed primers in the “5’ constitutional” exon (5’-FAM-ACTGTCACAGAGAAGCACAA-3’) and in the “3’ constitutional” exon (5’-TGCAGGCAGGGAATCATCAC-3’). PCR amplification was performed in 10 µl volume, using 10 ng of cDNA as input. Each PCR reaction consisted of .25 U of AmpliTaq Gold 360 (Life Technologies Corporation, Carlsbad, CA, USA), 1.5 mM MgCl₂, .2 mM each dNTP, and 1 µM of each primer. After an initial heating step at 95°C for 10 min to activate the polymerase, 25 PCR cycles were performed. Each cycle consisted of a denaturation step at 95°C for 30 s, an annealing step at 60°C for 30 s, followed by an extension step at 72°C for 30 s. After cycling, there was a final extension step at 72°C for 7 min.

The PCR amplifications were performed on one human brain sample (Catalog number: 636561 - Normal human brain cerebral cortex total RNA pooled from 3 male Asians, ages: 26-41) obtained from Clontech (Mountain View, CA, USA). In addition, we also amplified four whole blood TOP samples (individuals not carrying the splice site variant) in order to estimate the abundance of different isoforms in blood.

One microliter of the amplified products were mixed with 12 µl Hi-Di Formamide (Applied Biosystems) and .2 µl Genescan 500 LIZ dye Size Standard (Applied Biosystems), and subjected to capillary electrophoresis on the Applied Biosystems 3730-xl DNA Analyzer. The result files were analyzed in the GeneMapper 5 software (Applied Biosystems), see Figure S2.
**Figure S1.** Screenshot of IGV displaying the RNAseq read alignments of one sample of the Lister *et al.* dataset (12). Bottom pane shows the exonic structure of the different RefSeq isoforms of ANK3 in the regions surrounding the little exon. Middle pane displays the read alignments. Top pane displays in blue the exon junctions (junctions are laid out on different lines without regard for which transcript they belong to) and in gray bars the coverage of the different exons. The alternative splicing on the “medium” exon is clearly visible (middle pane) as well as the low frequency of the “little” exon isoform (middle and top pane).
Figure S2. Fragment analysis of labeled PCR product spanning “little” and “medium” exons. 10 ng cDNA input, 25 cycles, 10 s injection. Top four panels: cDNA derived from blood total RNA, four individuals from TOP sample not carrying rs41283526 (1 CTR, 1 BD, 1 MIX, 1 SCZ). Bottom panel: cDNA derived from human brain cerebral cortex total RNA (pooled from 3 Asian males, ages: 26-41). Exonic structure thumbnails match Figure 3 in main text.

There are two notable differences between blood and brain: 1) expression in brain of all isoforms is several orders of magnitude higher than in blood and 2) there is clear evidence of a fourth isoform in our blood data where the “little” exon is present between the constitutive exons without the “medium” exon. This fourth isoform has a very low relative abundance in brain (even less than “altSplice-2”), but in blood about the same relative abundance as altSplice-1.
Figure S3. Detail of the “little” exon in the UCSC genome browser (hg19). rs41283526 in bright red. Note: ANK3 is encoded on the reverse strand. Lower dark blue track is the 100 vertebrates basewise conservation by PhyloP.
Figure S4. Heatmap from brainspan.org of the 5’ exons of ANK3 (human developmental transcriptome): from blue (low expression) to red (high expression). For a visualization of all exons, please use brainspan.org. Each color block in the second line from the top represents a different brain structure and, within each of these blocks, the different individuals are plotted in order of increasing age. One can quite clearly see the increase in the expression of the “medium” exon from childhood to adolescence in the vast majority of the brain structures. The change in the expression of the “little” exon in early adolescence is less visible in this plot due to the very low level of expression relative to the other exons (see Figure 5 in the main text for more a more detailed view of this data).
Supplemental References


