SNP Variants at 16p13.11 Clarify the Role of the NDE1/miR-484 Locus in Major Mental Illness in Finland

Sinha, Vishal

2020

Sinha , V , Ortega-Alonso , A , Ukkola-Vuoti , L , Linnaranta , O , Zheutlin , A B , Torniainen-Holm , M , Therman , S , Tuulio-Henriksson , A , Jylhä , P , Kaprio , J , Hovatta , I , Isometsä , E , Cannon , T D , Lönnqvist , J , Paunio , T , Suvisaari , J & Hennah , W 2020 , ' SNP Variants at 16p13.11 Clarify the Role of the NDE1/miR-484 Locus in Major Mental Illness in Finland ' , Schizophrenia bulletin open , vol. 1 , no. 1 , sgaa055 . https://doi.org/10.1093/schizbullopen/sgaa

http://hdl.handle.net/10138/341232
https://doi.org/10.1093/schizbullopen/sgaa055

Schizophrenia Bulletin Open
doi:10.1093/schizbullopen/sgaa055

SNP Variants at 16p13.11 Clarify the Role of the *NDE1*/miR-484 Locus in Major Mental Illness in Finland

Vishal Sinha[1,2,*], Alfredo Ortega-Alonso[1–3], Liisa Ukkola-Vuoti[1,2], Outi Linnaranta[4,5], Amanda B. Zheutlin[6], Minna Torniainen-Holm[2,7], Sebastian Therman[2], Annamari Tuulio-Henriksson[3], Pekka Jylhä[2,8], Jaakko Kaprio[1,9], Iiris Hovatta[3,10], Erkki Isometsä[8], Tyrone D. Cannon[6], Jouko Lönnqvist[2,8], Tiina Paunio[8,11], Jaana Suvisaari[2], and William Hennah[*,1,2,]

[1]Institute for Molecular Medicine FIMM, University of Helsinki, Helsinki, Finland; [2]Department of Public Health Solutions, Mental Health Unit, Finnish Institute for Health and Welfare, Helsinki, Finland; [3]Department of Psychology and Logopedics, Medicum, University of Helsinki, Helsinki, Finland; [4]Bipolar Disorders Clinic, Douglas Mental Health University Institute, Montréal, Canada; [5]Department of Psychiatry, McGill University, Montréal, Canada; [6]Department of Psychology, Yale University, New Haven, CT; [7]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden; [8]Department of Psychiatry, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; [9]Department of Public Health, University of Helsinki, Helsinki, Finland; [10]SleepWell Research Program, Faculty of Medicine, University of Helsinki, Helsinki, Finland; [11]Department of Public Health Solutions, Genomics and Biomarkers Unit, Finnish Institute for Health and Welfare, Helsinki, Finland

*To whom correspondence should be addressed; Institute for Molecular Medicine FIMM, PO Box 20, FI-00014, University of Helsinki, Helsinki, Finland; tel: +358 50 318 3423 e-mail: william.hennah@helsinki.fi

Through copy number variations, the 16p13.11 locus has been consistently linked to mental disorders. This locus contains the *NDE1* gene, which also encodes microRNA-484. Both of them have been highlighted to play a role in the etiology of mental illness. A 4-SNP haplotype spanning this locus has been shown to associate with schizophrenia in Finnish females. Here we set out to identify any functional variations implicated by this haplotype. We used a sequencing and genotyping study design to identify variations of interest in a Finnish familial cohort ascertained for schizophrenia. We identified 295 variants through sequencing, none of which were located directly within microRNA-484. Two variants were observed to associate with schizophrenia in a sex-dependent manner (females only) in the whole schizophrenia familial cohort (rs2242549 *P* = .00044; OR = 1.20, 95% CI 1.03–1.40; rs881803 *P* = .00021; OR = 1.20, 95% CI 1.02–1.40). Both variants were followed up in additional psychiatric cohorts, with neuropsychological traits, and gene expression data, in order to further examine their role. Gene expression data from the familial schizophrenia cohort demonstrated a significant association between rs881803 and 1504 probes (FDR *q* < 0.05). These were significantly enriched for genes that are predicted miR-484 targets (*n* = 54; *P* = .000193), and with probes differentially expressed between the sexes (*n* = 48; *P* = .000187). While both SNPs are eQTLs for *NDE1*, rs881803 is located in a predicted transcription factor binding site. Based on its location and association pattern, we conclude that rs881803 is the prime functional candidate under this locus, affecting the roles of both *NDE1* and *miR-484* in psychiatric disorders.

Keywords: schizophrenia/psychoticdisorders/NDE1/gene expression/miR-484

Introduction

Copy number variations, both deletions and duplications, at the 16p13.11 chromosome locus have consistently been observed in individuals with intellectual disability,[1] developmental delay,[2] autism,[3,4] attention deficit hyperactivity disorder,[5] microcephaly,[6] epilepsy,[7,8] and schizophrenia.[9] In addition to these large-scale genomic aberrations, evidence for the psychiatric relevance of this locus has come from several previous family studies in the Finnish population. A study within a Finnish familial schizophrenia cohort, having identified association at the *DISC1* gene locus,[10] conditioned these 458 families with the DISC1 risk haplotype to assist the identification of additional loci. This provided evidence for linkage on the 16p13 locus (D16S764, LOD = 3.17) in those families that already had the *DISC1* HEP3 risk haplotype.[11] Furthermore, the 16p12 locus had previously been observed to be

linked with bipolar disorder in the Finnish families.[12] As this conditioned linkage was located 0.8Mb from the *NDE1* gene, and the fact that the NDE1 protein-peptide binds the DISC1 protein, it was investigated whether the *NDE1* gene specifically associates with schizophrenia. Using the same schizophrenia family cohort, it was observed that a 4 SNP tag-haplotype in *NDE1* comprising the CGCC alleles of rs4781678, rs2242549, rs881803, and rs2075512 associated with schizophrenia in females, but not in the whole cohort.[11] Additionally, this tag-haplotype was checked for association with the Wechsler Memory Scale-Revised (WMS-R) visual working memory endophenotype in 215 families from this schizophrenia cohort, where analysis of female-only offspring approached the statistical significance level (P = .055).[11] In follow-up studies of these findings, using additional information available within the Finnish familial schizophrenia cohort, a significant interaction was found between high birth weight (>4000 g), and one of the constituent SNPs rs4781678 of the tag-haplotype in increasing risk to schizophrenia.[13] While a study using genome-wide gene expression, demonstrated that the *NDE1* SNP rs2242549 associates with significant expression level differences of 2542 genes, and with early cessation of psychoactive medications metabolized by CYP2C19.[14] In vitro studies indicated that microRNA-484 regulation of CYP2C19 would lead to increased metabolism of these medications.[14] The differentially expressed genes were significantly enriched for predicted targets of the micro RNA miR-484,[14] which is encoded on exon 1 of the main transcript of the *NDE1* gene.

Taken together, the genetic evidence from Finnish families for major mental illnesses not only implicates the 16p13.11 genomic locus, but suggests that either, or both, *NDE1* and miR-484 could be a functional element under this region. Two studies aiming to further understand any functional element at this locus have separately highlighted miR-484 and *NDE1*. In vivo knockout expression changes in miR-484 causes neural progenitor proliferation and differentiation alterations, leading to behavioral changes in mice, specifically hyperactivity, suggesting the importance of microRNA-484 in neurogenesis.[15] On the other hand, a comprehensive study of 16p13.11 microduplications, using induced pluripotent stem cells to study the functional consequences of these mutations, noted neural precursor cell (NPCs) proliferation abnormalities, which could be repeated specifically by overexpression of *NDE1*, exclusive of miR-484.[16]

In the Finnish familial cohorts, the observed associations have been found through the study of SNPs designed to tag the haplotypic structure of *NDE1*.[11] While this helps to narrow down the region of interest, from that which the CNVs implicates,[1–9] it still highlights a broad locus that includes both *NDE1* and mir-484. Thus, the goal of the present study was to discern any functional variant, or variants, at this locus, and to determine if they can help to differentiate the functional elements underlying this locus in major mental illness. This differentiation is assisted by the use of endophenotypes and gene expression array data available in the Finnish cohorts.

Materials and Methods

The three-stage Sequencing and Genotyping Design

In order to fulfill our main aim of identifying potential functional variants at the *NDE1* locus we implemented a three-stage sequencing and replication study design (illustrated in supplementary figure 1). We have previously demonstrated that the design can be used to investigate our data sets in a statistically robust manner.[17] In the first stage, we sequenced 96 individuals from 20 families from the familial schizophrenia cohort (LC4 affected n = 42, unaffected n = 54), representing 79 independent chromosomes. Of these, 3 families comprising 12 individuals (12 independent chromosomes) were specifically included for being polymorphic at the rs2242549 locus and having gene expression data available. These were included as the association to schizophrenia previously observed at this locus demonstrated a sex-dependent pattern, meaning that we have reduced power to detect any level of association in the sequencing and stage 1 of genotyping. Thus, we concluded that the addition of individuals with information previously known to be relevant to this locus would provide other means for identifying variants of potential functional interest. From the sequencing stage, 3 identified variants (chr16_1574307, rs74646346, rs117876551) were selected, as they associated (P < .05) with schizophrenia and are predicted to be located in potential functional regions, according to UCSC genome browser build 19 (supplementary table 2). Additionally, we selected 2 SNPs that associated with the chosen gene expression probes, one from our prior observations (rs2242549),[14] and the other SNP (rs881803) was the only SNP to associate at a higher level than rs2242549, and located in a predicted functional element. All 5 SNPs were genotyped (S1) in a sub-cohort of 301 schizophrenia families (n = 1122) alongside 323 population-based controls. This S1 genotyping aimed to confirm that the variants are not sequencing artifacts, along with replicating any observation of association in an enlarged cohort. Two of the SNPs (chr16_1574307, rs74646346) were identified to actually be monomorphic in this cohort, while another SNP (rs117876551) did not pass our quality control criteria for genotyping. Hence, these 3 SNPs were removed from additional analyses. The 2 SNPs (rs2252549, rs881803) that remained were those selected based on their association to gene expression probes. Thus, even though they continued to display no association to schizophrenia, they were taken forward to the third stage (genotyping stage 2; S2),

where we would have a large enough sample size to test for association even taking into account sex differences. These were to be studied due to our prior observations at this locus.[11] This S2 genotyping consisted of 2733 individuals from multiple cohorts, described in the Study Samples section.

Study Samples

This research includes several Finnish clinical cohorts ascertained for different phenotypes, and includes cohorts of different epidemiological design. These have been described in detail in previous studies,[18–25] and additionally in the supplementary materials. The full description of how these cohorts have been utilized into the three-stage study design for variant identification and association replication used here has recently been published,[17] and is briefly described below.

Psychiatric Cohorts for Genetic Association Analysis. The psychiatric cohorts used here comprise; Finnish familial schizophrenia (SCZ $n = 2818$) and bipolar disorder (BPD $n = 650$) cohorts, twin pairs concordant and discordant for schizophrenia (TwinSCZ $n = 303$), 3 clinical study samples for psychotic disorders (first-episode psychosis $n = 125$; MMPN $n = 449$; HUPC $n = 383$; the diagnoses represented under each cohort are explained in the supplementary methods), a population cohort for anxiety disorders (ANX $n = 823$), and randomly selected population controls exclusive of any neuropsychiatric diagnoses (Controls $n = 1117$). These cohorts altogether contain 7024 individuals, with 6668 genotyped, including 1909 psychiatrically healthy controls. These healthy controls contain the 1117 individuals specifically selected for this study, and 792 individuals that had been included as controls from a number of the individual cohorts being used here (supplementary table 1). In S1 genotyping, 323 of these 1117 population controls were used in the analysis, after genotyping stage 2, all 1909 controls have been used in the analyses of individual cohorts and in the analysis of combined cohorts.

The SCZ and BPD cohorts include family members as affected individuals according to increasingly inclusive liability classes (LC), based on the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV).[26] The inclusion criteria for various LCs (LC 1–4) are explained in the supplementary materials. Other cohorts used a single diagnosis in their analysis, the diagnosis being for which the cohort was originally ascertained (supplementary table 1).

In order to increase the statistical power of the association analysis, we sought to analyze the cohorts jointly. To do this we redefined the diagnosis of each individual to represent if it fell under 2 broad phenotypes: psychotic disorder (affected: $n = 1896$; $n = 1780$ without MMPN), mood disorder (affected: $n = 1227$; $n = 778$ without MMPN), or both (affected: $n = 698$; $n = 295$ without MMPN). The ascertainment criteria for these joint cohorts are explained in supplementary materials. Both the cohorts naturally excluded individuals from the anxiety disorder cohort, where the diagnosis does not fit under either of the broad phenotypes used. Additionally, due to the opposing observations demonstrated in the individual cohort analysis, this analysis was performed both including and excluding the MMPN cases.

To further understand the role of any associated variants in the etiology of psychiatric disorders, we studied them with respect to intermediate traits of both neuropsychological and gene expression origin. The details of these neuropsychological traits have been described previously,[14,17,27] and can be found in the supplementary materials. Briefly, 14 quantitative cognitive endophenotypes[28–34] were available for 919 subjects (SCZ, $n = 811$; BPD, $n = 108$). Due to high levels of intercorrelation,[14,27] these endophenotypes have been grouped into 5 first-order cognitive factors and a second-order general ability factor (g-factor).[17] Moreover, we have computed association between the *NDE1* mutation and gene expression measures derived from the SCZ cohort, and checked replication of these findings in the FEP and TwinSCZ cohorts, and the GTEx database. The probe selection and statistical procedures implemented on our gene expression datasets are described in supplementary materials.

Statistical Methods Used in Association Analysis

The standard statistical methods for our three-stage study design have been described previously,[17] and can be found in the supplementary materials. However, since our prior observation with a *NDE1* haplotype displayed sex-dependent effects by associating with schizophrenia and the visual working memory endophenotype in females only,[11] we wanted to explore here whether such effects may affect the association at our 2 identified variants of interest. For this purpose, in addition to our analysis of these cohorts as a whole, we performed association analysis in sex-separated datasets in the combined SCZ cohort (S1+S2), the other independent psychiatric ascertained cohorts, and the analysis of the cohorts combined to study broad psychosis and mood disorder phenotypes. Two-point linkage and association analysis of the familial (SCZ, BPD), joint (Psychotic disorder, Mood disorder), and twin (TwinSCZ) cohorts were performed using Pseudomarker.[35] In cohorts containing related individuals, odds ratios were calculated using allele counts within the R package questionR.[36] However, calculating the odds ratios using related individuals will lead to a bias in the results. Association within our population-based cohorts of unknown family structure, including ANX, FEP, MMPN, and HUPC were generated using PLINK.[37] The odds ratios in these cohorts of unrelated individuals the PLINK command --ci was used.[37] The

QTDT program[38] was used to test association with our 14 separate cognitive endophenotypes, 5 first-order cognitive factors, and the *g*-factor. Details on model and parameter selection for Pseudomarker, PLINK, and QTDT can be found in supplementary material.

Results

Three-Stage Mutation Identification Process

We carried out targeted genome sequencing of 96 individuals (20 SCZ families, and 79 independent chromosomes, sequencing read depth = 67.97; mean coverage = 69.54). This identified 295 variants over the *NDE1* genomic locus (chr16:15727124-15830120). No variants were identified within a 1237bp (chr16:15736621-15737858) region that contained the miR-484 coding sequence (chr16:15737151-15737229). Nine variants associated with the broadest diagnostic class used within this sub-section of the SCZ cohort (LC4_SCZ $P < .05$), and 3 variants were selected to be followed-up by genotyping, as they were located in areas liable to histone modification according to UCSC bioinformatic prediction, that include methylation, phosphorylation, and acetylation (supplementary table 2). Additionally, to improve our chances to detect variants related to our prior observations at the *NDE1* locus,[11] we specifically selected 12 individuals, from 3 families, that were polymorphic at the rs2242549 locus, and had gene expression data available. This aided in the selection of 2 variants. The original observation (rs2242549), that showed association ($P < .05$) with the 3 gene expression probes predicted to be most significantly altered by this locus,[14] and the only variant (rs881803) with better associations than the original finding, and located in a predicted functional region, a transcription factor binding site (see section The Three-Stage Sequencing and Genotyping Design; supplementary table 2).

To confirm their presence and validate their association with schizophrenia, all 5 variants were first genotyped and checked for association in an enlarged sub-section of the SCZ cohort (SCZ $n = 1122$; control $n = 323$). No variants were associated with schizophrenia at this stage (supplementary table 3a). Two SNPs were actually monomorphic in this cohort, and one failed quality control in genotyping. Due to our prior observations of sex-dependent association, and the small sample sizes in the first 2 stages, the 2 SNPs that associated with gene expression levels were genotyped in the entire cohort regardless of no observable association in the S1 genotyping stage. Sex-dependent association analysis in the whole SCZ cohort demonstrated that these SNPs (rs2242549, rs881803; $r^2 = .75$) significantly associated with schizophrenia in females (LC3_SCZ rs2242549 additive $P = .00044$, OR = 1.20, ±95% CI 1.03–1.40; LC3_SCZ rs881803 additive $P = .00021$, OR = 1.20, ±95% CI 1.02–1.40), but not in males (LC3_SCZ rs2242549 additive $P = 0.67$, OR = 0.92, ±95% CI 0.80–1.07; LC3_SCZ rs881803 additive $P = .12$, OR = 0.87, ±95% CI 0.74–1.01) (table 1, a and b; figure 1, supplementary table 3, supplementary figure 2). No geographical differences were observed in our analyses.

In the studies of the other cohorts, we observed an association between rs881803 and psychosis within the

Table 1. Association Values for the *NDE1* SNPs rs2242549 (T>G) and rs881803 (T>C) Within the Finnish Familial Schizophrenia Cohort (SCZ)

(a) Whole Cohort

SNP (All)	Sequencing LOD *P*-value	Sequencing *P*-value	S1 *P*-value	S2 *P*-value	Combined (S1+S2) *P*-value	Odds Ratio (95% CI)	MAF SCZ Families	MAF Controls
rs2242549	.42	.16	.58	.45	.24	1.02 (0.92–1.14)	0.43	0.43
rs881803	.50	.58	.49	.97	.54	0.99 (0.89–1.10)	0.36	0.36

(b) Separated on Sex

SNP	Females				Males			
	Combined (S1+S2) *P*-value	Odds Ratio (95% CI)	MAF[a] SCZ Families	MAF[a] Controls	Combined (S1+S2) *P*-value	Odds Ratio (95% CI)	MAF[a] SCZ Families	MAF[a] Controls
rs2242549	**.00044**	**1.20 (1.03–1.40)**	0.46	0.44	.67	0.92 (0.80–1.07)	0.41	0.42
rs881803	**.00021**	**1.20 (1.02–1.40)**	0.39	0.37	.12	0.87 (0.74–1.01)	0.33	0.35

Note: *P*-values, odds ratios, and confidence intervals across (a) the whole cohort across the three-stage design and (b) separated by sex in the combined (S1+S2) cohort. Values are shown for liability class 3 under an additive genetic model. Values for other models and classes are provided in supplementary table 3.

[a]Minor allele frequency is calculated for the founders in the entire SCZ family cohort (combined S1+S2) based on familial genotypes where the opposing sex offspring have been excluded, and for 1018 female control individuals or 891 male control individuals.

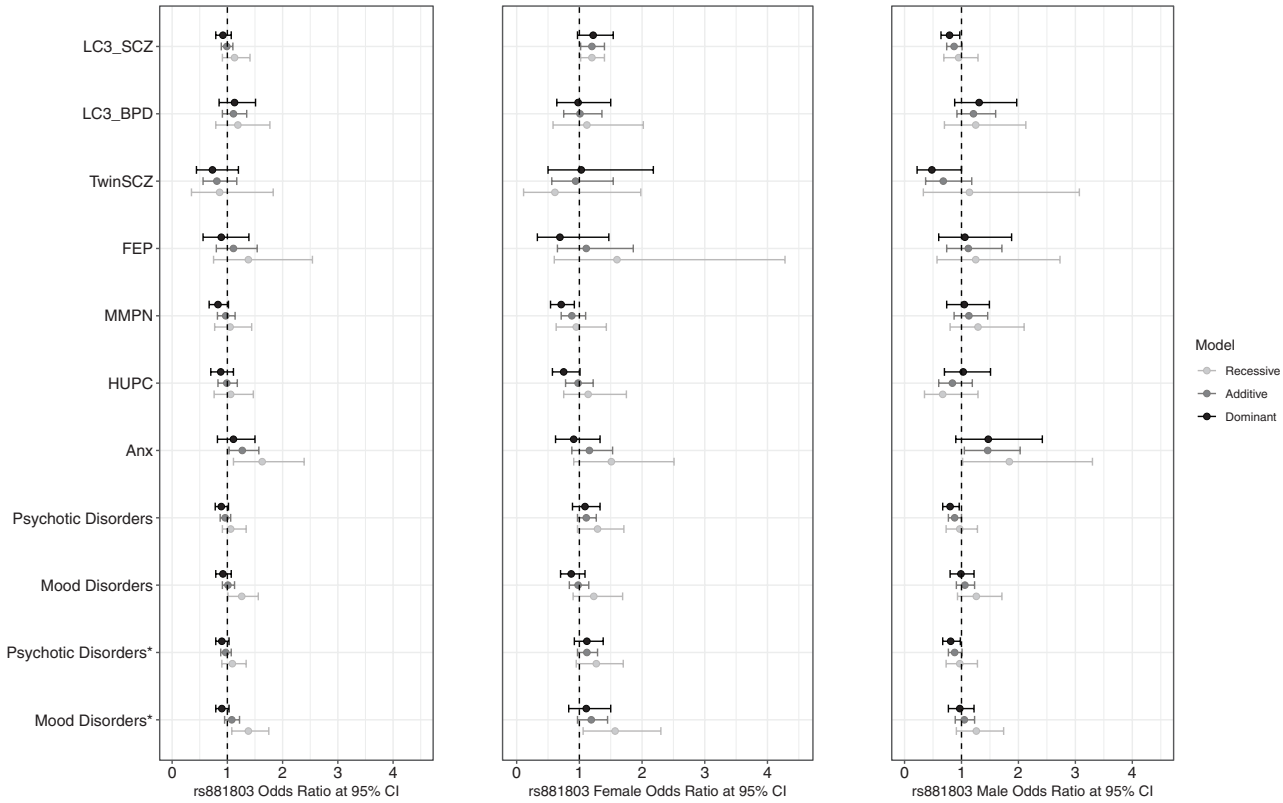

Fig. 1. Odds ratios, and their respective 95% confidence intervals, for rs881803 across the cohorts studied and the joint analyses. All plots represent the observations for the additive genetic model. For the schizophrenia and bipolar disorder family cohorts, the findings for their respective liability class 3's are shown. *Disorders have been analyzed excluding the MMPN cohort.

population-based MMPN cohort in females (rs881803 dominant P = .0097, OR = 0.71, ±95% CI 0.54–0.92) (supplementary tables 4 and 5); however, this was in the opposite direction when compared to the schizophrenia families, the C allele being protective rather than a risk factor. Our SNPs provided some degree of association with anxiety within our population-based anxiety disorder cohort regardless of sex (rs881803 recessive P = .013, OR = 1.63, ±95% CI 1.11–2.39; rs2242549 recessive P = .049, OR = 1.4, ±95% CI 1.00–2.00). Further, when the cohorts were combined for studying broad psychosis or mood phenotypes, both SNPs associated with the psychosis phenotype in females (rs2242549 additive P = .0040, OR = 1.12, ±95% CI 0.97–1.28; rs881803 additive P = .0041, OR = 1.12, ±95% CI 0.97–1.29), and with the mood disorder phenotype in the whole sample (rs2242549 recessive P = .037, OR = 1.18, ±95% CI 0.95–1.46; rs881803 recessive P = .0033, OR = 1.38, ±95% CI 1.08–1.75). Evidence for association was also observed in males between the SNP rs881803 and anxiety and mood disorder (supplementary table 5). Association data mining in the UK Biobank and FinnGen cohorts provide some support for these observations. In the UK biobank, as mined through the Oxford BIG database,[39] rs881803 associates with a diagnosis of schizophrenia (P = .0073, β = .00015), whereas rs2242549 is associated with "F43 reaction to severe stress and adjustment disorders" (P = .0047, β = −.00016). In FinnGen, rs881803 associates with disorders of psychological development (P = .0016, β = .19), and disorders of speech and language (P = .0049, β = .14), whereas rs2242549 associates with "Neurotic, stress-related and somatoform disorders" (P = .00042, β = .044) and anxiety disorders (P = .00051, β = .043).

Endophenotype Analysis

In order to investigate the role of the 2 identified variants roles in the etiology of psychiatric disorders, we studied them with regard to intermediate traits. To explore association within the cognitive domain, we analyzed 14 quantitative endophenotypes, 5 first-order factors, and the general g-factor for cognition. No significant associations were observed that would survive multiple testing correction (supplementary table 6). Neither when females were studied separately, nor when the SCZ cohort was studied separately.

Gene Expression Analysis

We next studied any impact these variants could have on gene expression. This was performed within the SCZ

cohort, using separate analysis of gene expression data generated within the TwinSCZ and FEP cohorts, alongside the GTEx database, in order to provide evidence of replication between the cohorts. In the SCZ cohort, regression analysis identified 1504 probes (n = 1326 unique genes) as significantly associated with rs881803 at a False Discovery Rate (FDR) of $q < 0.05$ (supplementary table 7). In the FEP cohort, 33 of the 206 genes studied were significantly associated ($P < .05$) with rs881803; 9 of the 33 were in common with those identified in the SCZ cohort: *ARNTl*, *GSK3β*, *IL6R*, *MAPK3*, *MYD88*, *PER2*, *SOCS3*, *STAT3*, and *STAT5A* (supplementary table 8). In the TwinSCZ cohort, only 2 probes out of 18,559 were significantly associated ($P < .05$) with rs881803, neither of which were identified in the original SCZ cohort (supplementary table 7). In the GTEx database, 59 of the 1086 genes (240 not recognized by GTEx) that were tested in whole blood were significant eQTLs with rs881803 (supplementary table 8).

Genes surviving FDR from the SCZ cohort were checked for enrichment with predicted targets of miR-484, and for genes significantly differentially expressed between the 2 sexes. Of the 555 genes predicted to be targeted by miR-484 (547 recognized by IPA), 54 (9.9%) were significantly associated with rs881803 (supplementary table 8; miR-484 P = .000193). In our SCZ cohort, only 14 genes were significantly different between the 2 sexes after FDR correction. These were not enriched for probes that significantly associated with rs881803. However, taking a number of P-value cutoffs ($P < .05$, $P < .01$, $P < .001$) prior to FDR correction shows a consistent 9–10% overlap between these genes and those associated with rs881803 (supplementary table 8, gender P = .00187, 48 out of 521 [9.2%]). Of these 605 genes (656 probes) that were significantly different between the sexes ($P < .05$) in the SCZ cohort, 305 genes (322 probes) were also noted to be significantly different between the sexes in the TwinSCZ cohort (supplementary table 8). These replicating 305 sex different genes had an 8% overlap with those significantly associated with rs881803 (supplementary table 8, P = .034, 24 out of 301 [8%]).

Discussion

Here we have followed up a continued line of evidence indicating that the *NDE1* gene locus is involved in the etiology of psychiatric disorders in Finland.[11–14] We aimed to identify any functional variant or variants at this locus, and determine if they help to differentiate between *NDE1* and/or miR-484 as the biologically relevant component at this locus. Through sequencing 96 related individuals (79 independent chromosomes) in 20 SCZ families, we identified 295 variants over the NDE1 genomic locus. None were immediately located within the mir-484 coding sequence. While this directly indicates that there is no common genetic variation in the microRNA in the Finnish population, it should be kept in mind that our sequencing cohort did not have sufficient power to identify rare mutations (minor allele frequency MAF<0.01) within these families. However, our previous study of *NDE1*[11] noted a haplotype that had a MAF of 0.30 in the founder schizophrenia families (control MAF = 0.19), suggesting that any functional mutation tagged by this haplotype will most likely be common (MAF > 0.05).

While the three-stage study design did not identify any variants that replicate in their association to schizophrenia, we were able to utilize our prior observations[14] to follow the role of 2 variants associated with gene expression changes. One of the 2 was rs2242549, the SNP that displayed the original association with the 3 gene expression probes in this familial schizophrenia cohort. The other SNP, rs881803, is a variant in high LD with rs2242549 (r^2 = .75), but is also annotated to be located within a Transcription Factor Binding Site (OCT1_06, POU2F1). Both selected SNPs associated with schizophrenia within the familial SCZ cohort (S1+S2), but only in a sex-dependent manner, increasing risk in females (supplementary table 3). Interestingly, both SNPs were a part of the originally identified *NDE1* haplotype that also displayed sex-dependent association with over-transmission to females.[11] In that study, both SNPs associated significantly with the broadest LC4 diagnosis in the SCZ cohort in females, yet the original haplotype's significance (P = .00046) surpassed the individual SNP observations (rs2242549 P = .018, rs881803 P = .0013).[11] Here, in this enlarged SCZ cohort, the SNP associations have strengthened in comparison with the original findings, as would be expected from a genuine observation. However, haplotypes were not tested in this study, as the aim of the study was to narrow our focus to individual variation.

The 2 SNPs identified were further investigated in clinical cohorts ascertained for psychiatric disorders, and with broader definitions of psychotic and mood domains. In the individual cohorts, rs881803 associated with anxiety disorders, and within the psychotic disorder MMPN female-only cohort. However, the C allele, while being initially observed as the risk allele, was protective in this MMPN cohort. The reason for this cohort-specific observation is most likely due to the much smaller sample size of the MMPN cohort, and thus could be a chance observation. When this variant was checked in large nationwide cohorts, we observed support not just for the association with SCZ, but also for the direction of risk indicated by our observations in the SCZ cohort. Both the UK biobank (n = 452 264; 590 with a diagnosis of schizophrenia) and FinnGen (n = 175 519; 595 with a diagnosis of speech disorder, n = 161 390; 15 509 with a diagnosis of mental and behavioral disorders) note association ($P < .05$) to diagnoses related to mental illness or neurodevelopment, with the C allele of rs881803 being the one that increases risk. It should be noted that

these data mining methods do not consider sex-specific effects. Due to the opposing effect seen in MMPN, when the combined cohorts were analyzed, they were additionally analyzed, excluding individuals from the MMPN cohort. Both SNPs highlighted sex-dependent associations in the etiology of psychotic disorders in females, and with mood disorders in general (supplementary table 5), demonstrating the role of this locus in the broader etiology of major mental illnesses. It should be noted that the models and conditions of these associations do not converge to any one particular inheritance pattern, with the provided odds ratios sometimes having their confidence intervals include unity. While this may be an artifact of including mixed etiological cohorts in the analysis here. The evidence from completely independent cohorts, FinnGen, and UK biobank, provides these findings with robust support. Despite these observations, recent population-based genome-wide consortia studies have failed to implicate *NDE1* in the etiology of psychiatric illnesses.[40,41] The possible explanation behind this could be that the *NDE1* SNPs being reported in our study have low odds ratio/effect size, confirming that the disease being studied is polygenic in nature, caused by a large number of interacting genes, with *NDE1* contributing only a small, additive increment to disease risk. Thus, to detect any significant association using population-based approaches, even larger sample sizes are required. Finally, analysis of our identified variants with cognitive endophenotypes again indicated suggestive association for both the SNPs with one of the verbal working memory endophenotypes (supplementary table 6 a–d); however, these associations did not survive multiple test correction.

We have previously reported association between rs2242549 and gene expression probes.[14] In the SCZ cohort analysis with rs881803 identified 1504 probes, representing 1326 genes, whose expression levels associated with the genotype at a 5% false discovery rate (FDR). These genes were significantly over-represented within the predicted targets of miR-484 (9.9% overlap, 54 out of 547 genes, *P*-value = .000193) (supplementary table 8). Additionally, these genes significantly overlap with those differentially expressed between the sexes in the SCZ cohort. Thus, demonstrating that some part of the functional consequences of variation at this genomic locus are likely to be through miR-484 moderated gene regulation, whilst also offering initial indications as to the biological changes that could be behind the observed sex differences in risk. When these 1326 genes were studied for evidence of replication in other Finnish cohorts, we identified 9 genes (including *GSK3β* a previously reported *DISC1* network interactor[42,43] that are significantly differentially expressed in the FEP cohort, and 59 that can be identified as eQTLs of rs881083 in the healthy GTEx population. However, we were not able to replicate these observations with rs881803 in the TwinSCZ cohort, when previously we had been able to replicate the associations between rs2242549 genotype and gene expression probes.[14] It should be noted that although these 2 SNPs are in high r^2 LD, their allele frequencies are different (rs2242549 MAF = 0.43; rs881803 MAF = 0.36). The SNP rs881803 has a lower frequency to rs2242549, to such an extent that within the TwinSCZ cohort there were not enough individuals homozygous for the minor allele to specifically test the recessive genetic model, which has been previously suggested as the most likely model at this locus.[14] The fact that some of the observed gene expression changes can already be observed in individuals undergoing their first episode of psychosis, suggests that these changes are not all related to the long-term consequences of these devastating disorders. While the fact that some of the gene expression changes associated with rs881803, can be observed in the GTEx database, suggests that some of the observed changes may contribute to natural variation in the healthy population.

In summary, our study identified 2 SNP variants at the 16p13.11 locus that associate with Finnish familial schizophrenia, in females. Exploration of their role in other major mental illnesses in Finland, alongside nationwide biobanks indicates that they play a broader role, by associating with both psychosis and mood domains of these disorders, as well as anxiety-related traits. Beyond diagnoses, exploration of the consequences of these variations demonstrated large-scale gene expression changes related to the rs881803 genotype. These were significantly enriched for genes differentially expressed between the sexes, thus beginning to reveal why our association observations are sex-specific, and for genes predicted to be targets of miR-484. Both SNPs are significant eQTLs for *NDE1* in the GTEx database,[44] with these changes being significant across multiple tissue types, including all brain regions studied (supplementary figure 4). We conclude that the SNPs identified here contribute to dysregulation of the whole region, both *NDE1* and miR-484, with both their downstream consequences increasing risk to mental disorders. The 2 SNPs being in high LD with each other makes it difficult to differentiate, which is the functional element underlying our findings. While we cannot rule out a role for rs2242549, rs881803 is annotated in the UCSC database to be located in a transcription factor binding site, making it the primary candidate. However, further biological studies will be needed to directly test the consequences of these variants.

Supplementary Material

Supplementary data are available at *Schizophrenia Bulletin Open* online.

Funding

This research has been supported and funded by the Academy of Finland (Grant numbers 128504, 259589,

and 265097), Marie Curie Initial Training Network EU FP7 (Grant Number 607616), and the Finnish Cultural Foundation (Ingrid, Toini and Olavi Martelius Grant 2018) for W.H., the Sigrid Juselius Foundation for J.L., and the Jalmari and Rauha Ahokas Foundation, and the Biomedicum Helsinki Foundation for V.S. M.T-H. is supported by the Academy of Finland (Grant Number 310295). J.S. is supported by the Finnish Cultural Foundation (J.S.), the Sigrid Juselius Foundation (J.S.) and the Academy of Finland (Grant numbers 278171 and 323035). J.K. is supported by the Academy of Finland (Grants 265240 and 312073), and O.L. by The Fonds de recherche du Québec (Grant numbers 252872 and 265693) and State Funding for University-Level Health Research (Hospital District of Helsinki and Uusimaa TYH2013332, TYH2014228, TYH2017128). The funders had no further role in the study design, in the collection, analysis, and interpretation of data, in the writing of the report, or in the decision to submit the article for publication. Finally, we thank the FinnGen project for providing access to their data.

Acknowledgments

We thank the Institute for Molecular Medicine Finland FIMM Technology Centre, sequencing and genotyping unit at the University of Helsinki for NGS library preparation, enrichment, sequencing, and sequence analysis. We thank Sarang Talwelkar for helping with the figures. V.S. and W.H. wrote the manuscript text and prepared the manuscript tables and figures; W.H. designed the study; T.P., J.S., J.L., I.H., P.J., E.I., O.L., A.T.H., S.T., T.D.C., and J.K. provided access to samples and data; V.S., A.O.A., L.U.V., W.H., A.B.Z., and M.T.H. performed the analysis. All authors have reviewed the manuscript and approved the final version to be published. W.H. has a co-appointment at Orion Pharma. T.D.C. reports that he is a consultant to Boehringer Ingelheim Pharmaceuticals and to Lundbeck A/S. A.B.Z. now works at Sema4, a health intelligence company. All other authors declare no potential conflicts of interest in relation to the subject of this study.

References

1. Mefford HC, Cooper GM, Zerr T, *et al.* A method for rapid, targeted CNV genotyping identifies rare variants associated with neurocognitive disease. *Genome Res.* 2009;19(9):1579–1585.
2. Coe BP, Witherspoon K, Rosenfeld JA, *et al.* Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat Genet.* 2014;46(10):1063–1071.
3. Ullmann R, Turner G, Kirchhoff M, *et al.* Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. *Hum Mutat.* 2007;28(7):674–682.
4. Levy D, Ronemus M, Yamrom B, *et al.* Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron.* 2011;70(5):886–897.
5. Williams NM, Zaharieva I, Martin A, *et al.* Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. *Lancet.* 2010;376(9750):1401–1408.
6. Hannes FD, Sharp AJ, Mefford HC, *et al.* Recurrent reciprocal deletions and duplications of 16p13.11: the deletion is a risk factor for MR/MCA while the duplication may be a rare benign variant. *J Med Genet.* 2009;46(4):223–232.
7. Heinzen EL, Radtke RA, Urban TJ, *et al.* Rare deletions at 16p13.11 predispose to a diverse spectrum of sporadic epilepsy syndromes. *Am J Hum Genet.* 2010;86(5):707–718.
8. de Kovel CG, Trucks H, Helbig I, *et al.* Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. *Brain.* 2010;133(Pt 1):23–32.
9. Ingason A, Rujescu D, Cichon S, *et al.*; GROUP Investigators. Copy number variations of chromosome 16p13.1 region associated with schizophrenia. *Mol Psychiatry.* 2011;16(1):17–25.
10. Hennah W, Varilo T, Kestilä M, *et al.* Haplotype transmission analysis provides evidence of association for DISC1 to schizophrenia and suggests sex-dependent effects. *Hum Mol Genet.* 2003;12(23):3151–3159.
11. Hennah W, Tomppo L, Hiekkalinna T, *et al.* Families with the risk allele of DISC1 reveal a link between schizophrenia and another component of the same molecular pathway, NDE1. *Hum Mol Genet.* 2007;16(5):453–462.
12. Ekholm JM, Kieseppä T, Hiekkalinna T, *et al.* Evidence of susceptibility loci on 4q32 and 16p12 for bipolar disorder. *Hum Mol Genet.* 2003;12(15):1907–1915.
13. Wegelius A, Pankakoski M, Tomppo L, *et al.* An interaction between NDE1 and high birth weight increases schizophrenia susceptibility. *Psychiatry Res.* 2015;230(2):194–199.
14. Bradshaw NJ, Ukkola-Vuoti L, Pankakoski M, *et al.* The NDE1 genomic locus can affect treatment of psychiatric illness through gene expression changes related to microRNA-484. *Open biology*. 2017;7.
15 Fujitani M, Zhang S, Fujiki R, Fujihara Y, Yamashita T. A chromosome 16p13.11 microduplication causes hyperactivity through dysregulation of miR-484/protocadherin-19 signaling. *Mol Psychiatry.* 2017;22(3):364–374.
16. Johnstone M, Vasistha NA, Barbu MC, *et al.* Reversal of proliferation deficits caused by chromosome 16p13.11 microduplication through targeting NFκB signaling: an integrated study of patient-derived neuronal precursor cells, cerebral organoids and in vivo brain imaging. *Mol Psychiatry.* 2019;24(2):294–311.
17. Sinha V, Ukkola-Vuoti L, Ortega-Alonso A, *et al.* Variants in regulatory elements of PDE4D associate with major mental illness in the Finnish population. *Mol Psychiatry.* 2019; doi:10.1038/s41380-019-0429-x
18 Ekholm JM, Pekkarinen P, Pajukanta P, *et al.* Bipolar disorder susceptibility region on Xq24-q27.1 in Finnish families. *Mol Psychiatry.* 2002;7(5):453–459.
19. Kaprio J, Koskenvuo M, Rose RJ. Population-based twin registries: illustrative applications in genetic epidemiology and behavioral genetics from the Finnish Twin Cohort Study. *Acta Genet Med Gemellol (Roma)*. 1990;39(4):427–439.
20. Cannon TD, Kaprio J, Lönnqvist J, Huttunen M, Koskenvuo M. The genetic epidemiology of schizophrenia in a Finnish twin cohort. A population-based modeling study. *Arch Gen Psychiatry.* 1998;55(1):67–74.

21. Mantere O, Saarela M, Kieseppä T, *et al.* Anti-neuronal anti-bodies in patients with early psychosis. *Schizophr Res.* 2018;192:404–407.
22. Aaltonen K, Näätänen P, Heikkinen M, *et al.* Differences and similarities of risk factors for suicidal ideation and attempts among patients with depressive or bipolar disorders. *J Affect Disord.* 2016;193:318–330.
23. Donner J, Pirkola S, Silander K, *et al.* An association analysis of murine anxiety genes in humans implicates novel candidate genes for anxiety disorders. *Biol Psychiatry.* 2008;64(8):672–680.
24. https://thl.fi/en/web/thl-biobank/for-researchers/sample-collections/health-2000-and-2011-surveys.
25. https://thl.fi/en/web/thl-biobank/for-researchers/sample-collections/thl-psychiatric-family-collections-1994–2008.
26. *DSM-IV-TR*. American Psychiatric Association; 2000.
27. Ukkola-Vuoti L, Torniainen-Holm M, Ortega-Alonso A, *et al.* Gene expression changes related to immune processes associate with cognitive endophenotypes of schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry.* 2019;88:159–167.
28. Cannon TD, Hennah W, van Erp TG, *et al.* Association of DISC1/TRAX haplotypes with schizophrenia, reduced prefrontal gray matter, and impaired short- and long-term memory. *Arch Gen Psychiatry.* 2005;62(11):1205–1213.
29. Hennah W, Tuulio-Henriksson A, Paunio T, *et al.* A haplotype within the DISC1 gene is associated with visual memory functions in families with a high density of schizophrenia. *Mol Psychiatry.* 2005;10(12):1097–1103.
30. Delis DC, Kramer JH, Kaplan E, Ober BA. *California Verbal Learning Test (CVLT)*. San Antonio, TX: The Psychological Corporation; 1987.
31. Wechsler D. *Manual for the Wechsler Adult Intelligence Scale-Revised (WAIS-R)*. San Antonio, TX: The Psychological Corporation; 1981.
32. Wechsler D. *WMS-R: Wechsler Memory Scale-Revised*. San Antonio, TX: Psychological Corporation; 1987.
33. Golden CJ. *Stroop Color and Word Test: A Manual for Clinical and Experimental Uses*. Chicago, IL: Stoelting Co; 1978.
34. Reitan RM. *Trail Making Test: Manual for Administration and Scoring*. Tucson, AZ: Reitan Neuropsychology Laboratory; 1986.
35. Gertz EM, Hiekkalinna T, Digabel SL, Audet C, Terwilliger JD, Schäffer AA. PSEUDOMARKER 2.0: efficient computation of likelihoods using NOMAD. *BMC Bioinf.* 2014;15:47.
36. https://juba.github.io/questionr/.
37. Purcell S, Neale B, Todd-Brown K, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559–575.
38. Abecasis GR, Cardon LR, Cookson WO. A general test of association for quantitative traits in nuclear families. *Am J Hum Genet.* 2000;66(1):279–292.
39. Elliott LT, Sharp K, Alfaro-Almagro F, *et al.* Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature.* 2018;562(7726):210–216.
40. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511:421–427.
41. Pardiñas AF, Holmans P, Pocklington AJ, *et al.*; GERAD1 Consortium; CRESTAR Consortium. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet.* 2018;50(3):381–389.
42. Gao FJ, Hebbar S, Gao XA, *et al.* GSK-3β phosphorylation of cytoplasmic dynein reduces Ndel1 binding to intermediate chains and alters dynein motility. *Traffic.* 2015;16(9):941–961.
43. Ogawa F, Murphy LC, Malavasi EL, *et al.* NDE1 and GSK3beta associate with TRAK1 and regulate axonal mitochondrial motility: identification of cyclic AMP as a novel modulator of axonal mitochondrial trafficking. *ACS Chem Neurosci.* 2016;7:553–564.
44. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45:580–585.