A genetic variant that disrupts MET transcription is associated with autism

Daniel B. Campbell*, James S. Sutcliffe†, Philip J. Ebert*, Roberto Militerni§, Carmela Bravaccio§, Simona Trillo¶, Maurizio Elia†, Cindy Schneider**, Raun Melmed††, Roberto Sacco***, Antonio M. Persico***, and Pat Levitt****

Departments of *Pharmacology and 1Molecular Physiology and Biophysics and 1Vanderbilt Kennedy Center for Research on Human Development, Vanderbilt University, Nashville, TN 37203; 2Department of Child Neuropsychiatry, II University of Naples, I-80131 Naples, Italy; 3Associazione Anni Verdi ONLUS, 00148 Rome, Italy; 4Unit of Neurology and Clinical Neurophysiopathology, Scientific Institutes for Research, Hospitalization and Health Care (IRCCS) Oasi Maria SS, 94018 Troina, EN, Italy; 5Center for Autism Research and Education, Phoenix, AZ 85012; 6Southwest Autism Research and Resource Center, Phoenix, AZ 85006; 7Laboratory of Molecular Psychiatry and Neurogenetics, University Campus Bio-Medico, I-00155 Rome, Italy; and 8IRCCS Fondazione Santa Lucia, 00179 Rome, Italy

Edited by Mary-Claire King, University of Washington, Seattle, WA, and approved September 1, 2006 (received for review June 23, 2006)

There is strong evidence for a genetic predisposition to autism and an intense interest in discovering heritable risk factors that disrupt gene function. Based on neurobiological findings and location within a chromosome 7q31 autism candidate gene region, we analyzed the gene encoding the pleiotropic MET receptor tyrosine kinase in a family based study of autism including 1,231 cases. MET signaling participates in neocortical and cerebellar growth and maturation, immune function, and gastrointestinal repair, consistent with reported medical complications in some children with autism. Here, we show genetic association (P = 0.0005) of a common C allele in the promoter region of the MET gene in 204 autism families. The allele association at this MET variant was confirmed in a replication sample of 539 autism families (P = 0.001) and in the combined sample (P = 0.000005). Multiplex families, in which more than one child has autism, exhibited the strongest allelic association (P = 0.000007). In case-control analyses, the autism diagnosis relative risk was 2.27 (95% confidence interval: 1.41–3.65; P = 0.0006) for the CC genotype and 1.67 (95% confidence interval: 1.11–2.48; P = 0.012) for the CG genotype compared with the GG genotype. Functional assays showed that the C allele results in a 2-fold decrease in MET promoter activity and altered binding of specific transcription factor complexes. These data implicate reduced MET gene expression in autism susceptibility, providing evidence of a previously undescribed pathophysiological basis for this behaviorally and medically complex disorder.

A genetic variant that disrupts MET transcription is associated with autism.

Autism is a complex, behaviorally defined neurodevelopmental disorder characterized by social deficits, language impairments, and repetitive behaviors with restricted interests. There has been a dramatic increase in the diagnosis of autism (1). The population prevalence of autism is debated; recent reports indicate that ~1 in 500 individuals have autism and as many as 1 in 166 individuals have an autism spectrum disorder (ASD) (1, 2). Here, we broadly use the term “autism” to refer to any individual with ASD; for the purposes of our genetic analyses, we use a binary code to designate as “affected” any individual diagnosed with autism or ASD and “unaffected” any individual lacking such a diagnosis. The etiology of this complex disease likely involves environmental factors, but autism is highly heritable. Twin studies demonstrate concordance rates of 82–92% in monozygotic twins and 1–10% concordance rate in dizygotic twins (3). Sibling recurrence risk (6–8%) is 35 times the population prevalence (1,4), making autism among the most heritable of all neuropsychiatric disorders.

The most widely accepted hypotheses regarding autism etiology include oligogenic inheritance and epistatic interactions among common vulnerability-conferring genetic variants and, possibly, gene–environment interactions. Genomewide linkage scans, an unbiased approach to localize genetic factors, have identified several chromosomal regions as promising locations for autism vulnerability genes, including peaks on chromosomes 2q, 7q, 15q, and 17q (5–8). This genetic approach to identify susceptibility genes is very powerful, but the heterogeneity present within autism families has led thus far to mixed success in identifying candidate genes (9, 10). In pursuing specific candidates, most studies have focused on genes expressed predominantly in the brain (11, 12). Our own evaluation of linkage peaks extended beyond genes with selective brain expression to consider the complex medical conditions seen in autism patients. In addition to the well known behavioral core features, some individuals with autism exhibit gastrointestinal, immunological, or nonspecific neurological symptoms (13–15). Although the degree to which individuals with autism exhibit more medical complications compared with typical individuals is debated, it is possible that autism vulnerability could include genes involved more broadly in multiple biological processes that impact the development and function of the brain and other organ systems in parallel.

We applied the convergence of developmental biological and genetic information to analyze the gene encoding the MET receptor tyrosine kinase (OMIM 164860; GenBank accession NM_000245; chromosome 7q31) as an autism candidate gene. MET is best understood for its role in the metastasis of a variety of cancers (16), due to somatic gain-of-function mutations, and in mediating hepatocyte growth factor (HGF)/scatter factor signaling in peripheral organ development and repair (17–19). MET signaling contributes to immune function (20–22) and gastrointestinal repair (18, 23, 24). Recent studies by our group and others revealed that MET also contributes to development of the cerebral cortex (25, 26) and cerebellum (27), both of which exhibit developmental disruptions in autism (28, 29). Hypomorphic MET/HGF signaling in the cerebral cortex results in abnormal interneuron migration from the ganglionic eminence and reduced interneurons in the frontal and parietal regions of cortex (25, 26). Hypomorphic MET/HGF signaling in the cerebellum causes decreased proliferation of granule cells and a concomitant reduction in the size of the cerebellum, particularly in the vermis (27). Both of these neuropathologic abnormalities are consistent with those observed in brains of individuals with autism (28, 29). We therefore pursued MET as an autism candidate gene.


The authors declare no conflict of interest.

This article is a PNAS direct submission.

Abbreviations: FBAT, family based association test; HBAT, haplotype-based association test; LD, linkage disequilibrium; TEXP, transmissions expected; T_OBS, transmissions observed.

See Commentary on page 16621.

†To whom correspondence should be addressed. E-mail: pat.levitt@vanderbilt.edu.

© 2006 by The National Academy of Sciences of the USA
frequencies (distinct LD blocks (Fig. 1): a 17-kb block at the 5′ linkage disequilibrium (LD) revealed that simplex and 452 multiplex) (Table 3). Analysis of intermarker original sample consisting of 204 families (178 simplex and 26 span the entire chromosome 7q31. Nine SNPs spanning the MET locus were chosen to perform association studies and Taqman Assays-on-Demand were used to determine genotype. The nine genotyping markers defined two distinct linkage disequilibrium blocks. Pairwise linkage disequilibrium ($D^*$) values are indicated. Pairwise $r^2$ values are provided in Table 1, which is published as supporting information on the PNAS web site.

**Results**

**Screen of the MET Gene for Variants in Autism.** To identify variants in the MET gene, we screened the 21 exons and key regulatory regions of the gene in 86 individuals with autism by using temperature gradient capillary electrophoresis and direct resequencing. Primers used to amplify the exonic regions of the MET gene are listed in Table 2, which is published as supporting information on the PNAS web site. Two rare nonsynonymous variants were identified in exon 14: C3095T, a nonconserved arginine-to-cysteine substitution at amino acid 988 (R988C), and C3162T, a threonine-to-isoleucine substitution at amino acid 1010 (T1010I). These same variants were reported in small cell lung cancer cell lines as functional somatic mutations (30). We directly resequenced 277 cases and 319 unrelated controls (Table 3, which is published as supporting information on the PNAS web site) to determine the frequencies of the two nonsynonymous exon 14 variants. For each of the variants, R988C and T1010I, the rare allele was present in five cases (1.8%) and two controls (0.6%). These differences are not significant either for genotypic ($\chi^2 = 1.773; df = 2; P = 0.412$) or for allelic frequencies ($\chi^2 = 1.762; df = 1; P = 0.184$). Thus, there is no genetic evidence to support autism association for these rare alleles. Synonymous SNPs were identified in exon 2 (rs11762213), exon 7 (rs13223756), exon 20 (rs41736), and exon 21 (rs2023748 and rs41737). The initial screen also identified variants in the promoter region (rs184953 and rs1858830) and a variant (rs41739) in the 3′ untranslated region of the MET gene.

**Family Based Association Analyses.** To determine a possible association between MET and autism, we tested for transmission disequilibrium in a two-stage study design by using nine markers that span the entire MET locus (Fig. 1). SNPs were genotyped in an original sample consisting of 204 families (178 simplex and 26 multiplex), followed by a replication sample of 539 families (87 simplex and 452 multiplex) (Table 3). Analysis of intermarker linkage disequilibrium (LD) revealed that MET contains two distinct LD blocks (Fig. 1): a 17-kb block at the 5′ end of the gene and an expansive 110-kb block that includes exons 2–21, the entire coding region of the MET gene. Transmissions of haplotypes within each LD block were examined with the haplotype-based association test (HBAT) (31). HBAT analysis revealed significant transmission distortion in LD block 1 ($\chi^2 = 17.521; df = 6; P = 0.008$), indicating the presence of an autism-associated variant in the MET promoter region.

The LD block 1 haplotype association supported the possibility of identifying a variant that disrupts MET gene regulation. We thus examined transmissions of single markers by using the family based association test (FBAT) (32). Parent-to-affected offspring transmissions observed ($T_{OBS}$) were compared with transmissions expected ($T_{EXP}$), generating a $P$ value representing the probability of observing the transmission disequilibrium by chance. We observed a significant overtransmission of the rs1858830 C allele to affected individuals (Fig. 2; see Tables 4–6, which are published as supporting information on the PNAS web site). Transmission disequilibrium for rs1858830 under a dominant model was significant in both the original 204-family sample ($T_{OBS} = 81, T_{EXP} = 65, P = 0.0005$) and in the replication sample of 539 families ($T_{OBS} = 225, T_{EXP} = 198, P = 0.001$); combined analysis of these samples was highly significant ($T_{OBS} = 306, T_{EXP} = 263, P = 0.000005$) (Fig. 2a). The rs1858830 variant is a common G/C SNP, situated just 20 bp 5′ to the MET transcriptional start site, and promoter variants often function as dominant mutations (33). However, FBAT analyses using an additive model yielded similar results: the original sample showed significant association ($T_{OBS} = 139, T_{EXP} = 119, P = 0.006$), a trend in the replication sample ($T_{OBS} = 490, T_{EXP} = 464, P = 0.072$) and again a significant association in the combined samples ($T_{OBS} = 629, T_{EXP} = 583, P = 0.005$) (Fig. 2b). Among the markers in LD with rs1858830, only the most informative based on allele frequency (rs437; minor allele frequency 0.295; pairwise $r^2 = 0.270$) was significantly associated with autism; less informative markers (rs184953 and rs40238; minor allele frequency <0.178; pairwise $r^2$ to rs1858830 < 0.164; Table 1) failed to show an association. No other marker consistently reached significant transmission disequilibrium after corrections for multiple comparisons (Fig. 2; Tables
Association of the rs1858830 C allele was restricted to multiplex families, with simplex families displaying no association under either the dominant model (multiplex: \( T_{\text{OBS}} = 236, T_{\text{EXP}} = 198, P = 0.000007);\) simplex: \( T_{\text{OBS}} = 70, T_{\text{EXP}} = 65, P = 0.202)\) or the additive model (multiplex: \( T_{\text{OBS}} = 494, T_{\text{EXP}} = 447, P = 0.001);\) simplex: \( T_{\text{OBS}} = 135, T_{\text{EXP}} = 136, P = 0.886).\) In addition, we used the Autism Genetic Resource Exchange database to identify 299 individuals with autism from 182 families who are positive for a narrow diagnosis, based on Autism Diagnostic Interview-Revised, Autism Diagnostic Observational Schedule, and a clinical diagnosis. The rs1858830 C allele was significantly overtransmitted to individuals with narrowly defined autism (\( T_{\text{OBS}} = 94, T_{\text{EXP}} = 75, P = 0.0002)\) (Table 9, which is published as supporting information on the PNAS web site). For comparison, exclusion of the 182 families with narrow diagnosis from the entire 743-family sample resulted in a significant but somewhat weaker association (\( T_{\text{OBS}} = 212, T_{\text{EXP}} = 188, P = 0.003)\). Thus, both subpopulations contribute to the association of the rs1858830 C allele in the combined sample (\( T_{\text{OBS}} = 306, T_{\text{EXP}} = 263, P = 0.000005)\).

**Case-Control Association of MET Promoter Variant rs1858830.** Genotype at the rs1858830 locus was determined in a group of 189 unrelated Italian and American control individuals. A single individual with autism was randomly selected from each of the pedigrees genotyped in the combined family based association sample. Significant differences in genotypic and allelic frequencies were detected between the individuals with autism and controls (genotypic: \( \chi^2 = 12.150); df = 2; P = 0.002)\); allelic: \( \chi^2 = 10.440); df = 1; P = 0.001)\) (Table 10, which is published as supporting information on the PNAS web site). Compared with the GG genotype, the autism diagnosis relative risk was 2.27 (95% confidence interval (CI): 1.41, 3.65) for the CC genotype and 1.67 (95% CI: 1.11, 2.49) for the CG genotype when analyzing cases and unrelated controls (Table 10). This relative risk may be biologically relevant in a polygenic disease such as autism. Thus, the rs1858830 C allele is common and overrepresented in individuals with autism.

**Transcription Assays.** Given the location of the rs1858830 G/C variant (20-bp 5' to the transcription start site), we hypothesized that the associated allele would affect transcription of the MET gene. To test this hypothesis, we generated two reporter constructs containing 726 bp of the human MET promoter, differing a fraction of the simplex family cases are more likely to be caused by rare chromosomal rearrangements or other sporadic events (34). Examination of transmissions of all nine markers to unaffected siblings showed no significant transmission distortion, indicating that the overtransmission of the rs1858830 C-allele is specific to autism and not due to altered viability.

To further understand the heritability of the MET promoter allele in our large sample (1,231 individuals with autism from 743 families), we examined the 265 simplex (one affected child, irrespective of number of siblings) and 478 multiplex (more than one affected child) families independently (Fig. 2c; Tables 7 and 8, which are published as supporting information on the PNAS web site). Within the framework of a genetically complex, heterogeneous, polygenic disorder like autism, common genetic predisposing factors are likely to be enriched in multiplex families, whereas
only at the rs1858830 nucleotide, and transfected them into mouse neural cell lines N2A and SN56 and the HEK cell line. The reporter construct containing the C allele produced less than half the luciferase activity than the construct containing the G allele ($P < 0.05$ for each of the three cell lines; Fig. 3). The 2-fold reduction in promoter activity indicates that the autism-associated rs1858830 C allele is less efficient in driving transcription than the G allele, demonstrating that the rs1858830 variant is a functional regulatory element of \textit{MET} transcription.

**Identification of Transcription Factors That Differentially Bind the rs1858830 Variant.** We next attempted to identify the mechanisms through which the rs1858830 \textit{MET} variant might influence transcription by examining this region for transcription factor consensus sequences. The transcription factor database TRANSFAC (35) predicted that the G and C alleles would differentially bind the transcription factors SP1 and AP2 (Fig. 4a). Indeed, in EMSA, a G allele-containing oligonucleotide probe robustly bound a single DNA–protein complex in HeLa nuclear extracts, whereas an oligonucleotide probe containing the C allele weakly bound at least two protein complexes, one similar in size to that bound by the G allele oligonucleotide probe and another that migrated more slowly (Fig. 4b). EMSA with human fetal brain nuclear protein similarly showed that the G allele oligonucleotide probe bound a single transcription factor complex more robustly than the C allele probe (Fig. 5a, which is published as supporting information on the PNAS web site). To determine the specific transcription factors involved in the DNA–protein complexes, we performed supershift assays with antibodies directed to the predicted transcription factors, SP1 and AP2, as well as the SP1-family member SP3 and a transcription factor identified in a preliminary screen, PC4. Incubation of the DNA–protein complex with specific antibodies revealed that the predominant protein in the complex is the SP1 transcription factor: Addition of SP1 antibody created a visibly supershifted band, representing a DNA–protein–antibody complex, upon incubation with the G allele oligonucleotide probe in HeLa nuclear extracts (Fig. 4c). Addition of SP1 antibody created a visibly supershifted band, representing a DNA–protein–antibody complex, upon incubation with the G allele oligonucleotide probe in HeLa nuclear extracts (Fig. 4c).
of SP1 antibody in supershift assays with the C allele oligonucleotide probe caused markedly decreased DNA–protein complex formation, indicating a specific interaction of the antibody with the DNA–protein complex. Similar results were observed in SP1-antibody supershift assays with human fetal brain nuclear protein (Fig. 5b). The antibody directed to PC4 consistently competed more effectively with the C allele probe–protein complex than with the G allele probe–protein complex (Figs. 4c and 5b), indicating a differential interaction of the PC4 transcription factor with the rs1858830 alleles.

Discussion

The genetic and molecular data reported here indicate genetic association of a common, functional variant of MET with autism with a calculated relative risk of 2.27. There are several unusually attractive aspects of these findings. First, neuropathological findings in autism indicate altered organization of both the cerebral cortex and cerebellum, both of which are disrupted in mice with decreased MET signaling activity. There is co-occurrence of autism with a number of neurological and cognitive disorders, including epilepsy, atypical sleep patterns, and mental retardation (36). Together with well known dysfunction of cortical information processing, the role of MET signaling in interneuron development is relevant as a central component of the hypothesized GABAergic pathophysiological changes in autism (37). Second, the rise in autism diagnosis likely represents changing diagnostic criteria, increased awareness and an increased incidence (1, 4). Although yet to be identified environmental factors likely contribute to the development of autism, heritability studies suggest that the impact of those factors must be imposed upon individuals genetically predisposed to the disorder. Only a limited number of disease-related functional alleles have been identified to date in autism cases, and they only account for a small fraction of cases (38). We hypothesize that the common, functionally disruptive rs1858830 C allele can, together with other vulnerability genes and epigenetic and environmental factors, precipitate the onset of autism. The existence of epistatic interactions among common genetic variants at several different loci is further supported by the association between the rs1858830 C allele and autism in multiplex families (39). The antibody directed to PC4 consistently competed more effectively with the C allele probe–protein complex than with the G allele probe–protein complex (Figs. 4c and 5b), indicating a differential interaction of the PC4 transcription factor with the rs1858830 alleles.

Screening for Variants in the MET Gene. Genomic DNA samples from 40 individuals with autism from the Italian sample and 46 individuals with autism from the Autism Genetic Resource Exchange Consortium were screened for exonic variants. Primers and amplification conditions used to amplify the 21 exons of the MET gene are listed in Table 2. Reveal temperature gradient capillary electrophoresis (SpectruMedix, State College, PA) was used to screen for variants in the exons of the MET gene. Amplicons identified as variant-positive then were directly resequenced to identify the variant.

SNP Genotyping. Genotyping was performed by using TaqMan SNP Genotyping Assays on the ABI Prism 7900HT and analyzed with SDS software. Assays-On-Demand SNP Genotyping were obtained from Applied Biosystems (Foster City, CA). Eight of the nine assays provided reproducible results; the Assay-on-Demand for rs1858830 consistently failed to give reliable genotypes from genomic DNA template. Neither a TaqMan Assay-by-Design nor an Epoch Eclipse Quencher assay (Molecular, San Diego, CA) was able to reliably provide rs1858830 genotype from genomic DNA, probably because of an inability to generate a specific amplicon within this ~85% GC region. We therefore generated a 652-bp amplicon, including rs1858830, from genomic DNA for each sample and used separately generated 652-bp amplicons as templates for Taqman Assay-on-Demand and Epoch Eclipse Quencher genotyping assays. To ensure proper genotype calls, we also genotyped rs1858830 in each sample by using Reveal temperature gradient capillary electrophoresis (SpectruMedix). If inconsistency in any of the three indirect genotyping assays was detected, then the genotype at rs1858830 was determined by direct resequencing.

Corrections for Multiple Comparisons. Appropriate corrections for multiple comparisons are an ongoing debate in human genetics. The presence of two distinct LD blocks indicates that a Bonferroni correction for multiple comparisons of two is appropriate; we consider significant only those associations with P < 0.025 (= 0.05/2). More stringent corrections for multiple comparisons are possible, but do not change the conclusions. An a priori design to independently test simplex and multiplex families as well as a post hoc decision to analyze the data by using two models of association could be argued to bring the appropriate Bonferroni correction factor to 8 (27). Thus, a very stringent correction for multiple comparisons would lead to a significance threshold of P < 0.006 (0.05/8). All associations at rs1858830 exceed this more stringent significance threshold in the original sample, combined sample, and multiplex families.

Transcription Assays. A 762-bp fragment of the MET promoter was cloned into the pGL4.10[luc2] luciferase reporter vector (Promega, Madison, WI). Luciferase assays were conducted by using the Dual-Glo Luciferase Assay kit (Promega) according to the manufacturer’s protocol.

Electrophoretic Mobility Shift and Antibody Supershift Assays. All reactions included double-stranded, 32P-labeled oligonucleotide probe at 50,000 cpm. EMSAs were performed by using the Promega Gel Shift Assay System, according to the manufacturer's
protocol. HeLa nuclear extract was purchased from Promega. Human fetal brain nuclear protein, obtained from a spontaneously aborted 22-week female fetus, was purchased from BioChain Institute, Inc. (Hayward, CA; catalog no. P2244035; lot no. A304059). Nuclear protein (5 μg) was incubated at room temperature either alone or with 100× molar excess unlabeled competitor probe for 20 min before addition of 32P-labeled probe, then incubated an additional 20 min at room temperature before loading on a 4% nondenaturing acrylamide gel. Supershift assays were performed identically except for the addition of a 60-min incubation at 4°C with 2 μg of antibody before loading on the gel.

Supporting Information. See Supporting Text for detailed methods and Figs. 6 and 7 and Table 17, which are published as supporting information on the PNAS web site, for additional data.