

# Genome-wide association identifies candidate genes that influence the human electroencephalogram

Colin A. Hodgkinson<sup>a,1</sup>, Mary-Anne Enoch<sup>a</sup>, Vibhuti Srivastava<sup>a</sup>, Justine S. Cummins-Oman<sup>a</sup>, Cherisse Ferrier<sup>a</sup>, Polina Iarikova<sup>a</sup>, Sriram Sankararaman<sup>b</sup>, Goli Yamini<sup>a</sup>, Qiaoping Yuan<sup>a</sup>, Zhifeng Zhou<sup>a</sup>, Bernard Albaugh<sup>c</sup>, Kenneth V. White<sup>a</sup>, Pei-Hong Shen<sup>a</sup>, and David Goldman<sup>a</sup>

<sup>a</sup>Laboratory of Neurogenetics, National Institute on Alcohol Abuse and Alcoholism, Rockville, MD 20852; <sup>b</sup>Computer Science Department, University of California, Berkeley, CA 94720; and <sup>c</sup>Center for Human Behavior Studies, Weatherford, OK 73096

Edited\* by Raymond L. White, University of California, Emeryville, CA, and approved March 31, 2010 (received for review July 23, 2009)

**Complex psychiatric disorders are resistant to whole-genome analysis due to genetic and etiological heterogeneity. Variation in resting electroencephalogram (EEG) is associated with common, complex psychiatric diseases including alcoholism, schizophrenia, and anxiety disorders, although not diagnostic for any of them. EEG traits for an individual are stable, variable between individuals, and moderately to highly heritable. Such intermediate phenotypes appear to be closer to underlying molecular processes than are clinical symptoms, and represent an alternative approach for the identification of genetic variation that underlies complex psychiatric disorders. We performed a whole-genome association study on alpha ( $\alpha$ ), beta ( $\beta$ ), and theta ( $\theta$ ) EEG power in a Native American cohort of 322 individuals to take advantage of the genetic and environmental homogeneity of this population isolate. We identified three genes (*SGIP1*, *ST6GALNAC3*, and *UGDH*) with nominal association to variability of  $\theta$  or  $\alpha$  power. *SGIP1* was estimated to account for 8.8% of variance in  $\theta$  power, and this association was replicated in US Caucasians, where it accounted for 3.5% of the variance. Bayesian analysis of prior probability of association based upon earlier linkage to chromosome 1 and enrichment for vesicle-related transport proteins indicates that the association of *SGIP1* with  $\theta$  power is genuine. We also found association of *SGIP1* with alcoholism, an effect that may be mediated via the same brain mechanisms accessed by  $\theta$  EEG, and which also provides validation of the use of EEG as an endophenotype for alcoholism.**

alcoholism | electroencephalogram | endophenotype | genetics | whole-genome association

Genetic studies of behavior and psychiatric disorders are hampered by etiologic heterogeneity of these complex phenotypes. Addiction vulnerability arises from both internalizing (emotional) and externalizing (dyscontrol) behavioral dimensions (1), and both of these broad aspects of behavior are strongly influenced by early life trauma and other gene/environment interactions (2). Etiologic heterogeneity dilutes power to detect genetic effects, and is a reason for failures to detect and replicate genome-wide associations (GWAS) in complex disorders. Increasing sample size does not remove underlying heterogeneity and can introduce additional confounds. In neuropsychiatry, these considerations have led to the use of intermediate phenotypes (or endophenotypes) that are heritable, relevant to disease, and have good measurement properties and assay variation more closely related to gene function (3) as surrogates to probe the underlying biology of complex disorders. Risk genes for schizophrenia have recently been identified using quantitative variables derived from functional magnetic resonance imaging in comparatively small cohorts (4). Therefore, an important strategy to increase power for GWAS of psychiatric disorders might be the use of endophenotypes.

The resting electroencephalogram (EEG) is a safely and inexpensively obtained phenotype relevant to normal behavioral variation and to psychiatric disease. The EEG recorded at the scalp is the sum of postsynaptic currents of synchronously depolarized, radially oriented pyramidal cells in the cerebral cortex, and

reflects rhythmic electrical activity of the brain. EEG patterns dynamically and quantitatively index cortical activation, cognitive function, and state of consciousness. EEG traits were among the original intermediate phenotypes in neuropsychiatry, having been first recorded in humans in 1924 by Hans Berger, who documented the  $\alpha$  rhythm, seen maximally during states of relaxation with eyes closed, and supplanted by faster  $\beta$  waves during mental activity. EEG can be used clinically for the evaluation and differential diagnosis of epilepsy and sleep disorders, differentiation of encephalopathy from catatonia, assessment of depth of anesthesia, prognosis in coma, and determination of brain death (5, 6). EEG also has moderate predictive value for personality variation and psychiatric disease including depression (7), bipolar disorder (8), attention-deficit/hyperactivity disorder (9), and obsessive-compulsive disorder (10). Increased  $\beta$  power is associated with both alcoholism and family history of alcoholism (11, 12),  $\theta$  power is altered in alcoholics (13–15), and reduced  $\alpha$  power has been associated with a family history of alcoholism and with alcoholism with comorbid anxiety disorders (16, 17). However, the EEG is not clinically useful for diagnosis of any specific psychiatric disorder.

The stability and heritability of the EEG make it suitable for genetic analysis. Under standardized conditions the resting EEG is a stable trait in healthy adults, with high test-retest correlations [e.g., 0.7 even at >10 months (18)]. The marked interindividual variability in the resting EEG spectral band power is largely genetically determined and heritability of EEG spectral power is uniformly high for all wave forms (19, 20).

We have performed GWAS in a sample of Plains American Indians, in which  $\alpha$  (8–13 Hz),  $\beta$  (13–30 Hz), and  $\theta$  (3–8 Hz) EEG spectral power are moderately heritable with high test-retest correlations over 2 years (15). Notably, this sample represents a population isolate evidencing a small but, as it turned out, useful degree of European admixture, and is genetically distinct from other Native American populations. In this dataset, common familial traits such as alcoholism and other psychiatric disorders do not themselves generate statistical signals approaching genome-wide significance. However, we were able to identify a set of genes and possible pathways that affect  $\alpha$  and  $\theta$  EEG wave forms that are relevant to some of these same complex behavioral traits. Two of the gene associations were replicated in a US Caucasian dataset.

## Results

**Genome-Wide Significant Loci for Resting EEG Power.** Five separate genomic regions, three for  $\theta$  power [Fig. 1A, all on chromosome

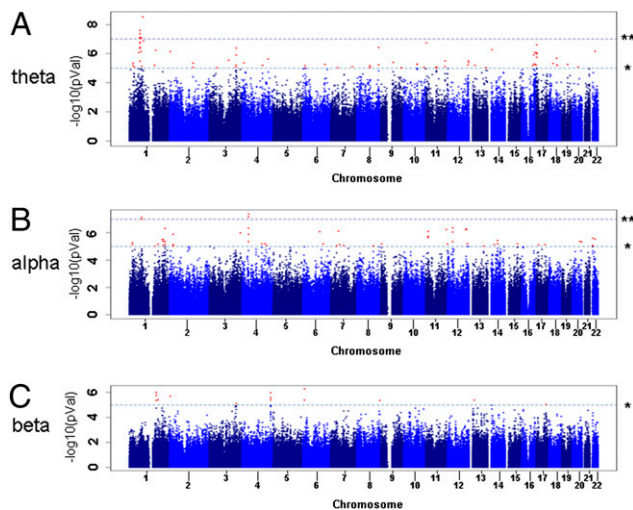
Author contributions: C.A.H., M.-A.E., and D.G. designed research; C.A.H., M.-A.E., J.S.C.-O., C.F., P.I., Z.Z., B.A., and K.V.W. performed research; S.S. contributed new reagents/analytic tools; C.A.H., M.-A.E., V.S., J.S.C.-O., C.F., P.I., G.Y., Q.Y., Z.Z., P.-H.S., and D.G. analyzed data; and C.A.H., M.-A.E., V.S., and D.G. wrote the paper.

The authors declare no conflict of interest.

\*This Direct Submission article had a prearranged editor.

<sup>1</sup>To whom correspondence should be addressed. E-mail: chodg@mail.nih.gov.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.0908134107/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.0908134107/-DCSupplemental).



**Fig. 1.** Manhattan plots for resting EEG spectral power:  $\theta$  (A),  $\alpha$  (B), and  $\beta$  (C). The three plots show association ( $-\log_{10}P$  value) (y axis) for individual SNPs (uncorrected for family structure) against chromosome position (x axis). Threshold  $P$  values of  $1 \times 10^{-7}$  (\*\*) and  $1 \times 10^{-5}$  (\*) are indicated by dotted lines.

(chr) 1] and two for  $\alpha$  power (Fig. 1B, chr 1 and 4), showed genome-wide significant association ( $P < 1.23 \times 10^{-7}$ ) to a single EEG trait (Table 1). No genome-wide significant associations were observed for  $\beta$  power (Fig. 1C) or for the complex psychiatric diagnosis of alcohol use disorder (*SI Text*), demonstrating that random genome-wide significant signals are not generated for common familial traits in a dataset of this size. Quantile-quantile plots show an excess of low  $P$  values for  $\theta$  power but little or no excess for  $\alpha$  power or  $\beta$  power (*SI Text*). Due to the population structure of the Plains Indians there exist close family relationships between study participants confirmed by  $\lambda$  values of 1.3 ( $\alpha$  and  $\theta$ ) and 1.1 ( $\beta$ ), and these could lead to spurious association. Empirically derived  $P$  values (Table 1) that correct for family structure for nominally significant single-nucleotide polymorphisms (SNPs) ( $P < 1 \times 10^{-5}$ )

remained suggestive of association, although no marker remained significant after Bonferroni correction.

**Genomic Regions Associated with  $\theta$  Power.** The three genomic regions associated with  $\theta$  power are all located at chr 1p31, each identifying a single candidate gene (Table 1). In Plains Indians, four significant (and two subthreshold) alleles are represented on two closely related low-frequency haplotypes (*SI Text*) that span almost the whole of the SH3-domain GRB2-like (endophilin)-interacting protein 1 gene (*SGIP1*), and an additional three subthreshold markers lie within another haplotype block that extends 3' from *SGIP1* into the adjacent *Tctex1*-containing 1 gene (*TCTEX1D1*). The association with *SGIP1* accounts for 8.8% of observed variation in  $\theta$  power. Two missense polymorphisms in exon 7 of *SGIP1*, rs17490057 and rs7526812, can be excluded as the functional locus because of their distributions within haplotypes. Four haplotypes, including a common one (frequency = 0.24) carrying the increased  $\theta$  power-associated alleles, are present in HapMap Caucasians (*SI Text*). A single marker significant for  $\theta$  power was found in both ST6 ( $\alpha$ -*N*-acetylneuraminyl-2,3- $\beta$ -galactosyl-1,3)-*N*-acetylgalactosamine- $\alpha$ -2,6-sialyltransferase 3 (*ST6GALNAC3*; rs6696780) and latrophilin 2 (*LPHN2*; rs12145665). *ST6GALNAC3* is an integral Golgi membrane protein which catalyzes the transfer of sialic acids to carbohydrate groups on glycoproteins and glycolipids. *LPHN2* is a G-protein-coupled receptor related to the receptor that binds black widow (*Latrodectus*) spider venom in synaptic membranes (21).

**Genomic Regions Associated with  $\alpha$  Power.** Two genomic regions show significant association to  $\alpha$  power. Although the *ST6GALNAC3* gene overlaps with our findings for  $\theta$  power, the two  $\alpha$  power-associated markers (rs410076 and rs172714) both lie within the third intron of *ST6GALNAC3* and are in linkage disequilibrium with each other, but not with the  $\theta$  power-associated rs6696780 located more than 169 kb away. The signals for association to  $\alpha$  and  $\theta$  power are therefore likely to be independent in this gene region.  $\alpha$  power was also associated with UDP-glucose dehydrogenase gene (*UGDH*) through two markers, rs7667766 and rs6817264, that are located in introns 1 and 9, respectively. An additional three SNPs adjacent to this region that show sub-

**Table 1.** Markers with significant association to EEG traits accompanied by subthreshold markers within significant regions

EEG frequency band	SNP	Location	Position	Gene	HWE $P$ value	Plains Indians		Plains Indians corrected $P$ value*	Caucasians $P$ value
						minor allele frequency	Plains Indians $P$ value		
Theta	rs6588207	1p31.3	66839146	<i>SGIP1</i>	0.5972	0.03	$4.24 \times 10^{-8}$	$7.1 \times 10^{-6}$	0.056
Theta	rs10889635	1p31.3	66848163	<i>SGIP1</i>	0.5526	0.03	$2.52 \times 10^{-8}$	$4.9 \times 10^{-6}$	0.013
Theta	rs6656912	1p31.3	66856259	<i>SGIP1</i>	1	0.02	$4.24 \times 10^{-7}$	$1.61 \times 10^{-5}$	ND
Theta	rs6681460	1p31.3	66895645	<i>SGIP1</i>	0.4659	0.03	$7.30 \times 10^{-8}$	$1.3 \times 10^{-5}$	ND
Theta	rs10789215	1p31.3	66923773	<i>SGIP1</i>	1	0.03	$8.51 \times 10^{-8}$	$1.2 \times 10^{-5}$	NS
Theta	rs2146904	1p31.3	66932924	<i>SGIP1</i>	1	0.03	$4.37 \times 10^{-7}$	$3.63 \times 10^{-5}$	ND
Theta	rs536410	1p31.3	66963672	<i>SGIP1</i>	0.8271	0.04	$1.69 \times 10^{-7}$	$1.43 \times 10^{-5}$	ND
Theta	rs2483704	1p31.3	66967941	<i>SGIP1</i>	1	0.04	$3.63 \times 10^{-7}$	$2.59 \times 10^{-5}$	ND
Theta	rs2916	1p31.3	66989285	<i>TCTEX1D1</i>	0.6424	0.03	$1.87 \times 10^{-7}$	$1.92 \times 10^{-5}$	ND
Theta	rs6696780	1p31.1	76504495	<i>ST6GALNAC3</i>	0.2083	0.02	$8.47 \times 10^{-8}$	$3.89 \times 10^{-5}$	ND
Alpha	rs410076	1p31.1	76774219	<i>ST6GALNAC3</i>	0.6564	0.35	$1.03 \times 10^{-7}$	$1.1 \times 10^{-6}$	NS
Alpha	rs172714	1p31.1	76787701	<i>ST6GALNAC3</i>	0.2563	0.35	$7.34 \times 10^{-8}$	$3.0 \times 10^{-7}$	NS
Theta	rs1245665	1p31.1	81800834	<i>LPHN2</i>	1	0.02	$3.00 \times 10^{-9}$	$7.1 \times 10^{-6}$	NS
Alpha	rs2608830	4p13	39135549	<i>RPL9/ILIAS</i>	0.5768	0.11	$1.29 \times 10^{-6}$	$6.8 \times 10^{-6}$	ND
Alpha	rs2259073	4p13	39135549	<i>UGDH</i>	0.4204	0.11	$4.55 \times 10^{-7}$	$2.5 \times 10^{-6}$	ND
Alpha	rs2687964	4p13	39151469	<i>UGDH</i>	0.6905	0.10	$4.54 \times 10^{-6}$	$1.39 \times 10^{-5}$	ND
Alpha	rs7667766	4p13	39172667	<i>UGDH</i>	0.6632	0.11	$7.04 \times 10^{-8}$	$8.0 \times 10^{-7}$	NS
Alpha	rs6827264	4p13	39194505	<i>UGDH</i>	0.5355	0.11	$4.23 \times 10^{-8}$	$7.0 \times 10^{-7}$	NS

HWE, Hardy-Weinberg equilibrium; ND, not determined; NS, not significant.

\*Empirical  $P$  values correcting for participant relatedness.

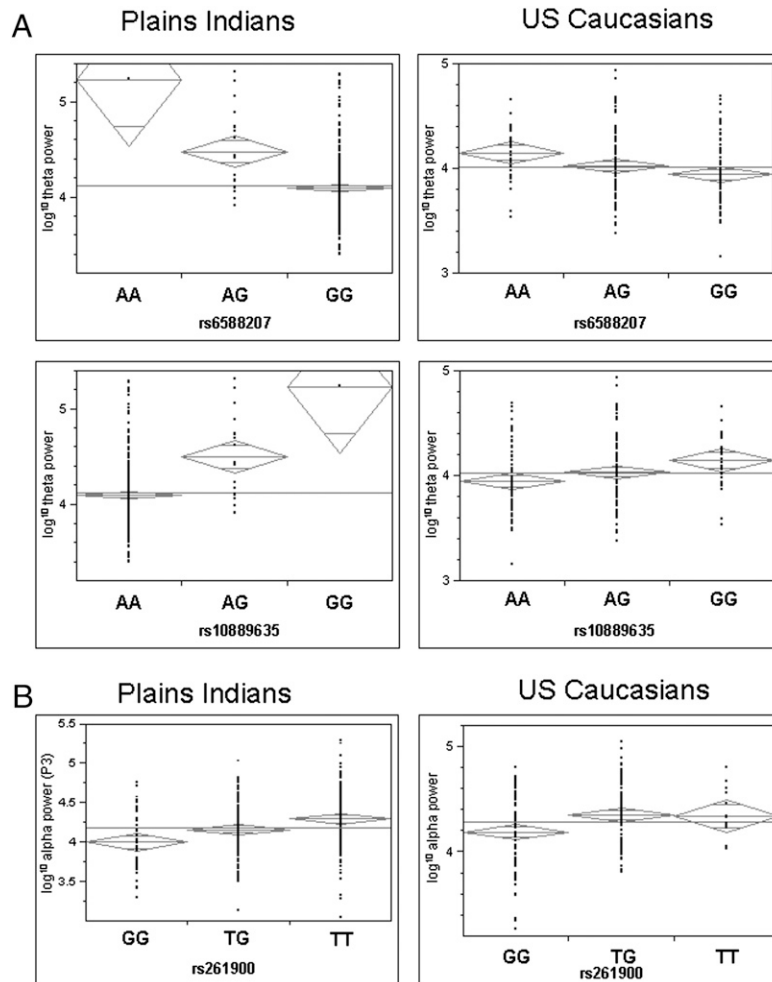
threshold association to  $\alpha$  power are located in the 5'-regulatory region of *UGDH*, and in the adjacent lipoic acid synthetase (*LIAS*) and ribosomal protein-like 9 (*RPL9*) genes. All five SNPs are within a single haplotype block extending almost 1 Mb (*SI Text*) and encompassing *UGDH*, *LIAS*, and *RPL9*, along with klotho  $\beta$  (*KLB*), a component of the receptor for fibroblast growth factor 19 (*FGF19*). *UGDH* is an integral Golgi membrane protein involved in posttranslational modification of extrinsic proteins, whose expression is up-regulated in response to TGF- $\beta$  and hypoxia, a risk factor for schizophrenia.

**Validation of *SGIP1* Association.** We estimated posterior probabilities of association (PPAs) for the top two hits in *SGIP1*: rs6588207 and rs10889635. Posterior probability of association is based on Bayesian method and can be interpreted directly as a probability (22). The  $\log_{10}(\text{BF})$  for the combination of associated markers was 4.11 and for rs10889635 was 4.40 [ $\sigma_a$  and  $\sigma_d$  as described by Servin and Stephens (23) for phenotypes normally distributed across genotypes]. We assumed a value of  $10^{-4}$  for  $\pi$  (prior probability of  $H_1$ ), a choice based on earlier evidence of linkage of EEG variation to this chr 1 region in the Collaborative Study on the Genetics of Alcoholism (COGA) dataset, and nominally significant enrichment for  $\theta$  power-associated proteins that mediate vesicle transport ( $P = 0.040$ ) based on gene ontology. A moderate PPA of 0.56

was observed for the region whereas rs10889635 showed a PPA of 0.72, indicating that this region is associated with  $\theta$  power among Plains Indians, justifying analysis of these markers in an independent sample after consideration of PPAs under defined prior assumptions.

**Replication in the United States. Caucasians.** The five genome-wide significant regions were tested in an independent US Caucasian dataset. Association of increased  $\theta$  power to the G allele of rs10889635 within *SGIP1* was replicated ( $P = 0.013$ ) (Fig. 2A), accounting for 3.5% of the variance in  $\theta$  power in this population, although the association does not remain significant after correction for multiple testing. Within *SGIP1*, the A allele of rs6588207, significant in Plains Indians, showed a similar trend ( $P = 0.056$ ) but did not reach significance. No *ST6GALNAC3*, *LPHN2*, or *UGDH* markers showed association in Caucasians.

**Subthreshold Associations with  $\theta$  and  $\alpha$  Power.** Other genomic regions showed association in the range of  $P = 1 \times 10^{-5}$ – $1.22 \times 10^{-7}$ . It is likely that several represent chance findings, because either no annotated gene is present in the region or the gene would not appear to affect neuronal reactivity (*SI Text*). All potential candidate genes identified for each EEG spectral power (at significant or subthreshold level) are listed by function in *SI Text*.



**Fig. 2.** Replication of EEG/SNP associations in Caucasians. (A) One-way analysis of variance (ANOVA) for rs6588207 and rs10889635 in *SGIP1*, against ( $-\log_{10}$ ) $\theta$  power averaged across posterior electrodes. The G allele of rs10889635 showed association to increased  $\theta$  power in both Plains Indians ( $P = 2.52 \times 10^{-8}$ ) and Caucasians ( $P = 0.013$ ). The A allele of rs6588207 showed association to increased  $\theta$  power in Plains Indians ( $P = 4.24 \times 10^{-8}$ ) with a similar trend in Caucasians ( $P = 0.056$ ). (B) One-way ANOVA for rs261900 in *BICD1*, against ( $-\log_{10}$ )  $\alpha$  power in Plains Indians (at the P3 electrode) and Caucasians ( $-\log_{10}$ )  $\alpha$  power averaged across posterior electrodes. The T allele showed association to increased  $\alpha$  power in both Plains Indians ( $P = 9.76 \times 10^{-9}$ ) and Caucasians ( $P = 0.0023$ ).

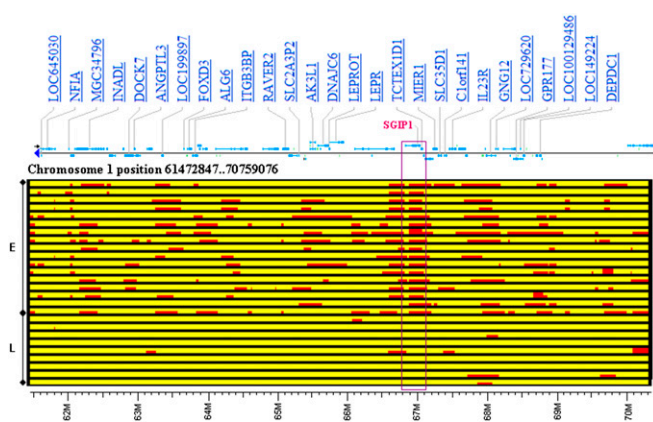
**Golgi Transport Candidates Among Subthreshold Signals.** Analysis of the gene ontology for all candidate genes suggested that a proportion of the genes are involved in similar cellular processes (*SI Text*). Like *SGIP1*, four of these genes, *BICD1*, *FAM125B*, *ANKRD27*, and *C11orf2*, are either involved in retrograde Golgi transport or at least contain a domain similar to those found in vesicular transport proteins. Additionally, *RTN1*, *RPH3AL*, and *CECR2* are involved in vesicle-mediated transport and neuroendocrine processes. *BICD1* (bicaudal D homolog 1) is particularly interesting in that although there was only subthreshold association [rs261900:  $P = 4.32 \times 10^{-7}$  (uncorrected)/ $4.5 \times 10^{-6}$  (empirical)] to  $\alpha$  power across posterior scalp electrodes, the T allele of rs261900 is also associated ( $P = 0.0023$ ) with increased  $\alpha$  power in Caucasians (Fig. 2B).

**$\beta$  EEG Power.** Several candidate genes associated to  $\beta$  power have functions that may affect neuronal electrical activity neurons. These include the  $\alpha 3$  glycine receptor gene (*GLRA3*), a brain-expressed synaptic receptor that inhibits neuroexcitability, with four SNPs in a single haplotype block. rs12027066 is near *ALD9A* (aldehyde dehydrogenase, 9 family, member A1, an enzyme involved in GABA synthesis). Genetic variation in GABA receptors has been associated with disorders for which EEG has been proposed as an intermediate phenotype, including alcoholism (24) and schizophrenia (25).

#### ***SGIP1* Association with $\theta$ Power Is Revealed by Admixture Mapping.**

The Plains Indian population was selected because of its characteristics as a population isolate with low admixture (European admixture; mean 5.23% and median 1.6%) and is genetically distinct even from other Indian tribes (*SI Text*). However, despite the isolate character of Plains Indians, analysis of the  $\theta$  power association on chr 1 suggests that this signal was detected via admixture mapping, making it feasible to identify the introgressed European-derived chromosomal region and extended haplotypes harboring the putative functional locus. The minor allele frequency (MAF) in Plains Indians for the significant SNP rs10889635 is low ( $>0.03$ ) compared with Caucasians (MAF = 0.41). When MAF is plotted against European ethnic factor score, a measure of European ancestry, the frequency of the minor G allele increases with increasing European admixture (*SI Text*), an effect seen also for  $\theta$  power-associated markers in *ST6GALNAC3* and *LPHN2*, and to a lesser extent in the more distant *SEPI5* gene. Localized admixture for the region surrounding *SGIP1* on chr 1 was assessed for the 17 Plains Indians carrying the G allele of rs10889635 and 9 Plains Indians with low European admixture (mean = 1.3%, median = 1.0%). In all carriers of the G allele, at least one copy of *SGIP1* was European-derived (Fig. 3). The relatively small size of the computed blocks of admixture suggests that introduction of European DNA into this population is not recent (26). The association of *SGIP1* to  $\theta$  power is unlikely to be a stratification artifact because there is no correlation between degree of European ancestry and EEG power in any of the frequency bands (*SI Text*), and because the association at *SGIP1* was replicated in Caucasians. Instead, this appears to be a serendipitous example of admixture mapping within a GWAS. The nonreplicating associations of *LPHN2* and *ST6GALNAC3* to  $\theta$  power are likely to arise as a consequence of this admixture, and probably do not represent true associations. For  $\alpha$  power, European ancestry does not predict the associated minor allele frequencies in *ST6GALNAC3* and *UGDH* (*SI Text*), indicating these associations may be valid but population-specific.

**EEG as an Endophenotype for Alcoholism.** To evaluate whether *SGIP1* is a candidate gene for alcoholism, in addition to EEG variation we analyzed significant *SGIP1* markers for association with alcohol use disorders (AUD) in the Plains Indian cohort (Table 2). Alleles significant (but subthreshold) for association with increased  $\theta$  power were underrepresented in individuals with AUD. Reduced  $\theta$  power in alcoholics had been reported for this



**Fig. 3.** Local admixture of chr 1 region (61.47–70.76 Mb) containing *SGIP1* in 17 Plains Indians (E) carrying the  $\theta$  power-associated G allele of rs10889635 compared with 9 Plains Indians (L) with low European admixture. The analysis shows a common European chromosomal segment at *SGIP1* in all Plains Indians carrying the G allele of rs10889635, indicating that the association signal ( $P = 2.52 \times 10^{-8}$ ) is in part due to admixture at the *SGIP1* locus. The only Plains Indian homozygous for the G allele of rs10889635 is indicated (t).

dataset (15). The present data indicate a relationship between *SGIP1* genotype and AUD. Significant SNPs at other loci did not show association to AUD except for SNPs at *LPHN2* and *ST6GALNAC3* that were linked to rs1088935 by their common European ancestry. Moreover, strength of these associations weakened with increasing distance from *SGIP1*.

#### **Discussion**

This GWAS in Plains American Indians identified two chromosomal regions, both previously implicated in linkage studies, containing multiple candidate genes for variability in resting EEG power. *SGIP1* and *ST6GALNAC3* lie within a linkage peak on chr 1p (95–115 cM), linked to  $\beta 2$  EEG spectral power (16–20 Hz) in the COGA (27, 28). *UGDH*, *LIAS*, *RPL9*, and *KLB* lie in a region of chr 4 previously linked to EEG in this Plains Indian dataset (15), together with an additional Native American population (29) and in the COGA dataset (30). Additionally, the study has identified *BICD1* located on chr 12p11 as a candidate gene for resting  $\alpha$  power variability.

Linkage studies have identified other chromosome regions potentially harboring genes involved in EEG variability. Low-voltage  $\alpha$  power mapped to chr 20q (31). Convergence of linkage was found for  $\alpha$ ,  $\beta$ , and  $\theta$  resting EEG power to chr 5q13-14 in this Plains Indian dataset, suggesting common loci that regulate EEG across the frequency spectrum (15). This finding was not replicated in this GWAS, most likely because of the relatively small effect of this locus on EEG power. In the present study, failure to replicate associations of *ST6GALNAC3* and *UGDH* in Caucasians may reflect cross-population heterogeneity in loci influencing the EEG. With regard to this point, the haplotype block structure of the chr 4 region containing *UGDH* is more complex in Caucasians compared with Plains Indians, and it is possible that the association signal arises from variation in other genes in the region including *KLB* and *LIAS*. *KLB* is an interesting candidate because FGF-19 signaling affects cortical development (32) and voltage-gated  $\text{Na}^+$  channels, thereby altering neuronal reactivity (33).

The EEG, although an intermediate phenotype, is itself a complex trait reflecting the rhythmic electrical activity of the brain, which is probably modulated by many loci. Discovery of a single locus accounting for 8.8% of variance in  $\theta$  EEG is therefore fortuitous. In US Caucasians this locus accounts for 3.5% of variance, the diminution due to the increased frequency of the informative alleles and relative haplotype frequencies.

**Table 2. Comparison of minor allele frequency for seven *SGIP1* SNPs between individuals with alcohol use disorder and controls shows significant underrepresentation of alleles associated with increased  $\theta$  power**

	rs6588207	rs10889635	rs66881460	rs10789215	rs2146904	rs536410	rs2483704
Alcohol use disorder	0.018	0.018	0.018	0.018	0.018	0.020	0.018
Controls	0.054	0.051	0.045	0.039	0.051	0.067	0.061
<i>P</i> value, 1 degree of freedom	0.005	0.009	0.028	0.076	0.009	0.001	0.002
$\chi^2$	7.843	6.78	4.824	3.139	6.87	10.903	10.024

The data from the significant and subthreshold genes suggest that many of the genes involved in determining EEG variability are involved in intracellular transport processes rather than cell-surface receptor activity, although two subthreshold genes were involved in glycine ion channel and GABAergic function. *SGIP1* is directly involved in vesicle formation at the plasma membrane through interactions with phospholipids and EPS15, an adaptor protein. Knockdown of *SGIP1* expression reduces clathrin-mediated endocytosis (34). Alterations in the efficiency of endocytosis might change neuronal activity via receptor turnover or neurotransmitter uptake. The rabphilin 3A-like (without C2 domains) gene (*RPHAL*) generated only a modest *P* value [ $6.2 \times 10^{-6}$  (uncorrected)/ $1.2 \times 10^{-5}$  (empirical)] but is involved in exocytosis and vesicle transport, and is on chr 17 located within a linkage peak identified for the maximum-drinks phenotype in the COGA study (35). Coincidentally, D17S1308, which gave the strongest linkage signal, maps adjacent to the vacuolar protein-sorting 53 homolog gene (*VPS53*), a gene involved in retrograde vesicle trafficking in the late Golgi. *BICD1* is also involved in vesicle trafficking and is essential for retrograde Golgi-to-endoplasmic reticulum transport via interaction with the adaptor protein Rab6a. Disruption of *BICD1*/Rab6a interaction uncouples vesicles from the dynein-dynactin transport machinery, resulting in the accumulation of vesicles at the cell periphery (36). *BICD1* also appears to have an important function in nuclear positioning in neurons during development. The involvement of *BICD1* with EEG variability is further suggested because *BICD1* interacts directly with *LIS1* (37), a protein that also interacts directly with *DISC1* (38). Variation at *DISC1* has been associated with alteration in the P300-evoked potential, another phenotype of neuronal reactivity (39).

Two other subthreshold genes, cadherin 13 (*CDH13*) and fibroblast growth factor 14 (*FGF14*), were identified in a GWAS for nicotine dependence (40) and more recently in a GWAS for alcohol dependence (41). *FGF14* potentially alters neuronal reactivity through interaction with voltage-gated  $\text{Na}^+$  channels (33). Analysis of response to contextual imagery in cigarette smokers suggested that alteration of EEG activity may be a common feature associated with craving (42). The adaptor-related protein complex 2,  $\alpha$  subunit 2 gene (*AP2A2*), which encodes another protein involved in endocytosis, is located close to D11S1984, which generated the maximum linkage signal to alcohol dependence in a southwestern Native American population (29). Although it appears that our main findings are for genes that solely influence EEG power, convergences of findings at the subthreshold level to previous results in addictions suggest that by using an intermediate phenotype we potentially identified genes that have a general influence on addiction.

The Plains Indians are a population isolate, offering advantages of reduced genetic and environmental heterogeneity. As described, this is a relatively nonadmixed tribe, with median European admixture of 1.5%. Although it is a family-based sample (*SI Text*), we were able to correct our analyses for the relatedness of the participants.

In conclusion, this GWAS suggests that neuronal excitability in the brain is determined in part by the ability to recycle neuronal membrane components. Replication of association of *SGIP1* to  $\theta$  power validates the use of the GWAS approach on intermediate phenotype datasets of modest size. The success of this study in identifying novel genes and cellular processes in determining EEG traits demonstrates that if there are loci of

moderate effect size, which is more likely for an intermediate phenotype in a population isolate, it may be unnecessary to have datasets numbering in the thousands, and that the analysis of accurately measured, heritable traits in relatively homogeneous populations is an informative alternative in the genetic analysis of complex psychiatric diseases and behavior.

## Methods

**Participant Recruitment. Plains American Indians.** Participants were recruited from a tribe in rural Oklahoma (15) (fully described in *SI Text*). Written informed consent was obtained according to a human research protocol approved by the Tribal Council and the human research committee of the National Institute on Alcohol Abuse and Alcoholism. EEG was obtained from 359 people. The alcohol use disorder (AUD) cohort comprised 225 cases and 156 controls. AUD was defined as DSM-III alcohol abuse (without dependence) (222 individuals) or DSM-III alcohol dependence (3 individuals).

The replication dataset (described in *SI Text*) consisted of 185 European Americans (104 women and 81 men: 61 healthy controls and 124 individuals with a variety of DSM-III diagnoses) recruited in Bethesda, MD (43). Written informed consent was obtained via a human research protocol approved by the NIAAA institutional review board.

**EEG Acquisition and Analysis.** Details of EEG methods and analysis are in Enoch et al. (15) and *SI Text*. Resting EEG was recorded using a fitted electrode cap in a field setting while the subject was seated with eyes closed and relaxed in a darkened noise-baffled room. To balance data quality against participant time, data were collected from six scalp electrodes (one frontal, two occipital, three parietal) selected to maximize information for  $\alpha$  and  $\theta$  power. Spectral power was determined for  $\theta$  (3–8 Hz),  $\alpha$  (8–13 Hz), and  $\beta$  (13–30 Hz), and log-transformed to normalize the distribution. For each frequency band, power was averaged for the five posterior electrodes to reduce variability due to individual electrode measurement and was justified by high correlations between  $\alpha$  and  $\theta$  power at the selected electrodes (15), coherence between these signals (44), and the higher  $\theta$  and  $\alpha$  activity in posterior and occipital regions (45). These data were used for association (see *SI Text* for means and ranges). No outliers were removed.

**Genotyping.** Samples (399) were genotyped using the Illumina HapMap500K. Fourteen were discarded due to low genotype call rate (<97%), and four were discarded after gender and allele transmission testing. Genotypes were called using BeadStudio 3.2 (Illumina). Thirty-five replicate pairs showed average genotype reproducibility of 0.99994. Following quality control, association with  $\log_{10}$  EEG power was performed on 322 individuals (137 male, 185 female) for whom EEG was available. Replication genotyping was performed by 5'-exonuclease assay.

**Statistical Analysis.** Association was tested using the standard linear regression model in PLINK (46). SNPs with MAF < 1% or a call rate < 90% were excluded, leaving 405,281 SNPs. Threshold for association was set at  $<1.23 \times 10^{-7}$  using the Bonferroni method. Posthoc analyses using sex and age as covariates both individually and together did not influence associations. *P* value inflation due to genetic relatedness was estimated using the GC program (47). To correct for family structure, QFAM family-based association tests for SNPs determined to be nominally significant were performed within PLINK with  $10^7$  rounds of permutation, yielding empirical *P* values. Following GWAS, we estimated PPAs for the two top *SGIP1* markers, rs6588207 and rs10889635. Bayes factors (BF) were calculated using Bayesian Imputation-Based Association Mapping (<http://quartus.uchicago.edu/~yguan/bimbam>). PPAs were calculated as described in Stephens and Balding (22). Enrichment in gene ontology terms was detected using GOTM (48).

**Local Admixture.** Localized admixture on chr 1 was detected using Local Ancestry in admixed Populations (LAMP) (49) v2.0 with an  $r^2$  cutoff of 0.95 to minimize exclusion of informative SNPs due to variable LD between Euro-

peans and American Indians. Computational runs using generation time estimates of 3, 5, 7, and 9 yielded similar estimates of local admixture. A recombination rate of  $1 \times 10^{-8}$  and seed  $\alpha$  values of 0.04 (Europeans) and 0.96 (American Indians) were used.

**Population Substructure and Admixture Determination.** Ancestry informative markers (186) (50) were typed in the EEG samples and the HGDP-CEPH Human Genome Diversity Cell Line Panel (<http://www.cephb.fr/HGDP-CEPH-Panel>). PHASE

Structure 2.2 (<http://pritch.bsd.uchicago.edu/software.html>) was run using ancestry informative marker data from the EEG and CEPH datasets simultaneously to identify population substructure and compute individual ethnic factor scores.

**ACKNOWLEDGMENTS.** We thank the study participants for their contributions, and Longina Akhtar and Elisa Moore for technical assistance. This study was funded by the intramural program of National Institute on Alcohol Abuse and Alcoholism.

1. Kendler KS, et al. (2008) The structure of genetic and environmental risk factors for DSM-IV personality disorders: A multivariate twin study. *Arch Gen Psychiatry* 65: 1438–1446.
2. Caspi A, et al. (2002) Role of genotype in the cycle of violence in maltreated children. *Science* 297:851–854.
3. Gottesman II, Gould TD (2003) The endophenotype concept in psychiatry: Etymology and strategic intentions. *Am J Psychiatry* 160:636–645.
4. Potkin SG, et al.; FBIRN (2009) A genome-wide association study of schizophrenia using brain activation as a quantitative phenotype. *Schizophr Bull* 35:96–108.
5. Niedermeyer E (1993) The normal EEG of the waking adult. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, ed Lopez de Silva F (Williams and Wilkins, Baltimore), pp 97–117.
6. Nunez PL, Srinivasan R (2005) *Electric Fields of the Brain* (Oxford Univ Press, New York), 2nd Ed, p 9.
7. Knott V, Mahoney C, Kennedy S, Evans K (2001) EEG power, frequency, asymmetry and coherence in male depression. *Psychiatry Res* 106:123–140.
8. Cook BL, Shukla S, Hoff AL (1986) EEG abnormalities in bipolar affective disorder. *J Affect Disord* 11:147–149.
9. Barry RJ, Clarke AR, Johnstone SJ (2003) A review of electrophysiology in attention-deficit/hyperactivity disorder: I. Qualitative and quantitative electroencephalography. *Clin Neurophysiol* 114:171–183.
10. Pogarell O, et al. (2006) Symptom-specific EEG power correlations in patients with obsessive-compulsive disorder. *Int J Psychophysiol* 62:87–92.
11. Rangaswamy M, et al. (2002) Beta power in the EEG of alcoholics. *Biol Psychiatry* 52: 831–842.
12. Pollock VE, Earleywine M, Gabrielli WF (1995) Personality and EEG beta in older adults with alcoholic relatives. *Alcohol Clin Exp Res* 19:37–43.
13. Rangaswamy M, et al. (2003) Theta power in the EEG of alcoholics. *Alcohol Clin Exp Res* 27:607–615.
14. Coutin-Churchman P, Moreno R, Añez Y, Vergara F (2006) Clinical correlates of quantitative EEG alterations in alcoholic patients. *Clin Neurophysiol* 117:740–751.
15. Enoch MA, et al. (2008) Common genetic origins for EEG, alcoholism and anxiety: The role of CRH-BP. *PLoS One* 3:e3620.
16. Finn PR, Justus A (1999) Reduced EEG alpha power in the male and female offspring of alcoholics. *Alcohol Clin Exp Res* 23:256–262.
17. Enoch MA, et al. (1999) Association of low-voltage alpha EEG with a subtype of alcohol use disorders. *Alcohol Clin Exp Res* 23:1312–1319.
18. Salinsky MC, Oken BS, Morehead L (1991) Test-retest reliability in EEG frequency analysis. *Electroencephalogr Clin Neurophysiol* 79:382–392.
19. Smit DJ, Posthuma D, Boomsma DI, Geus EJ (2005) Heritability of background EEG across the power spectrum. *Psychophysiology* 42:691–697.
20. van Beijsterveldt CE, Molenaar PC, de Geus EJ, Boomsma DI (1996) Heritability of human brain functioning as assessed by electroencephalography. *Am J Hum Genet* 58:562–573.
21. Lang J, Ushkaryov Y, Grasso A, Wollheim CB (1998)  $Ca^{2+}$ -independent insulin exocytosis induced by  $\alpha$ -latrotoxin requires latrophilin, a G protein-coupled receptor. *EMBO J* 17:648–657.
22. Stephens M, Balding DJ (2009) Bayesian statistical methods for genetic association studies. *Nat Rev Genet* 10:681–690.
23. Servin B, Stephens M (2007) Imputation-based analysis of association studies: Candidate regions and quantitative traits. *PLoS Genet* 3:e114.
24. Enoch MA (2008) The role of GABA(A) receptors in the development of alcoholism. *Pharmacol Biochem Behav* 90:95–104.
25. Lang UE, Puls I, Muller DJ, Strutz-Seebohm N, Gallinat J (2007) Molecular mechanisms of schizophrenia. *Cell Physiol Biochem* 20:687–702.
26. Stephens JC, Briscoe D, O'Brien SJ (1994) Mapping by admixture linkage disequilibrium in human populations: Limits and guidelines. *Am J Hum Genet* 55: 809–824.
27. Ghosh S, et al. (2003) Linkage mapping of beta 2 EEG waves via non-parametric regression. *Am J Med Genet B Neuropsychiatr Genet* 118B:66–71.
28. Porjesz B, et al. (2002) Linkage disequilibrium between the beta frequency of the human EEG and a GABA<sub>A</sub> receptor gene locus. *Proc Natl Acad Sci USA* 99:3729–3733.
29. Long JC, et al. (1998) Evidence for genetic linkage to alcohol dependence on chromosomes 4 and 11 from an autosome-wide scan in an American Indian population. *Am J Med Genet* 81:216–221.
30. Porjesz B, et al. (2002) Linkage and linkage disequilibrium mapping of ERP and EEG phenotypes. *Biol Psychol* 61:229–248.
31. Steinlein O, Anokhin A, Yping M, Schalt E, Vogel F (1992) Localization of a gene for the human low-voltage EEG on 20q and genetic heterogeneity. *Genomics* 12:69–73.
32. Borello U, et al. (2008) FGF15 promotes neurogenesis and opposes FGF8 function during neocortical development. *Neural Dev* 3:17.
33. Laezza F, et al. (2007) The FGF14(F145S) mutation disrupts the interaction of FGF14 with voltage-gated Na<sup>+</sup> channels and impairs neuronal excitability. *J Neurosci* 27: 12033–12044.
34. Uezu A, et al. (2007) SGIP1 $\alpha$  is an endocytic protein that directly interacts with phospholipids and Eps15. *J Biol Chem* 282:26481–26489.
35. Saccone NL, et al. (2000) A genome screen of maximum number of drinks as an alcoholism phenotype. *Am J Med Genet* 96:632–637.
36. Matanis T, et al. (2002) Bicaudal-D regulates COPI-independent Golgi-ER transport by recruiting the dynein-dynactin motor complex. *Nat Cell Biol* 4:986–992.
37. Claussen M, Suter B (2005) BicD-dependent localization processes: From *Drosophila* development to human cell biology. *Ann Anat* 187:539–553.
38. Brandon NJ, et al. (2004) Disrupted in Schizophrenia 1 and Nudel form a neurodevelopmentally regulated protein complex: Implications for schizophrenia and other major neurological disorders. *Mol Cell Neurosci* 25:42–55.
39. Blackwood DH, Muir WJ (2004) Clinical phenotypes associated with DISC1, a candidate gene for schizophrenia. *Neurotox Res* 6:35–41.
40. Uhl GR, et al. (2007) Molecular genetics of nicotine dependence and abstinence: Whole genome association using 520,000 SNPs. *BMC Genet* 8:10.
41. Treutlein J, et al. (2009) Genome-wide association study of alcohol dependence. *Arch Gen Psychiatry* 66:773–784.
42. Knott V, et al. (2008) EEG correlates of imagery-induced cigarette craving in male and female smokers. *Addict Behav* 33:616–621.
43. Enoch MA, et al. (1999) Association of low-voltage alpha EEG with a subtype of alcohol use disorders. *Alcohol Clin Exp Res* 23:1312–1319.
44. Winterer G, et al. (2003) EEG phenotype in alcoholism: Increased coherence in the depressive subtype. *Acta Psychiatr Scand* 108:51–60.
45. Porjesz B, et al. (2005) The utility of neurophysiological markers in the study of alcoholism. *Clin Neurophysiol* 116:993–1018.
46. Purcell S, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575.
47. Devlin B, Bacanu SA, Roeder K (2004) Genomic Control to the extreme. *Nat Genet* 36: 1129–1130, author reply 1131.
48. Zhang B, Schmoyer D, Kirov S, Snoddy J (2004) GOTree Machine (GOTM): A web-based platform for interpreting sets of interesting genes using gene ontology hierarchies. *BMC Bioinformatics* 5:16.
49. Sankararaman S, Sridhar S, Kimmel G, Halperin E (2008) Estimating local ancestry in admixed populations. *Am J Hum Genet* 82:290–303.
50. Hodgkinson CA, et al. (2008) Addictions biology: Haplotype-based analysis for 130 candidate genes on a single array. *Alcohol* 43:505–515.