Published in final edited form as:
Am J Med Genet B Neuropsychiatr Genet. 2011 March ; 156(2): 125–138. doi:10.1002/ajmg.b.31143.

"Replicated" genome wide association for dependence on illegal substances: genomic regions identified by overlapping clusters of nominally positive SNPs

Tomas Drgon, **Catherine Johnson**, **Michelle Nino**, **Jana Drgonova**, **Donna Walther**, and **George R Uhl***
Molecular Neurobiology Branch, NIH-IRP, NIDA, Baltimore, Maryland

Abstract

Declaring "replication" from results of genome wide association (GWA) studies is straightforward when major gene effects provide genome-wide significance for association of the same allele of the same SNP in each of multiple independent samples. However, such unambiguous replication may be unlikely when phenotypes display polygenic genetic architecture, allelic heterogeneity, locus heterogeneity and when different samples display linkage disequilibria with different fine structures. We seek chromosomal regions that are tagged by clustered SNPs that display nominally-significant association in each of several independent samples. This approach provides one "nontemplate" approach to identifying overall replication of groups of GWA results in the face of difficult genetic architectures. We apply this strategy to 1M SNP Affymetrix and Illumina GWA results for dependence on illegal substances. This approach provides high confidence in rejecting the null hypothesis that chance alone accounts for the extent to which clustered, nominally-significant SNPs from samples of the same racial/ethnic background identify the same chromosomal regions. There is more modest confidence in: a) identification of individual chromosomal regions and genes and b) overlap between results from samples of different racial/ethnic backgrounds. The strong overlap identified among the samples with similar racial/ethnic backgrounds, together with prior work that identified overlapping results in samples of different racial/ethnic backgrounds, support contributions to individual differences in vulnerability to addictions that come from both relatively older allelic variants that are common in many current human populations and newer allelic variants that are common in fewer current human populations.

Keywords

substance dependence; microarray

Introduction

Genome wide association (GWA) is a method of choice for identifying genes whose variants influence vulnerability to complex disorders. Declaring "replication" of individual results of genome wide association studies is straightforward when major gene effects provide associations between marker and phenotype that display the same phase and "genome wide" levels of significance (p *ca* 10^{-8}) in each of several independent samples. However, such

*To whom correspondence should be addressed at: Molecular Neurobiology, Box 5180, Baltimore, MD 21224 phone: (410) 550-2843 x 146, fax: (410) 550-1535, guhl@intra.nida.nih.gov.

Financial Disclosure
Authors report no biomedical financial interests or potential conflicts of interest

"template" replication for individual markers is unlikely to be achieved in many otherwise-reasonable samples for many phenotypes. Phenotypes and samples that display polygenic genetic architecture, allelic heterogeneity, locus heterogeneity and sample to sample differences in fine structure of linkage disequilibrium can provide especial difficulties for this "template" approach. These difficulties can be exacerbated when data comes from different genotyping platforms that do not assess allele frequencies for identical sets of SNPs. Much current genome wide association and linkage data suggests that we may have identified many or even most of the loci at which we might expect "template" analyses to identify reproducible genome wide significance in reasonably sized samples *(see references below)*. Much of the risk attributable to genetic influences on common phenotypes may well arise from polygenic influences whose properties are likely to provide many false negative results in searches for replicated "genome wide" significance in multiple independent samples, using "template" criteria for replication.

Vulnerability to heavy use and development of dependence on an illegal substance ("addiction vulnerability") appears to be such a trait. The substantial genetic influences on addiction vulnerability are documented by data from family, adoption and twin studies (Karkowski and others 2000; True and others 1999; Tsuang and others 1998; Uhl and others 1995). Twin studies also document shared heritable influences on vulnerability to dependence on addictive substances from different pharmacological classes (*eg* opiates and stimulants) (Karkowski and others 2000; Kendler and others 2000; Tsuang and others 1998). Combined data from linkage and initial GWA studies (Bierut and others 2007; Johnson and others 2006; Johnson and others 2008; Liu and others 2006; Liu and others 2005; Thorgeirsson and others 2008; Treutlein and others 2009; Uhl and others 2008b; Uhl and others 2008c; Uhl and others 2001) suggest that much of the genetic influence on vulnerability to substance dependence is likely to be polygenic. These prior studies have identified a number of genetic influences that appear to be shared among individuals from different racial/ethnic backgrounds. However, no prior work of which we are aware provides dense genome-wide data with which we can seek variants that may be more likely to replicate in independent samples of individuals from the same racial/ethnic group than they might be to replicate in samples of individuals from different racial/ethnic groups.

We have developed a "nontemplate" strategy that identifies overall replication of *sets* of genome wide association (GWA) results in the face of difficulties with genetic architectures, samples and genotyping methods (Drgon and others 2010; Johnson and others 2009; Liu and others 2006; Uhl and others 2008b). Such an approach can complement meta-analyses that seek to combine data from single markers whose significance in single samples does not achieve genome wide significance.

We now report application of this nontemplate strategy to identify overall replication of groups of results from GWA studies of four samples of individuals with dependence on illegal substances and matched controls (Drgon and others 2010), (http://www.ncbi.nlm.nih.gov/gap). Two of these samples display European-American and two display African-American genetic backgrounds. These data come from individual genotyping and multiple-pool genotyping approaches that use 1M SNP Illumina and Affymetrix platforms, respectively. The results focus attention on chromosomal regions that are identified by clusters of SNPs for which case *vs* control differences achieve nominal statistical significance in multiple samples from the same racial/ethnic group. We describe the high confidence with which this approach rejects the null hypothesis that nominally-significant SNPs from each sample from the same racial/ethnic group identify the same chromosomal regions with frequencies expected by chance. We note the more modest levels of confidence that this approach provides for identification of individual SNPs, individual chromosomal regions, individual genes and the overlap between data from samples of the

two racial/ethnic groups studied. We discuss this work in light of its technical and analytic limitations and in its similarities and differences with "template" GWA analyses and meta-analyses that seek reproducible associations of striking levels of significance at single SNP markers. The current "nontemplate" replication of sets of results may be useful in other settings in which the underlying properties of the disorder and of the samples create difficulties for searches for individual SNPs with replicated genome wide significance.

Materials and Methods

Subjects, genotyping and assignment of nominal significance of dependent vs control allele frequencies in each sample

1) NIDA/MNB—European-American and African-American research volunteers, largely non treatment seeking, came to the NIDA research facility in Baltimore, Maryland between 1990 and 2007 in response to advertisements and referrals from other research volunteers provided informed consents, self-reported ethnicity data, drug use histories *via* the Drug Use Survey and DSMIII-R or IV diagnoses (Diagnostic and Statistical Manual) and were reimbursed for their time as previously described (Drgon and others 2010; Persico and others 1996; Smith and others 1992; Uhl and others 2001). Genotypes were assessed in DNA pools using Affymetrix 6.0 arrays and methods that we have extensively validated, as previously described (Drgon and others 2010; Johnson and others 2006; Liu and others 2006; Liu and others 2005; Uhl and others 2001). Pooling 1) provided us with the maximal ability to protect the genetic confidentiality of subjects who volunteered for study of genetics of illegal behaviors, 2) allowed us to utilize DNAs from individuals who consented to participation in this study during time periods when consents did not explicitly describe studies using high densities of DNA markers, 3) allowed us to use methods that we have developed and validated in this and in previous work and 4) reduced costs. Many of these subjects would thus not have been available for studies that assessed substantial numbers of polymorphisms using individual genotyping, though many of the most recently-consented subjects (who constituted virtually all of the subjects in four DNA pools) consented to allow their DNAs to be used for unlimited genotyping. Nominal p values for each SNP were determined based on t tests that compared data from multiple abuser *vs* control pools that contained DNAs from 680 European-American and 940 African-American individuals who had mean ages of 32.8 and 34.0 and were 69.5 and 58.8% male, respectively, as described (Drgon and others 2010). In addition, to validate pooling results, we performed individual genotyping for the 155 African American research volunteers who constituted virtually all of the members of 8 DNA pools and who had consented to unlimited individual genotyping, using Affymetrix 6.0 arrays whose results all passed Affymetrix quality control standards and resulted in at least 98% call rates and employing Pearson correlation coefficients.

2) dbGAP samples from the family study of cocaine dependence, COGA and COGEND studies—Unrelated subjects who met DSM criteria for cocaine dependence and control subjects with no evidence for dependence on any addictive substance and who reported smoking fewer than 100 cigarettes in their lives were assembled from three studies and deposited in dbGAP. Family study of cocaine dependence subjects were recruited from treatment centers close to St. Louis. Mo; 55% of contacted subjects participated (Bierut and others 2008). Community-based comparison subjects were recruited through driver's license records from the Missouri Family Registry and were matched to cocaine dependent subjects based on date of birth, ethnicity, gender, and zip code. Eighty percent of screened and eligible comparison subjects participated. Other participants came from individuals who participated in the Collaborative Study on the Genetics of Alcoholism (Nurnberger and others 2004) and the Collaborative Study on the Genetics of Nicotine Dependence (Bierut and others 2007). Dependent individuals displayed DSM (Diagnostic and Statistical Manual

IV) dependence on cocaine as reflected in the dbGAP variable phv00066444.v1.p1. Controls displayed no DSM dependence on cocaine, nicotine, alcohol, marijuana, opioids or other drugs. We eliminated individuals who smoked more than 100 cigarettes in their lives from the control group. We identified 481 dependent and 1053 control unrelated European-American subjects and 516 dependent and 409 control unrelated African-American subjects for this analysis who averaged 40.2 and 37.2 years old and were 47.1 and 46.95 male, respectively. The numbers of cocaine-dependent African-American subjects who were also dependent on nicotine, cannabis, opioids, alcohol or other substances were: 383, 210, 86, 426 and 115, respectively; corresponding numbers for European-American subjects were: 366, 288, 141, 453 and 246.

Genotyping for these samples was performed using Illumina 1M SNP arrays at the Center for Inherited Disease Research (CIDR), with quality controls and principal components analysis (PCA) controls for racial/ethnic background available at the CIDR website (www.cidr.jhmi.edu). Genotypes from dependent and control individuals were selected from dbGAP files.

Primary p values for each SNP were based on χ^2 tests. In addition, to compare analyses with those for NIDA-MNB subjects, we also calculated mean allele frequencies for each SNP for each "pseudopool" that contained data for 20 individuals. We than performed t tests using mean and variance information from the multiple pseudopools that represented the entire sample in ways that paralleled the t tests for NIDA-MNB data noted above.

3) Identification of chromosomal regions containing clusters of SNPs with nominally-significant case vs control differences in single or multiple samples —We performed analyses based on previously-defined criteria using datasets of approximately 1million SNPs (Drgon and others 2010). We identified chromosomal regions of interest in individual samples by seeking regions in which at least 4 clustered SNPs displayed case *vs* control differences with nominal, $p < 0.05$ levels of statistical significance. We defined clustering based on separation of each clustered SNP from the nearest nominally-significant SNP by ≤ 10kb. We identified similarities between the results obtained from multiple samples by identifying the chromosomal regions that were tagged by such clustered, nominally positive SNPs in each of the samples of individuals from the same racial/ethnic group. We identified genes for which these chromosomal intervals lay within the exons of the gene and/or in 10kb of 5' or 3' flanking sequence.

4) Monte Carlo methods for assignment of levels of significance of: a) the extent of clustering in each sample and b) the degree to which clustered nominally-positive SNPs from multiple independent samples identify the same chromosomal regions—For each Monte Carlo trial, we randomly selected a number of "pseudo positive" SNPs from each dataset that matched the number that achieved nominal significance in the *bona fide* dataset. Thus, we constructed a list of autosomal SNPs assayed in each sample and assigned a number to each SNP that corresponded to its position on the list. To select the pseudo-positive SNPs for each trial of the European-American datasets, we selected 75,413 random numbers for the NIDA *(see below)* and 45,108 random numbers for the dbGAP datasets. For each trial, the SNPs identified by the positions on the list that corresponded to these randomly assigned numbers were then queried for the extent to which their results equaled or exceeded the results obtained for the actual dataset. In 10,000 such trials for each sample, we compared results concerning the extent of chromosomal clustering from these sets of pseudo-positive SNPs to those for the true positive SNPs. These empirical Monte Carlo p values thus addressed the null hypothesis that the true positive SNPs from single samples were randomly arrayed on the chromosomes. This Monte Carlo method thus samples from the actual datasets, providing a good method to

address this null hypothesis in light of differences in the linkage disequilibrium structure that might occur from sample to sample.

In 10,000 trials from pairs of samples, we compared the results from pseudo-positive SNPs to those for true positive SNPs. We compared the extent of overlap between chromosomal clusters identified in each independent sample. This analysis addressed the null hypothesis that all of the clustering in each sample was based only on linkage disequilibrium between sets of SNPs that was not related to phenotype. Put another way, the Monte Carlo p values that derive from these 10,000 trials addressed the null hypothesis that the clusters of SNPs found in the true data identified the same chromosomal regions in independent samples as often as anticipated by chance.

Secondary analysis of dbGAP data used permutation approaches as implemented in PLINK (v1.06) (http://pngu.mgh.harvard.edu/purcell/plink/) (Purcell and others 2007). We randomized assignment of the phenotypes to data derived from the current SNPs and analyzed the data from 1,000 permutation trials.

To assess the power of our current approach we used current sample sizes and standard deviations, power calculator PS v2.1.31 (Dupont and Plummer 1990; Dupont and Plummer 1998) and $\alpha = 0.05$.

5) Individual SNP genotyping to follow up SNPs with apparent genome wide significance in dbGAP samples—Several SNPs appeared to achieve “genome wide” levels of significance in dbGAP datasets, including SNPs rs3817222 and rs12734338 in the PPP1R12B gene. We sought individual genotypes for these SNPs and the additional SNP in this gene rs12741415 in NIDA samples using Sequenom MALDITOF assays and the oligonucleotide primers TGGATGAGCAGTCCTCTAAGA, CCACTTACATCCTTTGTCCAG and TCCATCCGAGAGAGGAGG for rs381722; CTGTGGTTACTGGAGTCTGG, TTAGTGCTATAGAACACTAGAAC and TCTGTTAACCAACCTCTGACT for rs12734338 and GGCTCGATTGCCTAATATGGT, AGCTCACCTACCATGTCTTTAA and GTATTTCCCAGACAAGATTGC for rs12741415.

Results

As noted elsewhere (Drgon and others 2010), variation among the allele frequency estimates between pools from individuals of the same phenotype for each racial/ethnic group from the NIDA/MNB samples was +/− 0.02 (standard error of the mean SEM). This represented 0.028 and 0.032 of the mean hybridization density measure for data from African- and European-American samples, respectively. “Pseudopools” constructed from groups of 20 of the dbGAP samples displayed standard error measurements that represented larger, 0.060 and 0.059, corresponding proportions of mean values.

European-American samples

For the NIDA/MNB European-American samples, 75,413 of the autosomal Affymetrix 6.0 SNPs displayed t values with $p < 0.05$ in comparisons between data from substance dependent *vs* control samples (Drgon and others 2010). For the dbGAP data from European-Americans, χ^2 tests displayed $p < 0.05$ for 45,108 autosomal Illumina SNPs. The more modest pool-to-pool variance in the NIDA datasets, as noted above, provided more nominally-positive results from these European-American samples than the dbGAP datasets. Use of t testing also appeared to contribute to the greater number of nominally-positive results in the NIDA datasets. When the pseudopool data from the dbGAP individuals was analyzed using t testing, 49,141 SNPs displayed nominal statistical significance.

Searches for genome wide significance in each European-American sample

We identified case *vs* control p values for t test results from NIDA/MNB samples and for χ^2 results from dbGAP samples from unrelated individuals. Permutation testing for the dbGAP European-American samples revealed $p < 0.001$ for the number of SNPs with nominal case *vs* control p values < 0.05. However, only a few of the p values reached the 10^{-8} level deemed necessary for genome wide significance (*see below*).

Searches for clustering of SNPs with nominally-significant case vs control differences in each European-American sample

We identified 2931 clusters of SNPs that displayed nominally significant, $p < 0.05$ case *vs* control differences for p values from t test results from NIDA/MNB samples and 2783 clusters for χ^2 results from dbGAP samples.

Searches for chromosomal regions identified by clustered SNPs with nominally-significant case vs control differences in both European-American samples

One hundred sixty two chromosomal regions contained clusters of nominally-positive SNPs from each of the two European-American samples.

None of 10,000 Monte Carlo simulation trials that each began with random sets of SNPs selected from each of the datasets identified as many overlapping regions as found in the true dataset. The overall Monte Carlo $p < 0.0001$ for the overlap noted in the true data thus provides very high levels of confidence that these independently-derived sets of results do not identify the same set of chromosomal regions by chance alone.

The chromosomal regions identified by data from both European-American samples and the genes that they contain are listed in Table 1A. The fraction of the genome occupied by these results is about 1.7 × that expected by chance, based on the fraction of the genome occupied by clustered nominally positive results from each of the two European-American samples *(data not shown)*.

African American samples

For the NIDA/MNB African-American samples, 83,330 SNPs displayed “nominally positive” t values with $p < 0.05$. 46,433 SNPs displayed nominally-significant case *vs* control χ^2 differences for dbGAP samples from individuals from this racial/ethnic group. The more modest pool-to-pool variance in the NIDA datasets, as noted above, provided more nominally-positive results from these African-American samples than the dbGAP datasets. Use of t testing also appeared to contribute to the greater number of nominally-positive results in the NIDA datasets. When the pseudopool data from the dbGAP individuals was analyzed using t testing, 48,811 SNPs displayed nominal statistical significance. Permutation testing for the dbGAP African-American samples revealed $p = 0.25$ for the number of SNPs with nominal case *vs* control p values < 0.05

Searches for genome wide significance in each African-American sample

We identified case *vs* control p values for t test results from NIDA/MNB pooled samples (Drgon and others 2010) and for χ^2 results from dbGAP samples from unrelated individuals. Few of these p values approached the 10^{-8} level deemed necessary for genome wide significance, though more reached *ca* 10^{-6}.

In NIDA samples for which there was adequate consent, we performed individual genotyping for the three PPP1R12B SNPs that appeared to achieve high levels of near-genome-wide significance in dbGAP samples. In these individually-genotyped NIDA

samples, rs12734338 initially appeared to provide a highly-significant difference in genotype frequencies between dependent and control individuals in 439 NIDA African American samples ($p = 0.15 \times 10^{-11}$). Neither rs3817222 nor rs12741415 provided any individuals with minor alleles, consistent with data in dbSNP, but inconsistent with HapMap data that indicated that rs3817222 minor allele frequencies were 0.186–0.256. BLAST searches for the PPP1R12B DNA sequences that surround these SNPs revealed a chromosome Y pseudogene that displayed high homology to PPP1R12B DNA. There were sequence differences between this pseudogene and the authentic PPP1R12B gene that correspond to the annotated PPP1R12B "SNPs". The suggestion, from this data, that the allelic polymorphism of this gene is based on differences between the PPP1R12B gene and the chromosome Y pseudogene was further reinforced by our analysis of gender association with the rs12734338 "SNP": 98.6 % of males were "heterozygous" and 96.7 % of females were "homozygous" for this "pseudo-SNP" *(data not shown)*.

Searches for clustering of SNPs with nominally-significant case vs control differences in each African-American sample

We identified clusters of SNPs that displayed nominally significant, $p < 0.05$ case *vs* control differences for p values from t test results from NIDA/MNB samples and for χ^2 results from dbGAP samples (3383 and 2053 clusters, respectively).

Searches for chromosomal regions identified by clustered SNPs with nominally-significant case vs control differences in both African-American samples

One hundred thirty six chromosomal regions were identified by clustered nominally-positive results from both of the two African-American samples. None of 10,000 Monte Carlo simulation trials that each began with random sets of SNPs selected from each of the datasets identified as many overlapping regions as found in the true dataset; hence Monte Carlo $p < 0.0001$.

The chromosomal regions identified by data from both African-American samples are listed in Table 1B. The fraction of the genome occupied by these results is about 2 × that expected by chance, based on the fraction of the genome occupied by clustered nominally positive results from each of the African American samples *(data not shown)*.

Searches for chromosomal regions identified by clustered SNPs with nominally-significant case vs control differences in all four samples

The clusters from both of the two African-American samples identified a single chromosome five region that is also identified by clusters from both of the two European-American samples. The 5' aspects of one large gene, CSMD1, were identified by overlapping clusters from both African-American and from both European-American samples. However, slightly different portions of the 5' end of this gene were identified in the data from the replicate European-American samples compared to the data from the replicate African-American samples. This relatively sparse overlap contrasts with the significant overall overlap between the Affymetrix datasets for the African-American *vs* European American NIDA/MNB samples (Drgon and others 2010) and the Illumina datasets for the African-American *vs* European American dbGAP samples. In the latter case, we can identify 150 chromosomal regions in which overlapping results between the two racial/ethnic groups are found in ways not found by chance in 10,000 Monte Carlo simulation trials (*data not shown*).

Validation of pooling vs individual genotyping for SNPs whose results provided the clusters

We compared individual *vs* pooled allele frequency estimations for the 12,184 SNPs that displayed minor allele frequencies > 0.1 and provided clustered, nominally positive results in data from the NIDA pooled samples. These correlations were more modest than those identified in validating studies for pooling that used larger ranges of expected allele frequencies (average 0.19 range for these genotypes *vs* 0.95 range for the SNPs used in validating studies). Nevertheless, the results from these SNPs displayed 0.68 Pearson correlation coefficients between data from pooled and individual genotyping.

Discussion

Genome-wide association data of increasing richness is available for a number of complex disorders. Several of these GWA datasets contain relatively robust results at “oligogenic” loci that can also be identified, in many cases, by linkage-based approaches (Hageman and others 2005; Haines and others 2005; Lambert and Amouyel 2007; McElroy and Oksenberg 2008). Even moderately secure GWA identification of “polygenic” influences on disease, however, is likely to require replicated data from multiple independent samples.

“Template” analyses seek SNPs that provide “genome wide significance” with the same phase of association in data from each of multiple independent samples. However, there have been no unanimous criteria for declaring replication of sets of data in circumstances in which no SNP achieves this level of statistical significance in each of multiple samples.

We have focused on identification of statistical significance for sets of chromosomal regions that are each identified by sets of nominally-significant SNPs from each of several independent samples. This approach identifies chromosomal regions and genes that are very likely, as a group, to display *bona fide* association with individual differences in vulnerability to develop dependence on addictive substances. This overall confidence derives from approaches that address distinct sets of null and/or alternative hypotheses to explain the results obtained. *First*, seeking chromosomal regions in each sample that are identified by at least 4 closely-spaced nominally-positive SNPs addresses the null hypothesis that the results obtained are randomly distributed across chromosomes. This initial process also addresses the alternative hypothesis that the nominally-positive SNPs are identified based on technical problems that result in misassignment of allele frequency differences to case *vs* control sample comparisons (or misassignment of variances in these values). The *second* way in which we seek replication identifies many of the same chromosomal regions based on their content of clustered, nominally positive results from each of several independent samples. This comparison addresses the null hypothesis that the clustering observed in each sample derives from stochastic case *vs* control differences in haplotype frequencies rather than case *vs* control differences that are truly related to differences in phenotypes. Thus, if clusters of nominally-positive SNPs lay near each other simply because a haplotype was overrepresented in one “case” or “control” group by chance, it is unlikely that this would also occur by chance in another independent sample nearly as often as we have observed. This comparison also provides additional support for the ability to reject the null and alternative hypotheses relating to assay noise.

There are a number of important limitations that come from these samples, these analyses, and from the application of this approach to these datasets. The NIDA/MNB samples, largely of individuals who were not seeking treatment, were recruited at a single site and compare dependent individuals with heavy levels of substance use to controls with modest or no substance use. These features might provide differences from the dbGAP samples which were recruited at a number of sites from largely treatment-seeking individuals (or

from pedigrees with treatment-seeking probands). The dbGAP samples compare dependent individuals to controls whose levels of substance use do not produce dependence, but whose levels of use might be substantial. In addition, these samples provide greater "pseudopool-to-pseudopool" variance than the pool-to-pool variance from NIDA samples, possibly reflecting the substantial site-of-collection to site-of-collection variance noted in another recent report concerning analyses of data from an overlapping subset of these individuals (Bierut and others 2010). Based on statistical considerations, the present analyses are likely to provide many false negative results. The power of each of these samples to detect small, polygenic influences is modest to moderate. The requirement for convergent identification of the same chromosomal region by data from both samples provides a likelihood of even more false negative results. Case *vs* control allele frequency differences in the NIDA/MNB samples were genotyped using multiple DNA pools and an Affymetrix 6.0 platform, providing t tests that use information about both mean differences and variances and not allowing imputation of alleles of SNPs that were not genotyped for most samples. The modest number of individuals who have consented to unlimited genotyping provides additional support for the correlations between individual and pooled genotyping, but is too modest to contribute precise data for case *vs* control analyses in the NIDA samples. Case *vs* control differences in the dbGAP samples were assessed using Illumina platform genotyping of individual samples, yielding χ^2 results without explicit assessment of variance. The requirement that at least 4 nominally-significant SNPs lie within 10kb of each other cannot be fulfilled in a number of chromosomal regions or in a number of genes in which the density of SNPs is too low to meet this stringent requirement (see Supplement of (Uhl and others 2008b) for list of the genes that cannot be assessed with these criteria using the Affymetrix platform). There are only about ¼ million autosomal SNPs that are shared between the *ca*. 900K and 1M autosomal SNPs evaluated by the Affymetrix and Illumina platforms, respectively, further exacerbating this problem in many genomic regions. Further, as noted in the follow up studies of SNPs in PPP1R12B, properties of the SNP assays, rather than of the SNPs themselves, may provide some of the more striking nominal levels of significance in these assays, whether they are genotyped in individual or pooled samples.

Despite these limitations, there is highly-significant overall convergence between two comparisons of NIDA/MNB and dbGAP GWA data that compare individuals who are dependent on at least one illegal substance to controls: one comparison in European-American subjects and another comparison in African-American subjects. For each of these comparisons, the degree to which clusters of nominally-positive SNPs identify the same chromosomal regions and genes is never found by chance in up to 10,000 Monte Carlo simulation trials.

This striking evidence for replication, defined in this fashion, also provides striking contrasts to results from attempts to identify replication (and/or generalization) in other ways. For example, results that seek to identify the extent to which the same SNPs display nominally-significant associations with the same phase in each of these replicate samples within each racial/ethnic group identify about as many SNPs with these properties as expected by chance (*data not shown*).

We have previously reported overlapping results from application of similar analyses to data from replicate samples of methamphetamine-dependent and control Asian samples (Uhl and others 2008c). None of the chromosomal regions identified by these results is labeled by each of the four samples on which we focus in the present report. However, the data from the dbGAP samples identify 15 chromosomal regions in which clustered, nominally-positive SNPs from these two Asian stimulant dependence samples are found along with clustered SNPs from at least one of the dbGAP samples (*data not shown*); six of these regions are

shared by the African-American dbGAP subsample, and nine by the European-American subsample.

We have previously reported the apparent success of "nontemplate" analyses that are similar to those used herein when applied to data from four independent case *vs* control samples for bipolar disorder (Johnson and others 2009). None of these bipolar *vs* control samples, individually, provided results with genome wide significance. These samples combined data from individual and pooled genotyping using different genotyping platforms. Despite these difficulties, the results of nontemplate analyses provided much more frequent identification of the same genomic regions and genes by clustered, nominally positive SNPs from multiple independent samples in bipolar disorder than we would anticipate by chance.

Studies that focus on identifying same-phase association with genome wide levels of significance in multiple independent samples appear most likely to succeed when oligogenic genetic architecture confers large association signals in each independent sample, when the same SNP sets are studied in each, when the disease exhibits little allelic or locus heterogeneity and when there are good matches between the fine patterns of linkage disequilibrium of the samples being studied. Apparent replication "failures" using this approach could thus relate to a number of features that include associations of modest magnitude, sample-to-sample differences in fine patterns of linkage disequilibrium, different amounts of information provided by markers that display population-specific differences in allele frequencies, allelic heterogeneity and locus heterogeneity.

Monte Carlo methods allow us to test the probabilities of chance clustering of nominally-positive SNPs and the chance of convergence between clusters identified in one sample with clusters identified in other samples. Our Monte Carlo approaches deploy an empirical method that uses the existing dataset as a source for "randomly selected" SNPs for each Monte Carlo trial. The results of these simulations provide strong overall confidence that these results are not due to chance. By contrast, these approaches provide absolutely unequivocal identification of few individual SNPs or genes. This lack of unequivocal identification of individual SNPs is consistent with polygenic/allelic heterogeneity current working models for the genetic architecture of vulnerability to substance abuse (Uhl and others 2008a; Uhl and others 2008b).

Previous analyses that have compared the MNB/NIDA European-American to African-American results have identified genomic regions that are labeled by clustered, nominally-positive SNPs from both samples, supporting roles for some allelic variants that are likely to have arisen relatively long ago in human history (Drgon and others 2010; Liu and others 2006; Liu and others 2005; Uhl and others 2001). Such identification of SNP markers whose allelic frequencies distinguish controls from addicts of different ethnicities supports "common disease/common allele" genetic architecture (Lander and Schork 1994) for part of the genetics of addiction vulnerability. However, the substantially greater convergence, noted here, for data from the same racial/ethnic groups also points to possibly-substantial roles for variants that have been accumulated more recently in human populations that have been more separate until relatively recently.

Genes identified by this work include those in several classes. When we compare the list of genes identified by European-American samples to functional classes as annotated in Gene ontology (GO), we find the greatest (*ca* 10^{-4}) statistical significance for *under*representation of the genes whose products are involved with nucleobase, nucleoside, nucleotide and nucleic acid and biosynthetic processes. For African American samples, there is significant *over*representation for "neuromuscular process", "cyanate metabolic process", "synaptic transmission", "neurological process", "transmission of nerve impulse", "regulation of cell

migration", "nerve-nerve synaptic transmission", "regulation of embryonic development and "regulation of cell motility" genes with nearly-significant corrected p values for "cell adhesion/neuron adhesion".

The findings presented here promise to add to the ongoing consideration of methods for comparing GWA datasets as they enhance understanding of genetic underpinnings of human addiction. For addictions, as for many complex disorders, such data provides an increasingly rich basis for improved understanding and for personalized prevention and treatment strategies.

Acknowledgments

This research was supported financially by the NIH Intramural Research Program, NIDA, DHHS. We are grateful for dedicated help with clinical characterization of NIDA/MNB subjects from Dan Lipstein, Fely Carillo and other Johns Hopkins Bayview support staff. Support for dbGAP data came from the NIH Genes, Environment and Health Initiative [GEI] (U01 HG004422)/ Gene Environment Association Studies (GENEVA) that received assistance with phenotype harmonization, genotype cleaning, and study coordination from the GENEVA Coordinating Center (U01 HG004446), assistance with data cleaning from the National Center for Biotechnology Information, and assistance with collection of datasets and samples by the Collaborative Study on the Genetics of Alcoholism (COGA; U10 AA008401), the Collaborative Genetic Study of Nicotine Dependence (COGEND; P01 CA089392) and the Family Study of Cocaine Dependence (FSCD; R01 DA013423). Support for genotyping dbGAP samples at the Johns Hopkins University Center for Inherited Disease Research came through (U01HG004438), the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse and the NIH contract "High throughput genotyping for studying the genetic contributions to human disease" (HHSN268200782096C). dbGaP datasets were obtained at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000092.v1.p1 through dbGaP accession number phs000092.v1.p.

References

Bierut LJ, Agrawal A, Bucholz KK, Doheny KF, Laurie C, Pugh E, Fisher S, Fox L, Howells W, Bertelsen S, et al. A genome-wide association study of alcohol dependence. Proc Natl Acad Sci U S A. 2010; 107(11):5082–5087. [PubMed: 20202923]

Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF, Swan GE, Rutter J, Bertelsen S, Fox L, et al. Novel genes identified in a high-density genome wide association study for nicotine dependence. Hum Mol Genet. 2007; 16(1):24–35. [PubMed: 17158188]

Bierut LJ, Strickland JR, Thompson JR, Afful SE, Cottler LB. Drug use and dependence in cocaine dependent subjects, community-based individuals, and their siblings. Drug Alcohol Depend. 2008; 95(1–2):14–22. [PubMed: 18243582]

Drgon T, Zhang PW, Johnson C, Walther D, Hess J, Nino M, Uhl GR. Genome wide association for addiction: replicated results and comparisons of two analytic approaches. PLoS ONE. 2010 (in press).

Dupont WD, Plummer WD Jr. Power and sample size calculations. A review and computer program. Control Clin Trials. 1990; 11(2):116–128. [PubMed: 2161310]

Dupont WD, Plummer WD Jr. Power and sample size calculations for studies involving linear regression. Control Clin Trials. 1998; 19(6):589–601. [PubMed: 9875838]

Hageman GS, Anderson DH, Johnson LV, Hancox LS, Taiber AJ, Hardisty LI, Hageman JL, Stockman HA, Borchardt JD, Gehrs KM, et al. A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. Proc Natl Acad Sci U S A. 2005; 102(20):7227–7232. [PubMed: 15870199]

Haines JL, Hauser MA, Schmidt S, Scott WK, Olson LM, Gallins P, Spencer KL, Kwan SY, Noureddine M, Gilbert JR, et al. Complement factor H variant increases the risk of age-related macular degeneration. Science. 2005; 308(5720):419–421. [PubMed: 15761120]

Johnson C, Drgon T, Liu QR, Walther D, Edenberg H, Rice J, Foroud T, Uhl GR. Pooled association genome scanning for alcohol dependence using 104,268 SNPs: validation and use to identify alcoholism vulnerability loci in unrelated individuals from the collaborative study on the genetics of alcoholism. Am J Med Genet B Neuropsychiatr Genet. 2006; 141(8):844–853. [PubMed: 16894614]

Johnson C, Drgon T, Liu QR, Zhang PW, Walther D, Li CY, Anthony JC, Ding Y, Eaton WW, Uhl GR. Genome wide association for substance dependence: convergent results from epidemiologic and research volunteer samples. BMC Med Genet. 2008; 9:113. [PubMed: 19094236]

Johnson C, Drgon T, McMahon FJ, Uhl GR. Convergent genome wide association results for bipolar disorder and substance dependence. Am J Med Genet B Neuropsychiatr Genet. 2009; 150B(2): 182–190. [PubMed: 19127564]

Karkowski LM, Prescott CA, Kendler KS. Multivariate assessment of factors influencing illicit substance use in twins from female-female pairs. Am J Med Genet. 2000; 96(5):665–670. [PubMed: 11054775]

Kendler KS, Karkowski LM, Neale MC, Prescott CA. Illicit psychoactive substance use, heavy use, abuse, and dependence in a US population-based sample of male twins. Arch Gen Psychiatry. 2000; 57(3):261–269. [PubMed: 10711912]

Lambert JC, Amouyel P. Genetic heterogeneity of Alzheimer's disease: complexity and advances. Psychoneuroendocrinology. 2007; 32 Suppl 1:S62–S70. [PubMed: 17659844]

Lander ES, Schork NJ. Genetic dissection of complex traits. Science. 1994; 265(5181):2037–2048. [PubMed: 8091226]

Liu QR, Drgon T, Johnson C, Walther D, Hess J, Uhl GR. Addiction molecular genetics: 639,401 SNP whole genome association identifies many "cell adhesion" genes. Am J Med Genet B Neuropsychiatr Genet. 2006; 141(8):918–925. [PubMed: 17099884]

Liu QR, Drgon T, Walther D, Johnson C, Poleskaya O, Hess J, Uhl GR. Pooled association genome scanning: validation and use to identify addiction vulnerability loci in two samples. Proc Natl Acad Sci U S A. 2005; 102(33):11864–11869. [PubMed: 16091475]

McElroy JP, Oksenberg JR. Multiple sclerosis genetics. Curr Top Microbiol Immunol. 2008; 318:45–72. [PubMed: 18219814]

Nurnberger JI Jr, Wiegand R, Bucholz K, O'Connor S, Meyer ET, Reich T, Rice J, Schuckit M, King L, Petti T, et al. A family study of alcohol dependence: coaggregation of multiple disorders in relatives of alcohol-dependent probands. Arch Gen Psychiatry. 2004; 61(12):1246–1256. [PubMed: 15583116]

Persico AM, Bird G, Gabbay FH, Uhl GR. D2 dopamine receptor gene TaqI A1 and B1 restriction fragment length polymorphisms: enhanced frequencies in psychostimulant-preferring polysubstance abusers. Biol Psychiatry. 1996; 40(8):776–784. [PubMed: 8894071]

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81(3):559–575. [PubMed: 17701901]

Smith SS, O'Hara BF, Persico AM, Gorelick DA, Newlin DB, Vlahov D, Solomon L, Pickens R, Uhl GR. Genetic vulnerability to drug abuse. The D2 dopamine receptor Taq I B1 restriction fragment length polymorphism appears more frequently in polysubstance abusers. Arch Gen Psychiatry. 1992; 49(9):723–727. [PubMed: 1355337]

Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, Manolescu A, Thorleifsson G, Stefansson H, Ingason A, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature. 2008; 452(7187):638–642. [PubMed: 18385739]

Treutlein J, Cichon S, Ridinger M, Wodarz N, Soyka M, Zill P, Maier W, Moessner R, Gaebel W, Dahmen N, et al. Genome-wide association study of alcohol dependence. Arch Gen Psychiatry. 2009; 66(7):773–784. [PubMed: 19581569]

True WR, Heath AC, Scherrer JF, Xian H, Lin N, Eisen SA, Lyons MJ, Goldberg J, Tsuang MT. Interrelationship of genetic and environmental influences on conduct disorder and alcohol and marijuana dependence symptoms. Am J Med Genet. 1999; 88(4):391–397. [PubMed: 10402507]

Tsuang MT, Lyons MJ, Meyer JM, Doyle T, Eisen SA, Goldberg J, True W, Lin N, Toomey R, Eaves L. Co-occurrence of abuse of different drugs in men: the role of drug-specific and shared vulnerabilities. Arch Gen Psychiatry. 1998; 55(11):967–972. [PubMed: 9819064]

Uhl GR, Drgon T, Johnson C, Fatusin OO, Liu QR, Contoreggi C, Li CY, Buck K, Crabbe J. "Higher order" addiction molecular genetics: convergent data from genome-wide association in humans and mice. Biochem Pharmacol. 2008a; 75(1):98–111. [PubMed: 17764662]

Uhl GR, Drgon T, Johnson C, Li CY, Contoreggi C, Hess J, Naiman D, Liu QR. Molecular genetics of addiction and related heritable phenotypes: genome-wide association approaches identify "connectivity constellation" and drug target genes with pleiotropic effects. Ann N Y Acad Sci. 2008b; 1141:318–381. [PubMed: 18991966]

Uhl GR, Drgon T, Liu QR, Johnson C, Walther D, Komiyama T, Harano M, Sekine Y, Inada T, Ozaki N. Genome-wide association for methamphetamine dependence: convergent results from 2 samples. Arch Gen Psychiatry. 2008c; 65(3):345–355. [PubMed: 18316681]

Uhl, GR.; Elmer, GI.; Labuda, MC.; Pickens, RW. Genetic influences in drug abuse. In: Gloom, FE.; Kupfer, DJ., editors. Psychopharmacology: The Fourth Generation of Progress. New York: Raven Press; 1995. p. 1793-2783.

Uhl GR, Liu QR, Walther D, Hess J, Naiman D. Polysubstance abuse-vulnerability genes: genome scans for association, using 1,004 subjects and 1,494 single-nucleotide polymorphisms. Am J Hum Genet. 2001; 69(6):1290–1300. [PubMed: 11704927]

Am J Med Genet B Neuropsychiatr Genet. Author manuscript; available in PMC 2012 March 1.

Table 1

Chromosomal regions and genes identified by clusters of SNPs that provide nominally-significant differences between individuals dependent on at least one illegal substance (MNB) or cocaine (dbGAP) and controls of European-American (Table 1A) or African-American (Table 1B) heritage.. Columns list: chromosome, chromosomal region identified by clustered nominally-significant associations in MNB samples, number of nominally-positive SNPs in the corresponding region in MNB samples, chromosomal region identified by clustered nominally-significant associations in dbGAP samples, number of nominally-positive SNPs in the corresponding region in dbGAP samples, and gene(s) identified by the overlapping clusters from the two datasets.

Table 1A:

chr	Region from MNB SNPs *(bp start and end)*		#MNB SNP	MNB SNP with p_{min} in region	p_{min} for MNB SNP	region from dbGAP SNPs *(bp start and end)*		#dbGAP SNP	dbGAP SNP with p_{min} in region	p_{min} for dbGAP SNP	gene(s)
1	22,403,721	22,436,892	13	rs926435	0.00425140	22,411,184	22,419,939	4	rs10917176	0.01722000	
1	68,372,474	68,386,908	9	rs2820487	0.00122278	68,354,587	68,404,090	19	rs3736934	0.00050580	GPR177
1	119,001,219	119,007,078	4	rs1742848	0.00673615	119,001,219	119,005,312	4	rs12143914	0.00572800	
1	168,537,794	168,553,713	4	rs6691470	0.00224958	168,545,502	168,609,913	16	rs10919359	0.00007620	
1	170,608,822	170,631,953	7	rs9425586	0.00114059	170,613,171	170,633,685	7	rs9425592	0.00145600	DNM3
1	199,383,345	199,385,625	4	rs7525970	0.00023529	199,376,607	199,388,146	4	rs8158	0.03283000	TMEM9
1	208,048,696	208,072,444	7	rs6540560	0.00695414	208,063,165	208,073,226	7	rs660975	0.00953700	C1orf107
1	215,665,471	215,672,528	5	rs12128013	0.00002786	215,661,970	215,670,747	6	rs1048126	0.01201000	GPATC2
1	218,996,115	219,012,012	5	rs447861	0.01636257	218,987,480	219,005,191	9	rs12407624	0.02106000	MOSC2
1	219,037,105	219,039,825	4	rs12033808	0.00116666	219,022,635	219,052,523	8	rs1389742	0.00440100	MOSC1,MOSC2
1	238,293,444	238,295,582	4	rs10926111	0.00440053	238,292,083	238,297,349	4	rs10495457	0.02269000	
1	244,061,674	244,078,022	4	rs12066580	0.01505649	244,056,488	244,072,421	14	rs7514038	0.00265400	SMYD3
1	246,129,313	246,145,669	5	rs11204548	0.00006478	246,127,164	246,142,892	6	rs11204553	0.00367600	OR2W3,OR2T8
2	624,905	633,303	4	rs13393304	0.01855858	628,144	649,958	8	rs17042288	0.01080000	
2	15,803,805	15,813,722	4	rs10460289	0.00507678	15,811,342	15,819,589	4	rs6705499	0.01577000	
2	18,998,937	19,027,306	5	rs16985758	0.00097855	19,015,666	19,040,397	5	rs16985798	0.01087000	FLJ41481
2	33,939,112	33,956,488	4	rs17014226	0.00100593	33,947,740	33,960,168	4	rs13432090	0.01374000	
2	35,937,831	35,952,465	4	rs7585354	0.00217892	35,930,844	35,974,162	11	rs1607614	0.00392200	
2	37,942,034	37,952,362	5	rs2565637	0.00073196	37,909,907	37,949,126	11	rs10206788	0.00332400	
2	105,611,105	105,631,845	6	rs4851821	0.00435854	105,622,619	105,633,659	5	rs4851823	0.00958500	
2	129,575,592	129,583,075	4	rs17049088	0.01424614	129,574,528	129,580,602	4	rs13428187	0.00148200	
2	166,871,909	166,879,436	5	rs6760472	0.00023737	166,876,339	166,885,605	4	rs6432901	0.01052000	SCN9A

Am J Med Genet B Neuropsychiatr Genet. Author manuscript; available in PMC 2012 March 1.

Table 1A:

chr	Region from MNB SNPs *(bp start and end)*		#MNB SNP	MNB SNP with p_{min} in region	p_{min} for MNB SNP	region from dbGAP SNPs *(bp start and end)*		#dbGAP SNP	dbGAP SNP with p_{min} in region	p_{min} for dbGAP SNP	gene(s)
2	192,527,275	192,541,198	4	rs6756470	0.00665492	192,529,143	192,540,026	6	rs12612396	0.01866000	TMEFF2
2	233,429,928	233,444,593	5	rs6731064	0.00006057	233,441,388	233,445,376	8	rs2592116	0.00107300	NGEF,UNQ830
3	3,069,450	3,079,303	4	rs163577	0.00731718	3,055,509	3,074,174	9	rs17600202	0.00508400	CNTN4
3	7,166,923	7,178,151	5	rs1499142	0.00760806	7,142,701	7,170,141	11	rs17824908	0.00137500	GRM7
3	15,274,441	15,283,244	4	rs9310472	0.00752533	15,270,368	15,283,244	4	rs1318937	0.00000016	CAP \N7
3	16,139,950	16,151,784	6	rs17041907	0.00016140	16,139,124	16,159,346	8	rs17041904	0.00165300	
3	54,887,170	54,902,941	4	rs17054513	0.00000519	54,860,213	54,893,478	7	rs9849795	0.00611900	CACNA2D3
3	58,533,096	58,550,150	4	rs17059410	0.00232310	58,533,809	58,541,032	6	rs13088795	0.00396600	FAM107A
3	76,530,357	76,553,163	8	rs3907672	0.01018072	76,553,163	76,556,830	4	rs6796472	0.00712900	
3	100,527,865	100,546,621	9	rs900060	0.01172881	100,534,837	100,543,049	4	rs9861463	0.00109800	
3	127,698,906	127,707,980	4	rs4679251	0.00634307	127,698,963	127,749,921	15	rs1799388	0.00260500	UROC1
3	144,099,828	144,129,850	12	rs4683702	0.00045855	144,112,552	144,120,180	4	rs2608077	0.00203800	
3	147,738,756	147,759,347	7	rs13074501	0.00733947	147,725,651	147,754,452	6	rs2587014	0.04025000	PLSCR1
4	6,083,725	6,111,793	10	rs11935825	0.00073626	6,093,633	6,110,127	5	rs6850751	0.01339000	JAKMIP1
4	12,271,875	12,280,659	4	rs13151462	0.00941052	12,263,493	12,291,255	8	rs7668124	0.00169800	
4	15,553,008	15,603,859	13	rs6824333	0.00002604	15,567,988	15,583,320	5	rs12504895	0.00084520	FGFBP2,PROM1
4	37,801,804	37,821,864	4	rs6854169	0.00021616	37,815,419	37,822,024	4	rs17580037	0.01405000	TBC1D1
4	75,904,150	75,914,700	5	rs6840306	0.00032172	75,900,241	75,906,167	4	rs4859417	0.02485000	BTC
4	110,831,986	110,848,343	5	rs5030551	0.00042129	110,823,233	110,831,986	4	rs3212153	0.00425400	CASP6,CCDC109B
4	178,492,904	178,512,644	5	rs2048077	0.00004146	178,482,163	178,494,510	4	rs987467	0.02244000	NEIL3
4	185,570,737	185,604,825	11	rs17585389	0.00531928	185,568,339	185,593,905	5	rs793810	0.00999800	IRF2
5	3,481,925	3,490,919	4	rs42742	0.00099008	3,481,015	3,488,141	4	rs251444	0.00439900	
5	4,691,885	4,705,364	5	rs11134064	0.00000493	4,698,455	4,716,347	4	rs2077369	0.01712000	
5	30,841,081	30,853,465	4	rs2330607	0.00947254	30,821,797	30,850,626	7	rs1547531	0.00737700	
5	36,474,931	36,496,015	5	rs13359394	0.00045106	36,495,832	36,508,747	4	rs2455280	0.00375300	
5	54,288,790	54,318,552	9	rs1508891	0.00088078	54,286,951	54,293,569	4	rs10491370	0.00253100	
5	71,293,156	71,303,134	4	rs16873479	0.00086331	71,281,566	71,293,943	4	rs1217745	0.00135500	
5	83,255,824	83,262,289	4	rs7725359	0.00416153	83,236,774	83,258,315	10	rs9791160	0.00188800	

Am J Med Genet B Neuropsychiatr Genet. Author manuscript; available in PMC 2012 March 1.

Table 1A:

chr	Region from MNB SNPs *(bp start and end)*		#MNB SNP	MNB SNP with p_{min} in region	p_{min} for MNB SNP	region from dbGAP SNPs *(bp start and end)*		#dbGAP SNP	dbGAP SNP with p_{min} in region	p_{min} for dbGAP SNP	gene(s)
5	110,754,871	110,761,539	4	rs17133238	0.01132727	110,758,456	110,770,337	4	rs523389	0.03461000	CAMK4
5	122,633,795	122,637,938	5	rs255628	0.00206409	122,634,922	122,668,972	7	rs2897737	0.00300600	
5	159,208,420	159,222,284	7	rs31689	0.00025307	159,195,539	159,215,437	8	rs12656957	0.00354100	
5	178,580,524	178,586,436	4	rs17079221	0.00575412	178,586,200	178,600,590	5	rs469568	0.00530800	ADAMTS2
6	17,170,349	17,175,020	6	rs9383242	0.00149357	17,156,118	17,175,124	5	rs13213842	0.00446300	
6	22,666,801	22,689,492	4	rs9460760	0.00550320	22,671,616	22,678,043	6	rs12193939	0.00339600	HDGFL1
6	39,396,068	39,414,041	4	rs16891904	0.00129830	39,409,896	39,446,413	17	rs3818308	0.00219800	KIF6
6	81,087,550	81,099,967	8	rs10455370	0.00019544	81,098,304	81,117,440	14	rs4502885	0.01434000	BCKDHB
6	167,007,415	167,038,192	6	rs874277	0.00350443	167,026,084	167,052,518	10	rs874277	0.00465700	RPS6KA2
6	167,600,547	167,619,694	6	rs1209349	0.00045238	167,603,059	167,638,386	15	rs3010558	0.00111400	UNC93A
7	7,867,768	7,876,393	4	rs6979457	0.01230576	7,866,490	7,874,760	5	rs7812102	0.02089000	
7	11,479,011	11,498,204	6	rs7793532	0.00747784	11,485,443	11,502,428	8	rs12539888	0.00184400	THSD7A
7	36,067,711	36,083,854	14	rs1986698	0.00005694	36,083,063	36,091,506	6	rs11975227	0.00947300	
7	50,491,702	50,517,353	11	rs11575492	0.00006757	50,468,934	50,501,033	9	rs10243511	0.00533400	DDC,FIGNL1
7	88,998,520	89,002,076	4	rs11972083	0.00028462	88,999,065	89,006,198	4	rs7787163	0.01037000	
8	3,035,516	3,042,040	4	rs17079607	0.00825003	3,023,076	3,040,999	4	rs2730048	0.03149000	CSMD1
8	4,100,361	4,118,520	5	rs9693235	0.00648816	4,092,064	4,112,474	6	rs779107	0.00463900	CSMD1
8	4,166,391	4,177,720	7	rs10094349	0.00710516	4,173,158	4,177,630	5	rs1847570	0.00817800	CSMD1
8	4,397,922	4,404,404	4	rs1526335	0.01126980	4,362,810	4,427,218	24	rs6996668	0.00002430	CSMD1
8	6,789,365	6,804,222	7	rs11137078	0.00017713	6,802,066	6,816,223	5	rs2702910	0.04002000	DEFA8P
8	10,415,815	10,426,892	4	rs6601481	0.01062660	10,426,159	10,433,464	5	rs6601483	0.01067000	UNQ9391
8	15,702,024	15,712,594	4	rs13438987	0.00708685	15,681,763	15,708,708	8	rs2721207	0.00929400	
8	18,433,022	18,440,225	4	rs7812358	0.01422666	18,425,515	18,440,225	4	rs3739396	0.00293000	PSD3
8	18,562,990	18,573,060	4	rs17695641	0.00001669	18,558,380	18,601,178	9	rs17695724	0.00165400	PSD3
8	18,586,760	18,599,333	4	rs974053	0.00317847	18,558,380	18,601,178	9			PSD3
8	58,003,214	58,007,822	5	rs6474089	0.00104125	57,967,108	58,003,814	9	rs17791051	0.00237600	
8	58,258,438	58,267,213	6	rs7843659	0.01294892	58,260,868	58,277,297	6	rs7463453	0.00925100	
8	62,639,801	62,651,888	4	rs16927574	0.00027789	62,623,810	62,642,488	6	rs4291265	0.00046220	ASPH

Table 1A:

chr	Region from MNB SNPs *(bp start and end)*		#MNB SNP	MNB SNP with p_{min} in region	p_{min} for MNB SNP	region from dbGAP SNPs *(bp start and end)*		#dbGAP SNP	dbGAP SNP with p_{min} in region	p_{min} for dbGAP SNP	gene(s)
8	124,991,629	125,041,794	12	rs16893054	0.00135755	124,996,892	125,008,826	4	rs16899060	0.00966600	C8ORFK23
8	127,648,673	127,671,391	5	rs2385684	0.00032344	127,670,209	127,680,387	4	rs2385694	0.00962100	
8	129,218,353	129,290,287	17	rs2608038	0.00176380	129,222,462	129,239,290	6	rs16893188	0.00388300	
8	129,218,353	129,290,287	17			129,285,356	129,299,376	8	rs1967314	0.00212200	
8	138,940,511	138,960,987	8	rs1511849	0.00935925	138,933,078	138,955,125	5	rs7003495	0.00996600	FLJ45872
9	2,896,806	2,920,875	5	rs12345926	0.00136364	2,890,843	2,921,341	6	rs4013197	0.00141900	
9	28,661,665	28,685,002	4	rs10968712	0.01706234	28,683,005	28,709,213	5	rs1438478	0.01349000	LINGO2
9	35,986,553	35,993,037	5	rs12006109	0.00025560	35,981,104	36,008,310	8	rs16932816	0.00029160	OR13C6P,OR13C7P
9	71,067,399	71,087,631	6	rs1538579	0.01729508	71,063,582	71,073,044	4	rs2039785	0.00812000	TJP2
9	77,493,107	77,503,221	5	rs4745437	0.00007905	77,492,323	77,497,877	4	rs10512050	0.00148800	
9	100,853,373	100,876,607	6	rs16918220	0.00087819	100,869,363	100,901,588	7	rs7034462	0.00865900	COL15A1
9	109,450,466	109,466,790	6	rs7870632	0.00375761	109,426,897	109,450,709	5	rs11794132	0.00394600	
9	113,010,883	113,038,389	14	rs700137	0.00003992	113,016,564	113,037,708	5	rs700131	0.02814000	
10	10,290,510	10,304,657	4	rs12240935	0.00011649	10,296,338	10,301,536	4	rs11256534	0.02209000	
10	12,619,610	12,632,973	4	rs10458806	0.00032252	12,631,718	12,635,821	5	rs6602595	0.00384600	CAMK1D
10	61,585,961	61,607,010	6	rs12355908	0.00022946	61,607,010	61,624,482	5	rs11814752	0.00431800	ANK3
10	72,174,962	72,201,406	6	rs16927943	0.00355390	72,153,016	72,184,617	9	rs1816002	0.00148600	ADAMTS14
10	72,174,962	72,201,406	6			72,197,459	72,210,540	7	rs10999530	0.02139000	C10orf27
10	72,997,579	73,016,551	4	rs7093128	0.01759826	72,979,535	73,002,134	5	rs12247922	0.00699100	CDH23
10	79,015,837	79,028,608	4	rs2673402	0.00120627	79,010,407	79,018,741	4	rs11002212	0.00323600	KCNMA1
10	79,793,193	79,801,448	4	rs7910471	0.00758798	79,801,448	79,818,231	5	rs10490996	0.00218800	
10	125,273,355	125,279,327	4	rs705172	0.00064930	125,278,294	125,284,639	4	rs2486032	0.01196000	
10	127,471,364	127,485,187	4	rs11244653	0.00333472	127,474,643	127,494,379	7	rs11244664	0.02392000	UROS
10	131,563,111	131,590,700	6	rs7916096	0.00196947	131,574,091	131,584,075	4	rs1334011	0.00135800	EBF3
11	7,234,696	7,239,825	4	rs10839752	0.00449163	7,216,630	7,280,083	23	rs12799959	0.00003960	SYT9
11	13,185,312	13,209,685	6	rs7117211	0.01818414	13,189,314	13,196,135	5	rs7112005	0.00175100	
11	75,658,173	75,686,316	7	rs11236682	0.01030137	75,638,688	75,663,802	13	rs4945056	0.00021350	
11	99,351,517	99,368,357	6	rs7103510	0.00140884	99,348,596	99,387,874	15	rs6590484	0.00329900	CNTN5

Am J Med Genet B Neuropsychiatr Genet. Author manuscript; available in PMC 2012 March 1.

Table 1A:

chr	Region from MNB SNPs *(bp start and end)*		#MNB SNP	MNB SNP with p_{min} in region	p_{min} for MNB SNP	region from dbGAP SNPs *(bp start and end)*		#dbGAP SNP	dbGAP SNP with p_{min} in region	p_{min} for dbGAP SNP	gene(s)
12	6,196,175	6,218,155	5	rs10849423	0.01440004	6,210,649	6,223,258	5	rs3181301	0.00436100	CD9
12	50,447,681	50,471,244	6	rs11834760	0.00446783	50,455,698	50,480,150	5	rs2099715	0.02340000	SCN8A
12	51,229,951	51,235,142	4	rs681387	0.01722113	51,224,587	51,246,335	5	rs584436	0.02329000	KRT71
12	68,341,365	68,357,160	5	rs11609972	0.00050239	68,311,754	68,381,836	16	rs796538	0.00462900	BEST3
12	74,665,362	74,685,000	6	rs2117047	0.00209276	74,665,362	74,685,878	5	rs2367446	0.01159000	
12	106,948,693	106,971,387	4	rs11113696	0.00003020	106,960,526	106,983,174	7	rs2374951	0.00852500	
12	110,118,664	110,143,457	6	rs1034603	0.00545252	110,129,183	110,133,727	6	rs1034603	0.01413000	CUX2
12	111,761,119	111,804,161	11	rs2384069	0.00273499	111,803,488	111,811,401	4	rs7958347	0.03495000	RPH3A
12	114,518,824	114,574,647	11	rs7308283	0.00241900	114,495,519	114,519,550	7	rs10735088	0.00349200	
12	114,518,824	114,574,647	11		1.00000000	114,532,466	114,546,369	6	rs12813282	0.00510100	
13	19,936,904	19,941,804	5	rs9315602	0.00179960	19,934,921	19,948,575	8	rs9509234	0.00052670	CRYL1
13	35,830,527	35,860,622	6	rs7989684	0.00052489	35,853,261	35,868,927	5	rs1935099	0.00946500	
13	91,720,738	91,727,447	7	rs7996483	0.01455500	91,672,385	91,722,251	15	rs16947178	0.00012110	GPC5
13	91,850,009	91,865,378	7	rs16947570	0.00046885	91,865,189	91,895,126	8	rs2149065	0.00165600	GPC5
14	22,047,062	22,069,382	5	rs11157651	0.00001672	22,048,615	22,091,560	11	rs412790	0.00585100	TRAJ…
14	32,295,737	32,307,125	4	rs17091659	0.00007471	32,287,714	32,298,632	5	rs2383376	0.01102000	AKAP6
14	52,852,578	52,907,458	27	rs10138849	0.00212501	52,883,170	52,889,727	5	rs1255309	0.02097000	
14	59,408,099	59,431,061	5	rs2882302	0.00108852	59,408,099	59,432,923	5	rs1951366	0.03292000	RTN1
14	72,786,527	72,791,053	7	rs17126352	0.00053080	72,786,268	72,798,710	5	rs8017937	0.02057000	PAPLN
15	29,128,656	29,135,666	5	rs10162727	0.00155011	29,117,572	29,130,036	5	rs2288242	0.00519000	TRPM1
15	78,392,966	78,407,553	5	rs17312725	0.00253795	78,389,782	78,394,769	4	rs3858961	0.00133600	
15	84,164,168	84,168,771	4	rs12907270	0.01138496	84,147,059	84,173,904	5	rs7170181	0.01905000	
15	91,532,308	91,561,784	6	rs6416582	0.00633997	91,537,949	91,553,882	8	rs1872052	0.00022430	UNQ9370
15	95,882,705	95,892,369	4	rs12912857	0.01286705	95,889,109	95,897,046	5	rs1500652	0.01793000	
15	99,375,750	99,389,091	4	rs4965764	0.00344841	99,363,718	99,378,615	4	rs2412004	0.01233000	LRRK1
16	13,664,264	13,673,793	4	rs7195434	0.02469525	13,672,242	13,714,719	13	rs12927885	0.00027820	
16	13,697,788	13,705,700	4	rs12927885	0.02151072	13,672,242	13,714,719	13			
16	17,900,348	17,902,894	5	rs9937528	0.00030556	17,891,568	17,942,205	10	rs1019807	0.00607700	

Am J Med Genet B Neuropsychiatr Genet. Author manuscript; available in PMC 2012 March 1.

Table 1A:

chr	Region from MNB SNPs *(bp start and end)*		#MNB SNP	MNB SNP with p_{min} in region	p_{min} for MNB SNP	region from dbGAP SNPs *(bp start and end)*		#dbGAP SNP	dbGAP SNP with p_{min} in region	p_{min} for dbGAP SNP	gene(s)
16	22,903,047	22,932,189	5	rs16975343	0.02802412	22,916,103	22,931,597	9	rs6497620	0.00180000	
16	26,545,758	26,553,480	4	rs7192315	0.00008835	26,540,129	26,553,877	7	rs237131	0.00995300	
16	72,303,484	72,331,638	9	rs8053939	0.00021607	72,282,721	72,326,475	14	rs2639311	0.00090850	
16	76,385,234	76,399,547	5	rs8055855	0.00064341	76,356,809	76,389,690	9	rs2914451	0.00206500	KIAA1576
16	76,665,475	76,674,781	6	rs16947096	0.00347086	76,669,764	76,681,421	5	rs12927327	0.00547300	
16	77,117,528	77,144,008	5	rs16948054	0.00272614	77,121,333	77,130,526	4	rs11645676	0.02285000	
16	79,205,051	79,220,547	4	rs13334600	0.01441691	79,194,516	79,210,822	6	rs8063688	0.01390000	CDYL2
16	79,474,356	79,491,623	4	rs889519	0.00138143	79,485,938	79,501,696	4	rs12446361	0.00359000	
16	81,717,335	81,727,918	4	rs8054536	0.00213238	81,722,037	81,731,139	5	rs7200240	0.01624000	CDH13
16	82,523,937	82,537,492	4	rs12051537	0.01162385	82,535,061	82,543,753	6	rs4782866	0.00462500	OSGIN1
16	85,297,250	85,308,696	5	rs177267	0.00390403	85,257,186	85,302,819	14	rs12919935	0.00038490	
17	6,102,642	6,106,744	4	rs12452660	0.02094881	6,087,870	6,102,642	4	rs3888640	0.00428800	
17	28,827,828	28,847,221	6	rs7221633	0.00013241	28,828,513	28,835,589	4	rs7214958	0.00799600	
17	51,729,141	51,736,750	5	rs17820092	0.00220747	51,710,283	51,729,610	4	rs7211966	0.00563400	ANKFN1
18	8,777,857	8,784,092	4	rs588397	0.00047409	8,778,967	8,810,886	9	rs8084760	0.00620500	KIAA0802
18	10,727,373	10,755,792	7	rs10775415	0.01410891	10,731,011	10,749,961	6	rs8086878	0.02024000	
18	43,270,667	43,289,577	4	rs7234972	0.00191113	43,268,749	43,287,484	7	rs4939797	0.00268300	
18	53,153,236	53,176,313	5	rs626109	0.00012961	53,164,072	53,225,658	20	rs221877	0.00598900	ST8SIA3
18	53,871,218	53,879,450	4	rs1573390	0.00994035	53,871,218	53,883,190	4	rs8097619	0.00254900	NEDD4L
18	55,287,091	55,289,937	5	rs12961264	0.01295180	55,285,808	55,298,331	9	rs644856	0.01538000	CCBE1
18	73,028,757	73,048,874	4	rs1562774	0.00005443	73,021,843	73,050,153	6	rs7243404	0.00078880	
19	56,057,764	56,066,404	4	rs2739460	0.01128400	56,056,192	56,068,842	4	rs8104556	0.01320000	KLK2,KLK3
19	58,276,230	58,293,226	6	rs10406237	0.00461575	58,272,963	58,286,618	4	rs3745178	0.00844300	ZNF160
20	8,381,577	8,388,510	4	rs2022452	0.00049261	8,356,544	8,386,789	5	rs2719795	0.00295100	PLCB1
20	38,902,918	38,913,320	4	rs16989370	0.00258077	38,904,841	38,930,253	8	rs875627	0.00209200	
20	42,459,198	42,465,058	4	rs6031586	0.00758220	42,464,223	42,475,778	4	rs8116574	0.02672000	HNF4A
21	40,057,196	40,094,216	12	rs2837220	0.00609864	40,058,582	40,064,802	6	rs8128850	0.00162400	
22	24,505,564	24,516,610	4	rs5752205	0.00144446	24,498,558	24,524,941	10	rs9624894	0.00597400	MYO18B

Am J Med Genet B Neuropsychiatr Genet. Author manuscript; available in PMC 2012 March 1.

Table 1A:

chr	Region from MNB SNPs *(bp start and end)*		#MNB SNP	MNB SNP with p_{min} in region	p_{min} for MNB SNP	region from dbGAP SNPs *(bp start and end)*		#dbGAP SNP	dbGAP SNP with p_{min} in region	p_{min} for dbGAP SNP	gene(s)
22	25,576,070	25,584,122	6	rs136544	0.02053141	25,557,454	25,581,367	8	rs136535	0.00244300	
22	35,035,050	35,049,131	4	rs5995281	0.00064906	35,034,987	35,063,884	14	rs8136069	0.00066410	MYH9
22	36,387,284	36,391,396	4	rs732857	0.00244107	36,384,208	36,404,380	8	rs12167604	0.01628000	PDXP

Table 1B:

chr	region from MNB SNPs *(bp start and end)*		#MNB SNP	MNB SNP with p_{min} in region	p_{min} for MNB SNP	region from dbGAP SNPs *(bp start and end)*		#dbGAP SNP	dbGAP SNP with p_{min} in region	p_{min} for dbGAP SNP	gene(s)
1	38,673,230	38,682,775	8	rs7530323	0.00182633	38,672,323	38,678,275	4	rs7530233	0.02940000	
1	55,521,558	55,522,237	4	rs4927218	0.00711474	55,515,425	55,530,395	5	rs10888912	0.00051570	
1	56,899,238	56,905,582	5	rs2796528	0.01599004	56,880,731	56,899,297	7	rs2746349	0.00682100	PRKAA2
1	77,588,107	77,602,693	6	rs10493604	0.00123856	77,582,434	77,612,068	5	rs17100475	0.00249000	AK5
1	87,386,807	87,404,362	5	rs6675309	0.00250523	87,385,946	87,399,117	4	rs7553864	0.00711000	
1	91,692,105	91,695,992	4	rs17131455	0.03028641	91,690,036	91,695,842	4	rs7526795	0.00921700	
1	100,960,780	100,995,245	8	rs6660837	0.00214425	100,948,239	100,968,247	5	rs6680254	0.01082000	VCAM1
1	107,639,114	107,669,240	7	rs17509160	0.00006763	107,638,682	107,647,621	4	rs10494068	0.01202000	NTNG1
1	119,812,845	119,825,411	5	rs17023786	0.01083654	119,795,007	119,815,398	7	rs6672903	0.00002160	
1	164,558,365	164,576,468	11	rs1343295	0.00086242	164,557,685	164,570,256	4	rs10918425	0.01093000	
1	175,396,660	175,414,782	5	rs12092285	0.00071788	175,395,853	175,409,818	4	rs2014384	0.00334800	ASTN,FAM5B
1	198,455,928	198,466,594	5	rs6427809	0.01928194	198,429,293	198,467,479	11	rs10919843	0.00173300	FAM58B
1	207,161,387	207,180,891	6	rs1320539	0.00120812	207,161,387	207,186,803	6	rs4129434	0.00355800	
1	225,031,957	225,044,955	4	rs12071493	0.00299142	225,018,938	225,037,652	5	rs12729579	0.00415200	
1	236,819,330	236,842,448	5	rs16837190	0.00236112	236,824,948	236,830,118	4	rs6698914	0.02440000	
2	42,969,929	43,006,396	9	rs1078100	0.00022284	42,974,788	42,989,296	4	rs4953675	0.00376700	
2	70,790,369	70,823,291	7	rs12621522	0.01574612	70,796,646	70,805,231	4	rs7597992	0.01524000	ADD2
2	105,953,667	105,963,513	4	rs2377342	0.00175814	105,951,160	105,955,449	4	rs6729693	0.02341000	
2	111,507,275	111,521,726	5	rs17482961	0.00299924	111,506,262	111,514,272	4	rs13012948	0.00210900	ACOXL
2	142,088,710	142,106,899	4	rs10928113	0.02228862	142,093,010	142,108,825	6	rs10190730	0.00300700	LRP1B

Table 1B:

chr	region from MNB SNPs *(bp start and end)*		#MNB SNP	MNB SNP with p_{min} in region	p_{min} for MNB SNP	region from dbGAP SNPs *(bp start and end)*		#dbGAP SNP	dbGAP SNP with p_{min} in region	p_{min} for dbGAP SNP	gene(s)
2	166,123,345	166,130,256	4	rs16850914	0.00340857	166,129,660	166,148,793	6	rs10803799	0.00017960	TAIP-2,TTC21B
2	205,043,875	205,056,347	5	rs1550910	0.00392530	205,054,912	205,058,894	4	rs4673300	0.00962500	
2	221,797,760	221,813,151	4	rs1368064	0.00352961	221,789,700	221,797,772	6	rs1368065	0.01004000	
3	8,976,672	8,987,903	5	rs369651	0.01369019	8,950,538	8,978,124	6	rs11922749	0.00581300	RAD18
3	14,470,459	14,486,703	5	rs17237132	0.00177376	14,469,859	14,481,638	7	rs3773176	0.01040000	SLC6A6
3	22,569,059	22,615,321	9	rs7429763	0.00481766	22,565,692	22,575,090	7	rs17011585	0.00002090	
3	22,569,059	22,615,321	9			22,589,646	22,601,911	4	rs12633771	0.01866000	
3	23,156,363	23,160,166	4	rs13317243	0.00052540	23,153,852	23,192,342	12	rs4858060	0.00003840	
3	24,598,884	24,631,782	7	rs17014858	0.00050497	24,618,577	24,632,679	5	rs11706529	0.00018420	
3	41,447,475	41,465,298	7	rs12054014	0.00247884	41,451,566	41,467,634	7	rs13059459	0.00078210	ULK4
3	111,451,489	111,462,018	4	rs7641787	0.00190343	111,431,424	111,451,489	4	rs9883805	0.00167000	
3	141,266,707	141,276,065	7	rs2350488	0.00021976	141,261,763	141,270,836	4	rs12496538	0.00396500	CLSTN2
4	28,246,176	28,269,954	5	rs9996729	0.00006797	28,268,952	28,274,987	4	rs6815271	0.00229500	
4	35,681,896	35,691,621	4	rs1510653	0.01976209	35,681,615	35,691,365	6	rs10010285	0.01869000	
4	93,684,320	93,699,740	9	rs17319672	0.00000872	93,698,298	93,730,844	10	rs7663835	0.00026950	GRID2
4	93,719,692	93,727,390	4	rs17019608	0.00344821	93,698,298	93,730,844	10			GRID2
4	96,544,322	96,580,176	7	rs3912477	0.00399730	96,567,001	96,598,044	12	rs265045	0.00686000	UNC5C
5	10,316,582	10,320,090	5	rs699113	0.01235791	10,302,790	10,320,915	9	rs544	0.02792000	CCT5
5	11,553,608	11,561,023	4	rs17218080	0.00061574	11,560,368	11,572,347	4	rs10038337	0.00494500	CTNND2
5	15,285,308	15,298,402	4	rs10513204	0.01908972	15,278,383	15,302,785	6	rs11133806	0.00498100	
5	36,484,652	36,492,055	4	rs2468513	0.00006454	36,471,233	36,495,341	5	rs17358298	0.00228500	
5	41,575,032	41,617,270	10	rs620876	0.00499106	41,608,479	41,624,460	4	rs583442	0.03793000	TCP1L2
5	112,511,503	112,544,608	10	rs6594693	0.00058306	112,510,548	112,529,682	5	rs10068491	0.00426800	MCC
5	117,910,641	117,923,508	4	rs7719858	0.01062481	117,922,957	117,949,143	7	rs10213999	0.00230800	
5	117,956,000	117,975,073	7	rs2043052	0.00763322	117,971,158	117,979,439	4	rs10519559	0.00978400	
5	147,243,171	147,266,206	8	rs6889356	0.00306666	147,265,906	147,266,651	4	rs2250145	0.00579900	MGC23985
5	151,176,039	151,185,293	4	rs10051560	0.01270991	151,176,325	151,189,083	4	rs7733241	0.02632000	GLRA1
5	167,361,210	167,387,579	5	rs17069636	0.00029339	167,355,762	167,377,584	4	rs17069578	0.01187000	ODZ2
5	169,470,056	169,484,203	6	rs4867917	0.01444342	169,462,713	169,474,921	5	rs10063424	0.00185200	FOXI1

Table 1B:

chr	region from MNB SNPs *(bp start and end)*		#MNB SNP	MNB SNP with p_{min} in region	p_{min} for MNB SNP	region from dbGAP SNPs *(bp start and end)*		#dbGAP SNP	dbGAP SNP with p_{min} in region	p_{min} for dbGAP SNP	gene(s)
6	12,225,541	12,239,453	4	rs2327514	0.00235903	12,228,076	12,247,206	5	rs2228211	0.00802000	HIVEP1
6	37,617,436	37,635,955	5	rs914351	0.00424126	37,609,945	37,618,910	4	rs2797777	0.02426000	FLJ45825
6	45,987,042	46,031,767	10	rs9367228	0.00911977	46,016,169	46,032,945	6	rs4714892	0.00042350	CLIC5
6	93,036,860	93,065,931	10	rs1020320	0.00070018	93,022,018	93,040,012	5	rs1822589	0.00492300	
6	129,838,313	129,856,864	6	rs17057464	0.00275262	129,830,351	129,838,730	5	rs6569603	0.00056950	LAMA2
6	148,822,606	148,840,150	5	rs6927662	0.00352616	148,823,419	148,848,906	9	rs11961740	0.00091830	SASH1
6	152,505,595	152,510,136	4	rs2747665	0.01626437	152,498,721	152,511,829	6	rs17082180	0.00073110	SYNE1
6	164,339,429	164,354,618	5	rs11962696	0.00054469	164,346,155	164,372,273	12	rs206003	0.00017690	
7	37,754,332	37,786,876	6	rs2709114	0.00351750	37,749,625	37,757,420	4	rs7780507	0.00972500	GPR141
7	103,559,035	103,584,883	7	rs194846	0.00030790	103,568,301	103,599,077	8	rs3808008	0.00947200	ORC5L
8	1,468,372	1,478,011	4	rs17681530	0.00386125	1,455,744	1,477,325	6	rs10503166	0.00219000	DLGAP2
8	3,438,465	3,449,401	6	rs2469359	0.00114465	3,419,770	3,440,295	6	rs7013570	0.00021750	CSMD1
8	3,537,344	3,551,589	6	rs17067079	0.00000979	3,543,065	3,557,725	10	rs17326670	0.00081010	CSMD1
8	5,152,475	5,160,210	4	rs7824050	0.00064229	5,137,521	5,156,796	4	rs1420838	0.00613500	
8	6,891,919	6,909,100	4	rs4448295	0.01744268	6,887,848	6,897,291	7	rs10867025	0.00608900	DEFA5
8	15,030,038	15,044,483	8	rs17575278	0.00019902	15,044,188	15,054,249	4	rs6530838	0.01897000	SGCZ
8	17,281,156	17,285,461	4	rs7460082	0.02311621	17,281,979	17,290,933	5	rs10088485	0.00482200	MTMR7
8	54,323,669	54,330,478	4	rs12675595	0.00008797	54,319,624	54,323,911	5	rs2303432	0.02985000	OPRK1
8	72,351,592	72,382,875	8	rs6989867	0.00080278	72,336,968	72,357,946	6	rs11991562	0.00544100	EYA1
8	91,022,384	91,031,953	4	rs13312938	0.00179646	91,027,598	91,040,111	5	rs3026268	0.00650500	NBN
8	139,358,561	139,377,143	7	rs1512407	0.00007918	139,357,180	139,371,129	5	rs1512406	0.00687200	
9	1,439,084	1,458,895	5	rs4142436	0.00055175	1,443,731	1,495,268	15	rs10124818	0.00233700	
9	1,785,030	1,788,805	4	rs3847228	0.00403372	1,786,760	1,789,943	4	rs16934200	0.02853000	
9	5,186,695	5,200,671	5	rs10491650	0.00008609	5,164,638	5,208,524	11	rs7850294	0.00437700	
9	7,050,854	7,088,109	9	rs2381545	0.00308869	7,050,825	7,081,794	9	rs17449018	0.00786400	JMJD2C
9	7,435,627	7,453,978	4	rs2997549	0.01025102	7,438,612	7,456,024	5	rs2997554	0.00175800	
9	8,351,136	8,362,565	7	rs10976994	0.00580524	8,360,763	8,376,063	6	rs1392521	0.00694000	PTPRD
9	9,196,006	9,222,484	6	rs12685122	0.00604373	9,218,094	9,228,929	4	rs4742571	0.01041000	
9	32,867,013	32,879,615	4	rs16918758	0.00489732	32,860,406	32,879,200	6	rs10813887	0.01238000	

Table 1B:

chr	region from MNB SNPs *(bp start and end)*		#MNB SNP	MNB SNP with p_{min} in region	p_{min} for MNB SNP	region from dbGAP SNPs *(bp start and end)*		#dbGAP SNP	dbGAP SNP with p_{min} in region	p_{min} for dbGAP SNP	gene(s)
9	86,974,347	86,985,714	5	rs1387929	0.00466538	86,959,005	86,974,347	5	rs1931101	0.00116700	
9	121,591,631	121,603,681	4	rs11792644	0.00838348	121,588,889	121,606,307	4	rs2151641	0.00389900	
10	74,540,513	74,570,789	9	rs6480671	0.00122779	74,519,332	74,553,223	7	rs6480668	0.02900000	
10	79,882,845	79,900,632	9	rs12763437	0.00333686	79,891,113	79,894,711	4	rs10740479	0.00584900	
10	100,128,973	100,149,126	5	rs4345897	0.00444606	100,134,191	100,145,953	5	rs942812	0.00917700	C10orf33
10	123,963,836	123,982,634	4	rs3752956	0.00474689	123,965,713	123,974,432	4	rs6585804	0.01989000	TACC2
10	130,305,059	130,314,680	4	rs10829448	0.00336612	130,300,924	130,319,479	4	rs12359499	0.01333000	
11	4,550,165	4,583,486	8	rs11033006	0.00029178	4,548,315	4,553,041	4	rs10768096	0.01066000	C11orf40
11	11,346,019	11,359,653	6	rs7131111	0.00464535	11,344,701	11,360,293	4	rs4533032	0.03032000	GALNTL4
11	12,241,317	12,255,138	5	rs11022270	0.00932509	12,253,357	12,271,091	6	rs7106205	0.00324200	
11	20,586,739	20,621,980	7	rs2298826	0.00015712	20,620,979	20,646,062	6	rs4923627	0.00240900	SLC6A5
11	94,021,355	94,035,517	4	rs12786215	0.00130691	94,014,342	94,039,094	7	rs4127396	0.00314200	
11	101,965,568	101,975,466	4	rs1711433	0.00007807	101,965,987	101,990,940	9	rs1711410	0.00134600	MMP20
11	112,578,791	112,596,098	4	rs17115160	0.00347427	112,573,849	112,582,751	5	rs965560	0.01576000	NCAM1
11	121,528,137	121,554,885	6	rs588354	0.00713157	121,539,353	121,556,909	4	rs11218544	0.00146200	
11	122,121,115	122,128,473	4	rs1540113	0.00129108	122,111,119	122,128,603	4	rs4935804	0.00053370	STS-1
12	2,345,664	2,357,262	4	rs10774040	0.00073318	2,343,619	2,359,651	4	rs880342	0.00695700	CACNA1C
12	4,970,508	4,982,410	4	rs11063455	0.01164648	4,968,898	4,977,809	5	rs1014665	0.00953200	
12	21,479,435	21,491,311	4	rs11046048	0.01997843	21,470,347	21,512,686	14	rs2110165	0.01154000	FLJ22028
12	51,572,319	51,601,000	5	rs2131161	0.00335765	51,571,353	51,585,102	6	rs7964223	0.00493500	KRT8
12	52,379,381	52,408,340	11	rs2277370	0.00220727	52,380,303	52,390,299	5	rs12309211	0.00025460	CALCOCO1
12	72,409,569	72,423,208	4	rs17112634	0.00560549	72,381,821	72,409,569	5	rs7299914	0.00909200	
12	96,551,822	96,563,336	4	rs1376345	0.00525089	96,541,866	96,555,782	5	rs1901242	0.00821800	
12	126,937,406	126,963,369	12	rs2699066	0.00032841	126,958,313	126,965,920	4	rs4034627	0.00940100	
13	94,684,855	94,734,859	13	rs9590213	0.00015363	94,691,883	94,714,801	15	rs4258481	0.00091180	ABCC4
13	101,624,080	101,636,884	4	rs9554852	0.00779793	101,633,794	101,650,449	5	rs554393	0.00467000	FGF14
13	102,832,714	102,855,295	4	rs9519026	0.00288507	102,824,275	102,834,437	4	rs1929081	0.00935600	
14	33,479,907	33,484,416	5	rs1680692	0.00181106	33,465,851	33,480,643	4	rs996347	0.03129000	EGLN3
14	83,250,117	83,258,316	6	rs12587990	0.00010936	83,245,454	83,261,998	5	rs10873350	0.04028000	

Table 1B:

chr	region from MNB SNPs *(bp start and end)*		#MNB SNP	MNB SNP with p_{min} in region	p_{min} for MNB SNP	region from dbGAP SNPs *(bp start and end)*		#dbGAP SNP	dbGAP SNP with p_{min} in region	p_{min} for dbGAP SNP	gene(s)
14	85,086,089	85,096,268	4	rs985620	0.00019794	85,074,724	85,098,018	8	rs1955418	0.00203300	FLRT2
14	85,481,243	85,490,806	4	rs10130564	0.00214065	85,474,713	85,484,033	5	rs2373038	0.00057550	
14	93,382,087	93,405,207	6	rs10139611	0.00062020	93,401,569	93,408,486	5	rs11622652	0.00555300	
15	24,761,221	24,768,832	4	rs4887529	0.00107167	24,768,052	24,784,705	4	rs28551016	0.01719000	GABRA5
15	55,531,225	55,551,279	6	rs1280420	0.00475293	55,542,924	55,555,274	4	rs2922219	0.02513000	CGNL1
15	59,264,774	59,293,721	6	rs7167741	0.01522023	59,272,657	59,278,332	4	rs7177846	0.01107000	RORA
15	93,113,075	93,120,231	5	rs11073388	0.00419825	93,114,079	93,123,625	5	rs2199734	0.01306000	
16	6,695,415	6,722,737	6	rs17141047	0.00299366	6,689,567	6,710,879	9	rs7195574	0.00045400	A2BP1
16	27,283,288	27,304,096	6	rs13337730	0.00011074	27,267,076	27,284,411	7	rs3024613	0.00581200	IL4R
16	47,988,550	47,995,541	4	rs27807	0.03801566	47,990,325	48,002,247	4	rs2278026	0.01335000	MGC33367
16	76,126,270	76,134,741	5	rs17771897	0.00296411	76,127,850	76,134,916	4	rs277522	0.02168000	
17	3,154,169	3,161,666	4	rs9898721	0.02078650	3,159,529	3,161,064	4	rs9911226	0.00701900	OR3A4
17	44,633,488	44,638,057	5	rs8077444	0.02228124	44,628,966	44,651,340	7	rs658979	0.00361600	ABI3,GNGT2
17	65,826,027	65,840,454	5	rs16975551	0.00019999	65,802,728	65,826,901	7	rs11868369	0.00023950	
17	66,739,190	66,753,642	4	rs16976490	0.01802707	66,747,639	66,750,867	4	rs16976482	0.03162000	
18	7,013,371	7,026,249	7	rs7231835	0.00226836	7,013,371	7,022,221	5	rs475598	0.00450600	LAMA1
18	42,478,970	42,518,576	9	rs16960377	0.00085181	42,462,780	42,478,970	4	rs16939868	0.01107000	
18	52,007,225	52,023,316	4	rs11151630	0.00068356	51,999,397	52,009,049	4	rs12606608	0.00598200	FLJ45743
18	63,645,346	63,664,475	6	rs2318817	0.00452456	63,636,269	63,655,293	8	rs3907154	0.00369900	
18	66,187,616	66,202,453	4	rs17082526	0.00209317	66,183,437	66,201,794	10	rs7244215	0.00008850	
18	71,933,136	71,949,147	4	rs11664972	0.00461757	71,943,758	71,952,541	4	rs12954572	0.00895100	
18	74,015,536	74,017,210	4	rs10468814	0.00433017	73,987,910	74,019,487	7	rs4798896	0.00310900	
20	19,997,088	20,023,996	5	rs9808594	0.00056896	20,001,676	20,012,701	4	rs6046605	0.00938500	C20orf26
20	48,912,167	48,930,175	4	rs6096138	0.00545080	48,907,768	48,921,399	5	rs6020802	0.00131400	BCAS4
20	58,212,371	58,233,187	7	rs11696896	0.00256062	58,223,398	58,242,908	4	rs11907714	0.02357000	
21	26,419,368	26,465,912	12	rs9984764	0.00160837	26,422,204	26,447,377	6	rs2830072	0.00922800	APP
21	30,117,231	30,124,346	7	rs388700	0.01343888	30,118,644	30,167,132	13	rs468879	0.00006290	GRIK1
21	40,446,501	40,459,247	6	rs11911749	0.00032530	40,435,728	40,454,200	4	rs8130732	0.00670100	DSCAM
22	15,689,881	15,706,432	5	rs2075120	0.00016219	15,695,102	15,706,432	4	rs165611	0.00015620	CECR8

Table 1B:

chr	region from MNB SNPs *(bp start and end)*		#MNB SNP	MNB SNP with p_{min} in region	p_{min} for MNB SNP	region from dbGAP SNPs *(bp start and end)*		#dbGAP SNP	dbGAP SNP with p_{min} in region	p_{min} for dbGAP SNP	gene(s)
22	29,169,346	29,215,980	10	rs5753152	0.00021758	29,185,674	29,194,610	4	rs5753158	0.00046370	SEC14L3
22	35,733,248	35,738,081	5	rs130598	0.00785100	35,727,496	35,747,143	7	rs916213	0.01242000	MGC35206,MPST,TST