Statistical Analysis of pleiotropy between obesity and substance dependence

Abstract

To examine the shared genetic risk variants of obesity and substance dependence, a genome-wide data of 1829 individuals was used. The single locus association analysis and mixed effect liner model were applied, the top significant SNP for obesity located in FTO gene, and the top significant SNP for substance dependence located in the OPRM1 gene. Then, the univariate and bivariate methods was conducted to estimate the heritability and genetic correlation. The heritability for obesity and substance dependence is 0.26 and 0.22 respectively. But in this dataset, there is no significant evidence of genetic correlation. Finally, the central nervous system annotation data and pleiotropy information was integrated to prioritizing GWAS results.

Introduction

Obesity has a profound impact on individual physical health and on public health care. The prevalence of obesity in recent years is particularly alarming: More than one-third of U.S. adults (34.9%) are obese with a body mass index of 30 or greater. [1]

Obesity normally occurs when the consumed calories exceed the expended calories, this imbalanced energy situation can result from complex genetic and environment factors.[2] For a long time, drug and alcohol dependence has been linked to obesity.[3] A study suggests that both food and drugs abuse can be influenced by taste, pleasure, habits, social interactions, convenience, availability, and stress. [4]

Aside from environment factors, a fast growing consensus is that obesity and substance addition might have similar neurobiological framework in brain.[5] Neurotransmitters respond to psychoactive substances are also sensitive to the reinforcing properties of food.[6] One study found obese individuals had significantly fewer dopamine D2 receptors than the normal weight counterparts when compared the metabolic activity in the brains of ten obese individuals with ten normal weight individuals.[7] These findings suggest that individuals with fewer D2 receptors have to eat more in order to experience the rewarding properties of food intake. Another research suggests that substance addition has the similar dopamine system mechanisms as obesity, as for the individuals with chronic drug and alcohol use disorders, the D2 receptor availability is also significantly reduced in the brains. [8]

Several studies have been done to explore the shared genetic susceptibilities of obesity and substance addition. The most widely studied of these has been the dopamine receptor D2 (DRD2) gene.[9-11] However, many large analyses have failed to replicate these finding. [12,13]

Given our limited understanding of the genetic determinants and metabolic syndrome of obesity and substance addition, the shared risk gene cannot be identified. This paper first performed genome-wide association analyses in 2379 European Americans to identify genetic variants associated with obesity and substance dependence status. Then, mixed linear model was conducted to estimate the total variance in liability explained by SNPs through the estimation of genetic relationship matrix. Next, a bivariate REML method was applied to estimates the genetic correlation explained by SNPs between the two traits with genome-wide single-nucleotide polymorphism data.

Method

Data Collection

The GCD samples include 2379 European Americans. All subjects were recruited for studies of the genetics of drug dependence or alcohol dependence. The sample consisted of small nuclear families originally collected for linkage studies, and unrelated individuals. Subjects were recruited at five US clinical sites: Yale University School of Medicine (APT Foundation, New Haven, CT, USA), the University of Connecticut Health Center (Farmington, CT, USA), the University of Pennsylvania Perelman School of Medicine (Philadelphia, PA, USA), the Medical University of South Carolina (Charleston, SC, USA) and McLean Hospital (Belmont, MA, USA). [14]All subjects were interviewed using an electronic version of the semi-structured assessment for Drug Dependence and alcoholism (SSADDA)[15] to derive diagnoses for major psychiatric traits according to DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, 4th edition) criteria[16,17].

The final step also included SAGE data (dbGaP study accession phs000092.v1.p1). In this study, case subjects were identified as having a lifetime history of alcohol dependence using DSM-IV criteria. Control subjects were required to report a history of drinking because alcohol use is required to develop alcohol dependence. Control subjects had no significant alcohol-dependence symptoms. Because of the genetic overlap between alcohol and drug dependence, a diagnosis of drug dependence was an exclusionary criterion for control subjects.

The data recruited families with multiple members as well as unrelated individuals of European descent. The subjects came from 9 different sites across the U.S. [18]

Genotyping and quality control

Samples for GCD were genotyped on the Illumina HumanOmni1-Quad v1.0 microarray containing 988,306 autosomal SNPs. The samples in SAGE study were genotyped on the Illumina Human1M containing 1040107 SNPs.

Sample quality control including removing related individuals, misidentified individuals, as well as individuals with genotype failure rate of 0.02. After sample quality control, the number of GCD samples reduced from 2379 to 1828, the number of SAGE samples reduced from 2668 to 1840. (Table 1.).

For SNP quality control, SNPs with allele frequency > 1% were analyzed. A SNP call rate of 98% was required. Hardy–Weinberg equilibrium (HWE) was tested, and SNPs that deviated from HWE ($P < 10^{-6}$) were excluded. After SNP quality control, the SNP number in GCD samples is 805782; the SNP number in SAGE samples is 845871.

	SEX	N	AGE				BMI				SUB_DEP			
			mean	SD	min	max	mean	SD	min	max	mean	SD	min	max
GCD	Male	1097	38.1	10.7	17	71	27.5	5.1	18	59	18.4	7.0	0	35
	Female	731	37.8	10.3	16	76	27.4	6.7	16	59	17.7	7.3	0	35
SAGE	Male	780	52.2	11.2	30	92	27.5	4.6	19	47	9.56	7.7	0	34
	Female	1060	50.1	10.5	29	86	26.4	6.6	16	71	5.36	5.6	0	33

Table1. Samples Characteristics

Measurement of BMI and substance dependence

BMI was calculated in each study by dividing height (in meters) by weight (in kilograms) squared.

Substance dependence includes alcohol dependence, nicotine dependence, cocaine dependence, opioid dependence and marijuana dependence. The substance dependence is described through 7 addition symptoms: tolerance to substance; withdrawal from substance; use substance in larger amounts or over longer period

than intended; persistent desire or unsuccessful efforts to cut down or control substance use; great amount of time spent in activities necessary to obtain; use or recover from the effects of substance; gave up or reduced important social, occupational, or recreational activities because of substance use. The answer yes means 1 and answer no means 0. For each substance dependence, the score was determined by symptoms account from 0 to 7 by adding up the number of symptoms together. The total substance score is the sum of five substance dependence score ranging from 0 to 35.

Single locus association analysis

Principal components analysis was conducted using the multi-dimensional scaling to identify population stratification of the GWAS subjects. A pruned subset of SNPs was generated with linkage disequilibrium (R^2) smaller than 0.2. The pair-wise IBS (Identify-by-state) matrix was calculated. The first four MDS factors were adjusted in the GWAS analyses to control for population substructure. GWAS analyses were performed on all subjects after quality control with additive genetic model, adjusted by the first four MDS factors, sex, and age using PLINK.[19]

$$BMI = \beta_0 + \beta_1 SNP + \beta_3 SEX + \beta_4 AGE + \beta_5 SD + \beta_6 PC1 + \beta_7 PC2 + \beta_8 PC3 + \beta_9 PC4$$

$$SD = \beta_0 + \beta_1 SNP + \beta_3 SEX + \beta_4 AGE + \beta_5 PC1 + \beta_6 PC2 + \beta_7 PC3 + \beta_8 PC4$$

Mixed linear model based association analysis

Different from single-association analysis, the mixed linear model (MLM) treats all SNPs as random effect in the following form[20]:

$$y = X\beta + Wu + \varepsilon$$

y is an vector of phenotypes, with $\text{var}(y) = V = WW'\sigma_u^2 + I\sigma_\varepsilon^2$

β is a vector of fixed effects, including age, sex and four eigenvectors from principal component analysis. X is the SNP genotype indicator variable coded as 0, 1 or 2. u is a vector of SNP effects with $u \sim N(0, I\sigma_u^2)$. I is an identity matrix, ε is a vector of residual effects with $\varepsilon \sim N(0, I\sigma_\varepsilon^2)$. W is a standardized genotype matrix with the ij^{th}

element $w_{ij} = (\chi_{ij} - 2p_i)/\sqrt{2p_i(1-p_i)}$, where χ_{ij} is the number of copies of the reference allele for the i^{th} SNP of the j^{th} individual and p_i is the frequency of reference allele. Define $A = WW'/N$, A can be interpreted as the genetic relationship matrix (GRM) between individuals.[21] σ_g^2 is the variance explained by all the SNPs, i.e. $\sigma_g^2 = N\sigma_u^2$, with N being the number of SNPs, then $y = X\beta + Wu + \varepsilon$ is equivalent to:

$$y = X\beta + g + \varepsilon$$

g is a vector of the total genetic effects of the individuals with $g \sim N(0, A\sigma_g^2)$. g is the accumulated effect of all SNPs and treated as random effect in this model.

Heritability of obesity and substance dependence

The σ_g^2 is the variance explained by all the SNPs, which can be estimated by restricted maximum likelihood (REML)[22]approach, relying on the GRM estimated from all the SNPs. GCTA was used to estimate the genetic relatedness between all the individuals and removed one of each pair of samples with estimated genetic relatedness > 0.025.

SNP coheritabilities and SNP correlations (r_g SNP)

A bivariate extension of these genome-wide methods estimates the genetic correlation (r_g SNP) explained by SNPs between two phenotypes.[23]

$$y_1 = X_1\beta_1 + Z_1g_1 + \varepsilon_1$$

$$y_2 = X_2\beta_2 + Z_2g_2 + \varepsilon_2$$

The Z incidence matrices relate observations with the vectors of random additive genetic effects. The variance-covariance matrix of phenotypic observations across the two traits is:

$$V = \begin{pmatrix} Z_1AZ_1' + I\sigma_{g1}^2 & Z_2AZ_1'\sigma_{g12} \\ Z_1AZ_2'\sigma_{g12} & Z_2AZ_2' + I\sigma_{g2}^2 \end{pmatrix}$$

Where σ_{g12} is the additive genetic covariance captured by SNPs between the two traits. Individuals contributing to the two traits were unrelated, such that the covariance between environmental effects was assumed to be zero. The genetic correlation coefficient r was calculated with

$r = \sigma_{g12} / (\sigma_{g1}\sigma_{g2})$.

The test statistic is estimated by dividing the square of the estimated genetic correlation coefficient by its approximate sampling variance and calculated a p-value from this test statistic assuming that it is distributed as a chi-square with 1 degree of freedom.

Integrative analysis of two GWAS datasets and functional annotations

Numerous studies have shown that functionally annotated SNPs are more important in explaining the association of phenotype and genotypes.[24,25] In this study, the gene preferentially expressed in the central nervous system was used as annotation data. The p-values were available for 805782 SNPs in GCD samples, 845871 SNPs in SAGE samples. The number of overlapping SNPs for these two datasets is 466115. Consequently, 63274 (13.6%) of the genome-wide SNP markers were annotated. The P-value from these two datasets is:

p-value from GCD: $P_{11,}P_{21,}P_{31}\cdots\cdots P_{m1}$

p-value from SAGE: $P_{12,}P_{22,}P_{32}\cdots\cdots P_{m2}$

P_{jk} is the p-value of j-th SNP in the k-th GWAS datasets. The latent variables $Z_j = \left[Z_{j00}, Z_{j10}, Z_{j01}, Z_{j11} \right]$, indicates the association between j-th SNP and the two phenotypes. The model can be extended to the following model:

$$\pi_{00} = \Pr(Z_{j00} = 1) : (P_{j1} \mid Z_{j00} = 1) \sim u[0,1], (P_{j2} \mid Z_{j00} = 1) \sim u[0,1]$$

$$\pi_{10} = \Pr(Z_{j10} = 1) : (P_{j1} \mid Z_{j10} = 1) \sim Beta(\alpha_1, 1), (P_{j2} \mid Z_{j10} = 1) \sim u[0,1]$$

$$\pi_{01} = \Pr(Z_{j01} = 1) : (P_{j1} \mid Z_{j01} = 1) \sim u[0,1], (P_{j2} \mid Z_{j01} = 1) \sim Beta(\alpha_2, 1)$$

$$\pi_{11} = \Pr(Z_{j11} = 1) : (P_{j1} \mid Z_{j11} = 1) \sim Beta(\alpha_1, 1), (P_{j2} \mid Z_{j11} = 1) \sim Beta(\alpha_2, 1)$$

To incorporate annotation data, Suppose $A \in \text{»}^M$, where A indicates whether the j-th SNP is functionally annotation. $A_j = 1$ means the j-th SNP is annotated, $A_j = 0$ means

the j-th SNP is not annotated. Then the relationship between Z_j and A_j is:

$$q_{00} = \Pr(A_j = 1 \mid Z_{j00} = 1)$$

$$q_{10} = \Pr(A_j = 1 \mid Z_{j10} = 1)$$

$$q_{01} = \Pr(A_j = 1 \mid Z_{j01} = 1)$$

$$q_{11} = \Pr(A_j = 1 \mid Z_{j11} = 1)$$

q_{00} can be explained as the proportion of null SNPs in the annotation data; q_{10} can be explained as the proportion of null SNPs in the annotation data; q_{01} can be explained as the proportion of null SNPs in the annotation data; q_{11} can be explained as the proportion of null SNPs in the annotation data. The joint distribution of Pr(P, A) can be written as:

$$\Pr(P,A) = \prod_{j=1}^{M}\left[\sum_{l\in\{00,10,01,11\}} \Pr(Z_{jl}=1)\Pr(P_j, A_j \mid Z_{jl}=1)\right]$$

$$= \prod_{j=1}^{M}\left[\sum_{l\in\{00,10,01,11\}} \pi_l \Pr(Z_{jl}=1)\prod_{d=1}^{D}\Pr(A_{jd} \mid Z_{jl}=1)\right]$$

Using Expectation-Maximization (EM) algorithm[26], the parameters (π_0, π_1, α) can be estimated.

Results

Bonferroni correction is used for multiple comparisons; the threshold of genome-wide significance was 6.2×10^{-8} for SSADDA. The genomic inflation factor λs were 1.02 and 1.007 for obesity and substance dependence, respectively. The quartile-quartile plots are shown in Figures 1. The Manhattan plots are shown in Figures 2.

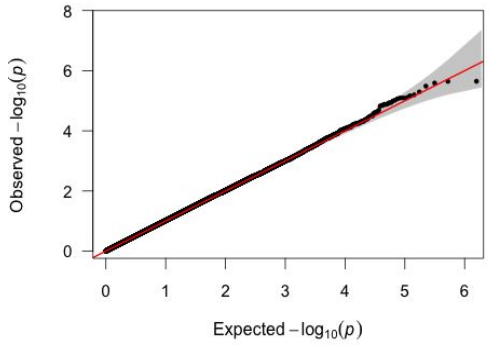


Figure 1. (A)

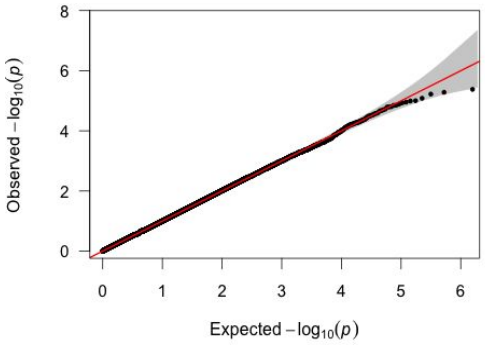
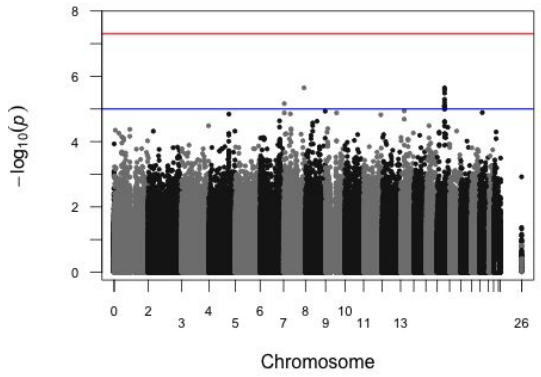


Figure 1. (B)

Figure 1. Quantile-quantile plots of SNPs in single genomic analysis. The gray shaded areas in the Q–Q plots represent the 95% confidence bands around the p-values under the null hypothesis. (A), obesity in GCD sample; (B), substance dependence in GCD sample.



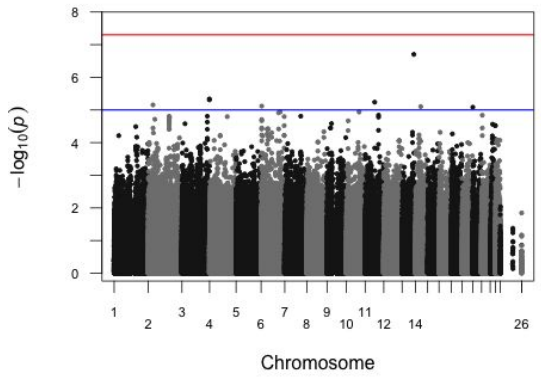


Figure 2. (A)

Figure 2. (B)

Figure 2. Manhattan plots of SNPs in single genomic analysis. SNPs are plotted on the x-axis according to their position on each chromosome against association with obesity on the y-axis(shown as −log10 p-value). The red line indicates the threshold for genome-wide significance and the blue line the threshold for suggestive hits. (A), obesity in GCD sample; (B), substance dependence in GCD sample.

In GCD samples, 94 SNPs exceeded genome wide significant threshold with $p < 1 \times 10^{-5}$ when using obesity as outcome(Table 2.). One SNP, rs1121980, with $p = 2.26 \times 10^{-6}$, on chromosome 16 in the FTO gene, was preciously reported to have the strongest association with early onset obesity in a study of Caucasians.[27] Another four SNPs, rs9940128, rs9939973, rs9937053, rs6050136, located in the same block in the FTO gene with linkage disequilibrium (r^2>=0.88 for all), were the third, fourth, fifth and eighth most significant SNPs. (Figure 3.)

SNP	Chr	BP	A1	Near Gene	Obesity		Sub_dep	
					Beta	P-value	Beta	P-value
rs7794085	7	148347171	A	NA	3.496	2.25E-06	0.3057	0.7535
rs1121980	16	53809247	A	FTO	0.9207	2.26E-06	0.3675	0.1519
rs9940128	16	53800754	A	FTO	0.914	2.57E-06	0.3512	0.1701
rs9939973	16	53800568	A	FTO	0.903	3.27E-06	0.3729	0.1445
rs9937053	16	53799507	A	FTO	0.8836	5.08E-06	0.3754	0.141
rs9930333	16	53799977	C	FTO	0.8752	6.36E-06	0.316	0.2153
rs10270741	7	4911079	A	RADIL	-2.094	6.87E-06	-0.147	0.8105
rs8050136	16	53816275	A	FTO	0.8758	7.94E-06	0.5359	0.0377
rs9922708	16	53831146	A	FTO	0.8713	8.08E-06	0.5385	0.0879
rs9936385	16	53819169	G	FTO	0.8743	8.10E-06	0.6045	0.0189

Table 2. Top Ten Most Significant SNPs in obesity association test results.

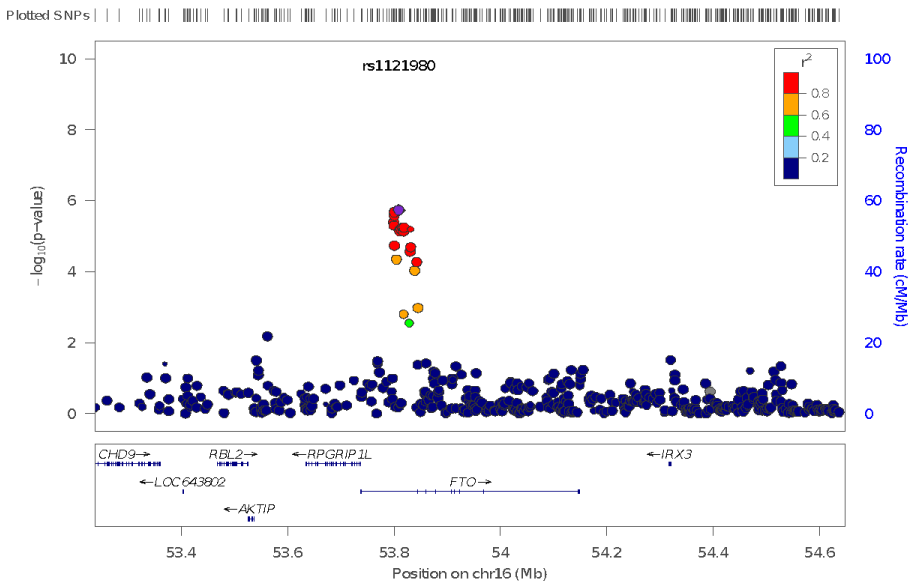


Figure 3. Regional Manhattan plots rs1121980 association regions in European Americans: FTO gene in chromosome 16. The single nucleotide polymorphisms are color coded based on the linkage disequilibrium with the rs1121980.

In GCD samples, 80 SNPs exceeded genome wide significant threshold with $p < 1 \times 10^{-5}$ when using substance dependence as outcome. (Table 3.) The top one significant SNP, rs2010884, with $p = 4.14 \times 10^{-6}$, on chromosome 6 in the OPRM1 gene, was preciously reported to have the association with opioid dependence and alcohol dependence.[28,29] (Figure 4.)

SNP	Chr	BP	A1	NearGene	Sub_dep		Obesity	
					Beta	P-value	Beta	P-value
rs2010884	6	154458207	A	OPRM1	-1.39	4.18E-06	-0.072	0.7538
rs4572747	3	184700356	A	VPS8	3.396	5.23E-06	0.3194	0.5774
rs10916348	1	228785805	A	RHOU	-1.435	5.99E-06	0.2889	0.2351
rs4269065	3	184708802	G	VPS8	3.431	8.26E-06	0.3381	0.5679
rs10852860	17	3416172	A	TRPV3	-1.141	9.94E-06	-0.219	0.2679
rs2291737	3	184717703	A	VPS8	3.401	1.02E-05	0.3975	0.5022
rs1241163	1	103355765	A	COL11A1	1.369	1.11E-05	0.0859	0.7192
rs12114032	8	20847332	A	LOC286114	-1.1	1.17E-05	0.0572	0.7658

rs12005505	9	9893740	C	PTPRD	2.007	1.29E-05	0.1159	0.7423
rs6658934	1	228796151	G	RHOU	-1.382	1.40E-05	0.3237	0.1846

Table 3. Top Ten Most Significant SNPs in substance dependence association test results.

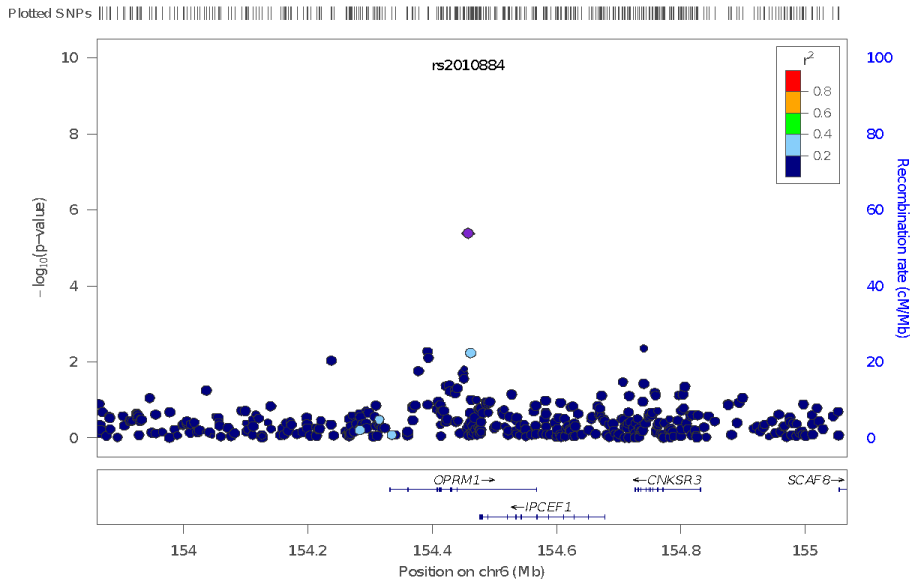


Figure 4. Regional Manhattan plots rs2010884 association regions in European Americans: OPRM1 gene in chromosome 6. The single nucleotide polymorphisms are color coded based on the linkage disequilibrium with the rs2010884.

Next, mixed linear model based association analysis was applied to the GCD data set. The quartile-quartile plots are shown in Figures 5. The Manhattan plots are shown in Figures 6.

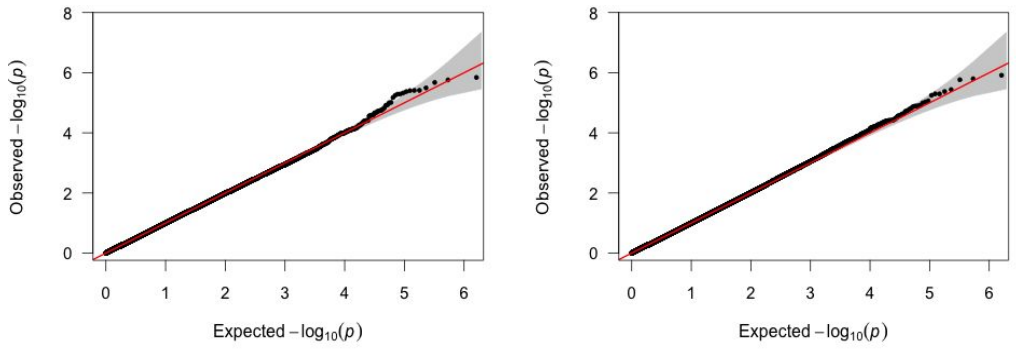
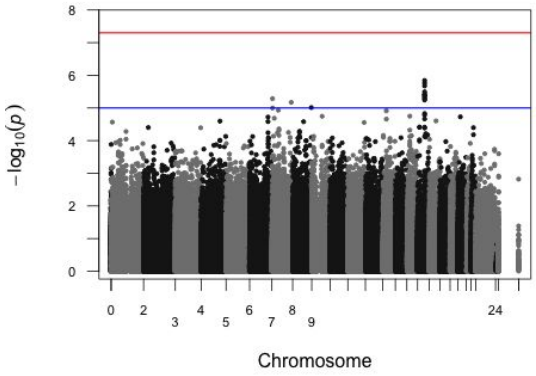


Figure 5. (A) Figure 5. (B)

Figure 5. Quantile-quantile plots of SNPs in mixed linear model analysis. The gray shaded areas in the Q–Q plots represent the 95% confidence bands around the p-values

under the null hypothesis. (A), obesity in GCD sample; (B), substance dependence in GCD sample.



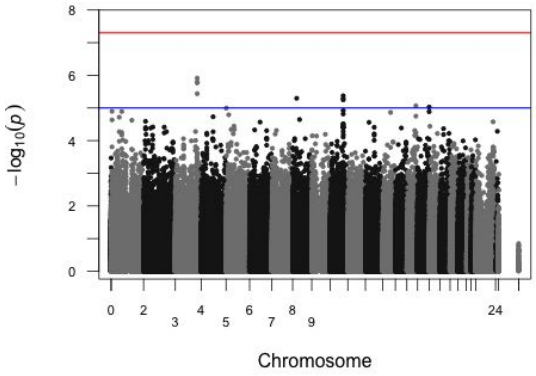


Figure 6. (A) Figure 6. (B)

Figure 6. Manhattan plots of SNPs in mixed linear model analysis. SNPs are plotted on the x-axis according to their position on each chromosome against association with obesity on the y-axis(shown as −log10 p-value). The red line indicates the threshold for genome-wide significance and the blue line the threshold for suggestive hits. (A), obesity in GCD sample; (B), substance dependence in GCD sample.

Using the mixed linear model, in the association test with obesity, 83 SNPs exceeded genome wide significant threshold with $p < 1 \times 10^{-5}$. The top significant SNP is rs1121980, with $p = 1.44 \times 10^{-6}$. In the association test with substance dependence, 70

SNPs exceeded genome wide significant threshold with $p < 1 \times 10^{-5}$. The top significant SNP is rs2010884, with $p = 6.57 \times 10^{-6}$.

The GCTA method provides a lower-bound estimate of narrow heritability. The output is shown in Table 4. The results can be interpreted as the fraction of variance that could be explained by the additive linear effect of all autosomal SNPs. The results suggest that 25.95% (SE=0.16, P=0.0438) of the variance in BMI in the GCD samples can be attributed to genetic differences that are captured by the current SNP microarrays;
For substance dependence, only 21.56% (SE=0.16, P=0.08) of the variance in substance dependence symptom count can be attributed to genetic differences.

	Phenotype	N	Hg	SE	LRT	P-value
GCD	Obesity	1828	0.2595	0.16	2.917	**0.0438**
	Sub_dep	1828	0.2156	0.16	1.890	0.0846

Table 4. Estimated variance explained by all autosomal SNPs for obesity and substance dependence.

Table 5 shows the pearson correlation coefficient between BMI and substance dependence is 0.2408, but because of the large error, the correlation coefficient is not significant different from 0.

	N	rG	(SE)	P-value
BMI/SUB_DEP	1828	0.2408	0.41	0.71

Table5. Estimated genetic correlation from bivariate analyses.

To visually check the genetic relationship between obesity and substance dependence, the histogram of P value from single locus association analysis was plotted. (Figure 7.)

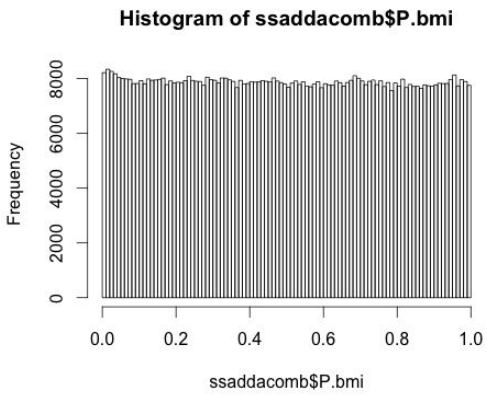


Figure 7. A

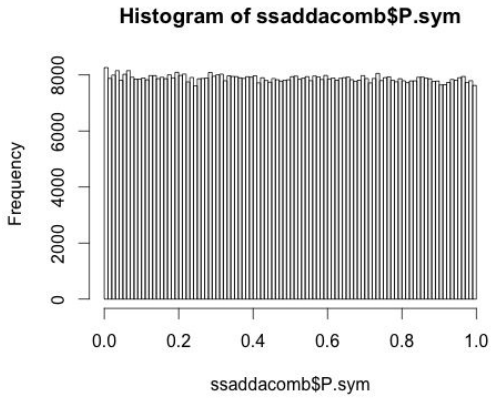


Figure 7.B

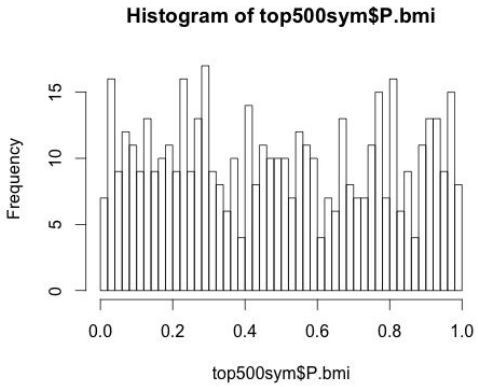


Figure 7. C

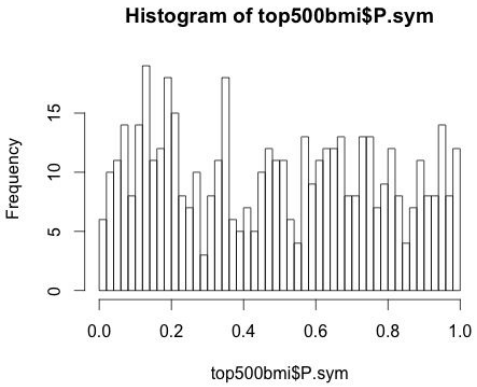


Figure 7.D

Figure 7. Histogram of P value from single locus association analysis. (A), P value of all SNPs in association test with obesity; (B), P value of all SNPs in association test with substance dependence; (C), P value of obesity test of top 500 significant SNPs in substance dependence association test; (D), P value of substance dependence test of top 500 significant SNPs in obesity association test.

Figure 7.C and Figure 7.D shows a flat tendency, which means no pattern of enrichment when take the other phenotype into consideration.

From the results of incorporating annotation data analysis (Table 8.), the ratio between $\hat{q}_{11}$ and $\hat{q}_{00}$ is 2.29 (s.e. 14.6), which is bigger than 1, implies that the annotated SNPs are strongly enriched in SNPs shared by both phenotypes, but because of the large error, we cannot make this conclusion. The estimates parameter $\hat{\pi}_{11} = 0.02(s.e.0.049)$, indicates the shared GWAS signal was very week.

	00	10	01	11
$\hat{\pi}$	0.911(0.086)	0.046(0.053)	0.04(0.084)	0.02(0.049)
$\hat{q}$	0.126(0.013)	0.213(0.094)	0.268(0.086)	0.288(1.843)

Table 8. Results for BMI and substance dependence with the CNS set as annotation.

Discussion

In this study, GWAS analyses of obesity and substance dependence were first applied in 1828 EAs. In the single locus association analysis, there is genome-wide significant evidence of association to the obesity for the SNPs located in FTO genes, the genome-wide significant evidence of association to substance dependence is the top SNP located in the OPRM1 gene. These results are consistent with the results of mixed linear model.

The h_g^2 estimated for obesity is 0.26 (0.16 s.e.), less than the heritability estimated from a twin studies, 0.37. [30] The h_g^2 estimated for substance dependence is 0.22 (0.16 s.e.), which is consistent with the heritability estimated from a twin studies, 0.1 to 0.3. [31] The results show that common SNPs make an important contribution to the overall variance for both phenotypes, implying that additional individual, common SNP associations can be discovered as sample size increases.

The estimates of SNP-based genetic correlation (r_g SNP) between obesity and substance dependence reflect the genome-wide pleiotropy of variants tagged by common SNPs. From the results of GCD samples, obesity and substance lack of overlap, which indicating that the two phenotypes may not share any genetic risk factors. When compared the results of association test of obesity and substance dependence, we found that no SNPs ranked among the top 500 most significant overlapped between the two phenotypes. With the annotation data includes, the results did not show any pleiotropy between obesity and substance dependence in this data set.

Due to the relatively small sample size, we likely missed true shared risk variants.

Also, the sample was recruited primarily to study substance dependence traits; it is possible that the findings are not applicable to obesity in other populations. In addition, the size of the GWAS and samples was small, and thus the statistical power was low.

1 Ogden, C. L., Carroll, M. D., Kit, B. K. & Flegal, K. M. Prevalence of childhood and adult obesity in the United States, 2011-2012. *JAMA : the journal of the American Medical Association* **311**, 806-814, doi:10.1001/jama.2014.732 (2014).

2 Swinburn, B. A. *et al.* The global obesity pandemic: shaped by global drivers and local environments. *Lancet* **378**, 804-814, doi:10.1016/S0140-6736(11)60813-1 (2011).

3 Holden, C. 'Behavioral' addictions: do they exist? *Science* **294**, 980-982, doi:10.1126/science.294.5544.980 (2001).
4 Volkow, N. D. & Wise, R. A. How can drug addiction help us understand obesity? *Nature neuroscience* **8**, 555-560, doi:10.1038/nn1452 (2005).
5 Gearhardt, A. N. *et al.* Neural correlates of food addiction. *Archives of general psychiatry* **68**, 808-816, doi:10.1001/archgenpsychiatry.2011.32 (2011).
6 Wise, R. A. Drug self-administration viewed as ingestive behaviour. *Appetite* **28**, 1-5, doi:10.1006/appe.1996.0059 (1997).
7 Wang, G. J. *et al.* Brain dopamine and obesity. *Lancet* **357**, 354-357 (2001).
8 Volkow, N. D. & Fowler, J. S. Addiction, a disease of compulsion and drive: involvement of the orbitofrontal cortex. *Cerebral cortex* **10**, 318-325 (2000).
9 Munafo, M. R., Matheson, I. J. & Flint, J. Association of the DRD2 gene Taq1A polymorphism and alcoholism: a meta-analysis of case-control studies and evidence of publication bias. *Molecular psychiatry* **12**, 454-461, doi:10.1038/sj.mp.4001938 (2007).
10 Zuo, Y. *et al.* DRD2-related TaqIA polymorphism modulates motivation to smoke. *Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco* **11**, 1321-1329, doi:10.1093/ntr/ntp141 (2009).
11 Doehring, A. *et al.* Genetic variants altering dopamine D2 receptor expression or function modulate the risk of opiate addiction and the dosage requirements of methadone substitution. *Pharmacogenetics and genomics* **19**, 407-414, doi:10.1097/FPC.0b013e328320a3fd (2009).
12 Smith, L., Watson, M., Gates, S., Ball, D. & Foxcroft, D. Meta-analysis of the association of the Taq1A polymorphism with the risk of alcohol dependency: a HuGE gene-disease association review. *American journal of epidemiology* **167**, 125-138, doi:10.1093/aje/kwm281 (2008).
13 Munafo, M. R., Timpson, N. J., David, S. P., Ebrahim, S. & Lawlor, D. A. Association of the DRD2 gene Taq1A polymorphism and smoking behavior: a meta-analysis and new data. *Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco* **11**, 64-76, doi:10.1093/ntr/ntn012 (2009).
14 Gelernter, J. *et al.* Genomewide linkage scan for cocaine dependence and related traits: significant linkages for a cocaine-related trait and cocaine-induced paranoia. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **136B**, 45-52, doi:10.1002/ajmg.b.30189 (2005).
15 Pierucci-Lagha, A. *et al.* Diagnostic reliability of the Semi-structured Assessment for Drug Dependence and Alcoholism (SSADDA). *Drug and alcohol dependence* **80**, 303-312, doi:10.1016/j.drugalcdep.2005.04.005 (2005).
16 Runeson, B. S. & Rich, C. L. Diagnostic and statistical manual of mental disorders, 3rd ed. (DSM-III), adaptive functioning in young Swedish suicides.

Annals of clinical psychiatry : official journal of the American Academy of Clinical Psychiatrists **6**, 181-183 (1994).

17 Pierucci-Lagha, A. *et al.* Reliability of DSM-IV diagnostic criteria using the semi-structured assessment for drug dependence and alcoholism (SSADDA). *Drug and alcohol dependence* **91**, 85-90, doi:10.1016/j.drugalcdep.2007.04.014 (2007).

18 Bierut, L. J. *et al.* A genome-wide association study of alcohol dependence. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 5082-5087, doi:10.1073/pnas.0911109107 (2010).

19 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**, 559-575, doi:10.1086/519795 (2007).

20 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics* **88**, 76-82, doi:10.1016/j.ajhg.2010.11.011 (2011).

21 Stranden, I. & Garrick, D. J. Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *Journal of dairy science* **92**, 2971-2975, doi:10.3168/jds.2008-1929 (2009).

22 Zhang, Q., Zhang, Y., Liu, Z. & Haussmann, H. [Comparison of MIVQUE and REML with Monte Carlo simulation]. *Yi chuan xue bao = Acta genetica Sinica* **22**, 424-430 (1995).

23 Lee, S. H. *et al.* Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nature genetics* **44**, 247-250, doi:10.1038/ng.1108 (2012).

24 Yang, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nature genetics* **43**, 519-525, doi:10.1038/ng.823 (2011).

25 Nicolae, D. L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS genetics* **6**, e1000888, doi:10.1371/journal.pgen.1000888 (2010).

26 Ott, J. Counting methods (EM algorithm) in human pedigree analysis: linkage and segregation analysis. *Annals of human genetics* **40**, 443-454 (1977).

27 Dina, C. *et al.* Variation in FTO contributes to childhood obesity and severe adult obesity. *Nature genetics* **39**, 724-726, doi:10.1038/ng2048 (2007).

28 Patel, S. D. *et al.* Coming to grips with complex disorders: genetic risk prediction in bipolar disorder using panels of genes identified through convergent functional genomics. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **153B**, 850-877, doi:10.1002/ajmg.b.31087 (2010).

29 Kimura, M. & Higuchi, S. Genetics of alcohol dependence. *Psychiatry and clinical neurosciences* **65**, 213-225, doi:10.1111/j.1440-1819.2011.02190.x (2011).

30 Llewellyn, C. H., Trzaskowski, M., Plomin, R. & Wardle, J. Finding the missing heritability in pediatric obesity: the contribution of genome-wide complex trait analysis. *International journal of obesity* **37**, 1506-1509, doi:10.1038/ijo.2013.30 (2013).

31 Vrieze, S. I., McGue, M., Miller, M. B., Hicks, B. M. & Iacono, W. G. Three mutually informative ways to understand the genetic relationships among behavioral disinhibition, alcohol use, drug use, nicotine use/dependence, and their co-occurrence: twin biometry, GCTA, and genome-wide scoring. *Behavior genetics* **43**, 97-107, doi:10.1007/s10519-013-9584-z (2013).