Genome-wide association study identifies common genetic risk factors for alcohol, heroin and methamphetamine dependence

Yan Sun[1#], Suhua Chang[2#], Zhen Liu[1], Libo Zhang[1], Fan Wang[3,4], Weihua Yue[2], Hongqiang Sun[2], Zhaojun Ni[2], Xiangwen Chang[1], Yibin Zhang[1], Yang Chen[1], Lin Lu[1.2], Jie Shi[1,5]*

1 National Institute on Drug Dependence, Peking University, Beijing 100191, China
2 National Clinical Research Center for Mental Disorders, Peking University Sixth Hospital /Institute of Mental Health and Key Laboratory of Mental Health, NHC (Peking University), Beijing 100191, China
3 The Second Affiliated Hospital, Xinjiang Medical University, Urumqi, 830063, China
4 Beijing Hui Long Guan Hospital, Beijing 100096, China
5 Beijing Key Laboratory on Drug Dependence Research, Beijing 100191, China

#Equally contributed to this work

Running title: Common genetic risk factors for addiction

Abstract / Article: 278 / 3856

*Correspondence:

Prof. Jie Shi (shijie@bjmu.edu.cn)

National Institute on Drug Dependence, Peking University, 38 Xueyuan Road, Haidian District, Beijing 100191, China. Tel: 86-10-8280-1593, Fax: 86-10-6203262

Abstract

Background: Common molecular and cellular foundations underlie different types of substance dependence (SD). However direct evidence for common genetic factors of SD is lacking. Here we aimed to identify specific genetic variants that are shared between alcoholism, heroin and methamphetamine dependence.

Methods: We first conducted a combined case-control genome-wide association analysis (GWAS) of 521 alcoholic, 1,026 heroin and 1,749 methamphetamine patients and 2,859 healthy controls. We then replicated the significant loci using an independent cohort (146 alcoholic, 1,045 heroin, 763 methamphetamine and 1,904 controls). Second, we examined the genetic effects of these identified SNPs on gene expression, addiction characteristics and brain images (gray and white matter). Furthermore, we investigated the effects of these genetic variants on addiction behaviors using self-administration rat models.

Results: We identified and validated four genome-wide significant loci in the combined cohorts in the discovery stage: *ADH1B* rs1229984 ($P=6.45\times10^{-10}$), *ANKS1B* rs2133896 ($P=4.09\times10^{-8}$), *AGBL4* rs147247472 ($P=4.30\times10^{-8}$) and *CTNNA2* rs10196867 ($P=4.67\times10^{-8}$). Association results for each dependence group showed that *ADH1B* rs1229984 was only associated with alcoholism, while the other three loci were associated with heroin, methamphetamine addiction and alcoholism respectively. Variants that were strongly linked to rs2133896 affected *ANKS1B* gene expression, heroin use frequency and interacted with heroin dependence to affect gray matter of the left calcarine and white matter of the right superior longitudinal fasciculus. In addition, the reduced *anks1b* expression in the ventral tegmental area increased addiction vulnerability for heroin and methamphetamine in self-administration rat models.

Conclusion: Our findings revealed several novel genome-wide significant SNPs and genes that synchronously affected the vulnerability and phenotypes for alcoholism, heroin and MA dependence. These findings could shed light on the root cause and the generalized vulnerability for SD.

Keyword: genome-wide association study (GWAS); common genetic risk factors; alcoholism;

heroin; methamphetamine; *ANKS1B* rs2133896; imaging; self-administration rat models

1. Introduction

Substance dependence (SD), which includes prescribed or illicit drug use, is a complex disease that is affected by genetics, environmental factors and pharmacological effects. Nicotine, alcohol, cannabis, opioid and amphetamine-type stimulants are the most commonly used addictive substances worldwide[1]. The prevalence of SD to a great extent is affected by the availability of addictive substances and the national culture. The transition to different types of drug addiction is common as several new types of addictive substances emerge. Different types of SD have common effects on the reward circuits, as well as to cellular and molecular adaptations[2]. The occurrence of cross-tolerance and cross-dependence between different types of SD is high[3]. Individuals, who are susceptible to drug exposure and become dependent, have several common characteristics, such as high impulsivity/novelty seeking and behavioral disinhibition[4]. Treatment methods for the diverse types of addiction disorders are limited. Hence, general and cost-effective preventive strategies for SD are required.

Genetic factors contribute significantly to all aspects of SD. The heritability for most addictive disorders has moderate to high pleiotropy and could be considered as the aggregate effect of risk SNPs[5]. Several studies support the shared genetic effects across different types of SD[6]. Generalized genetic vulnerability may attribute to at least 20% etiology for the different SD types[7] and may contribute to the high comorbidity for poly-substance abuse[8,9]. Identified addiction-related genes are mostly related to cell adhesion and memory processes[10]. With regards to genetic studies, Li et al. found that genetic variants in several genes, i.e. aldehyde dehydrogenases, *GABRA2,* and *ANKK1*, were strongly associated with dependence to various substances[11]. By performing gene network analysis, Reyes-Gibby et al. identified *ERK1/2* to be strongly linked to smoking, alcohol, and opioid addiction[12]. However, direct evidence for these common genetic factors in addictions is still lacking.

In this study, we aimed to assess the common genetic risk factors for alcoholism, methamphetamine and heroin dependence, and examined their effects on addiction-related phenotypes. We present a two-stage and case-control genome-wide association analysis (GWAS) using a combined cohort for alcoholism, heroin and methamphetamine dependence. This approach would have greater power in identifying common genetic etiologies for

addiction. Furthermore, to better understand the mechanism of association, we sequentially examined the genetic effects of these cross-addiction variants on gene expression, addiction characteristics, brain images and addictive behaviors using self-administration rat models.

2. Methods

2.1 Subjects

All study subjects were of Chinese Han ethnicity and were older than 16 years of age and able to understand the contents of the questionnaire. This study was approved by the Peking University Institutional Review Board and was performed in accordance with the relevant guidelines and regulations. All individuals signed a written informed consent form and were paid for their participation.

2.1.1 GWAS discovery sample

1) <u>Heroin abusers</u>: We recruited 1,026 heroin abusers (737 males, 289 females) from five drug addiction treatment centers in China's Guangdong and Hubei provinces (Yichang, Sanshui, Shenzhen and Huizhou) from Jan. 2011 to Mar. 2014. Using self-reported questionnaires and government records, heroin was their primary drug of abuse. Confirmatory urine tests were positive for heroin at the time of treatment entry. Individuals met the DSM-IV criteria for heroin dependence as evaluated by trained interviewers. Substance use characteristics (onset age for regular abuse, dose per time (g) and usage frequency per day within the past one-year) were recorded (Table 1). Study patients had no history of poly-substance abuse (continuous use of other opioid drugs for no more than one month, and use of other kinds of addictive drugs for no more than three times within the past three years), with the exception of nicotine dependence. Alcoholism was also excluded by selecting patients with a score < 4 in the Michigan Alcoholism Screening Test. All study subjects had no history of other major psychiatric disorders and medical conditions (e.g., cardiovascular, endocrinological, oncological, or autoimmune diseases) based on self-reports.

2) <u>Methamphetamine abusers</u>: A total of 1,749 methamphetamine (MA) abusers (1,255 males,

494 females) were recruited from three addiction treatment centers in China's Guangdong province (Dongguan, Shenzhen and Sanshui) from Dec. 2016 to Mar. 2017. Using self-reported questionnaires and government records, MA was their primary drug of abuse and urine tests confirmed MA use at time of treatment entry. Patients met the DSM-V criteria for methamphetamine dependence and were evaluated by trained interviewers. Substance use characteristics (onset age for regular abuse, dose per time (g) and usage frequency per day within the past one-year) were recorded (Table 1). Patients did not use other non-amphetamine-type stimulants (e.g., opioid and cocaine) for more than three times per year within the past three years, with the exception of nicotine. Alcoholism was also excluded by selecting patients with a score < 4 in the Michigan Alcoholism Screening Test. All study subjects had no history of other major psychiatric disorders and medical conditions based on self-reports.

3) <u>Alcohol abusers</u>: A total of 521 male alcoholic inpatients were recruited from twelve psychiatric hospitals in northern China (Beijing, Inner Mongolia Autonomous Region, Shandong, Tianjin, Jilin, Liaoning and Heilongjiang provinces) from Sept. 2009 to Nov. 2011. All enrolled patients sought treatment for alcohol dependence (AD) and were clinically ascertained as AD based on DSM-IV by experienced psychiatrists. The severity of alcoholism was measured by criterion counts based on the Chinese version of the Michigan Alcoholism Screening Test (MAST)[13]. Patients had not used any other addictive substances except for nicotine based on self-reports. All study subjects had no history of other major psychiatric disorders and medical conditions (e.g., cardiovascular, endocrinologic, oncologic, or autoimmune diseases) based on self-reports.

4) <u>Healthy controls:</u> 2,859 healthy controls (HCs) were recruited from local communities through advertisements and community centers. All HCs did not have a history of substance abuse (except nicotine), psychiatric disorders or any current major medical conditions (e.g., cardiovascular, endocrinological, oncological, or autoimmune diseases) based on self-report questionnaires.

2.1.2 Genetic replication cohorts

1) <u>Heroin and methamphetamine abusers</u>: The replication sample included 1,045 male heroin abusers and 763 male methamphetamine abusers from Sanshui addiction treatment hospital in Guangdong province of China. They were enrolled in Mar. 2018. The inclusion criteria (except for updating the DSM-V criteria for heroin dependence) and exclusion criteria were consistent with the discovery cohort.

2) <u>Alcoholics</u>: 146 male AD inpatients were recruited from the Peking University Sixth Hospital and four other psychiatric hospitals from Anhui, Shangdong and Henan provinces in China on 2018. The inclusion (except for updating the DSM-V criteria for alcoholism) and execution criteria were similar to the discovery cohort.

3) <u>Healthy controls</u>: The 1,904 healthy controls (HCs) were recruited from the local community. The recruitment method and inclusion criteria were similar to the controls in the discovery cohort.

2.1.3 Imaging analysis subset

Magnetic resonance imaging (MRI) scans and analysis were conducted for a subset of patients in the heroin discovery cohort, which consisted of 65 male heroin patients and 69 male healthy controls. They all had education levels that were higher than primary school. Additional exclusion criteria for imaging included left-handedness and contraindications for MRI (e.g., claustrophobia, dentures, head trauma, and metal implants).

2.2 Genetic analysis

2.2.1 Genotype, Quality control and Imputation

Genomic DNA was extracted from 5 mls of peripheral blood from all study participants and then genotyped (concentration $\geq$ 50ng/ul, total quantity $\geq$ 2ug, absorbance 260/280 ratio between 1.7 and 2.0). Genotyping for the discovery cohort was performed using the Illumina Global Screening Array-24 v1.0 BeadChip (Illumina, Inc., San Diego, CA, USA). For details of the quality control, please see supplemental methods.

Genotype imputation was performed using the pre-phasing/imputation stepwise approach in IMPUTE2[14] and SHAPEIT[15]. The imputation reference set consisted of 2,186 phased

haplotypes from the full 1000 Genomes Project Integrated Phase 1 Release (March 2012)[16]. Imputed SNPs with info < 0.6 or SNPs with minor allele frequency < 0.01 were removed.

2.2.2 Genetic association analysis

Association analysis for the discovery stage was performed using SNPTEST[17] (-frequentist 1, -method score) (https://mathgen.stats.ox.ac.uk/genetics_software/snptest/ snptest.html) for the data after imputation. Gender, age and top 10 principle components (PCs) from principle component analysis (PCA) were used as covariates. For each subgroup substance dependence, (MAs vs controls, heroin addicts vs controls, alcoholics vs controls), the top 10 PCs were obtained from the PCA for the specific subgroup cohorts. All association test p-values were two-sided and were reported without correction for multiple testing. Genome-wide significance was considered as $P < 5 \times 10^{-8}$. Manhattan and Q-Q plots were generated using the R package, qqman. Regional plots were generated using LocusZoom[18].

2.2.3 Replication stage

Significant SNPs from the discovery stage were genotyped in the replication cohort using the Agena MassArray Analyzer 4.0 (Agena Bioscience, Inc., San Diego, CA, USA). Association analysis for the replication stage cohort was performed with PLINK 1.9[19] using an additive model in logistic regression with age as covariate. The sample size weighted meta-analysis in METAL[20] was used for analysis of the two stages.

2.2.4 Bioinformatics and addiction characteristics association analysis

Gene expression profiles of the significant genes were obtained from GTEx[21]. To investigate the effect of rs2133896 on gene *ANKS1B*, the e-QTL was examined using the UK Brain Expression Cohort data set (GSE46706)[22]. Expression plot was generated using BRAINEAC (http://peana-od.inf.um.es:8080/UKBECv12/) by searching for target genes, transcript selection, and then stratifying its expression by target variants. Gene-level association analysis and gene-set association analysis were performed using the MAGMA software (Multi-marker Analysis of GenoMic Annotation, http://ctglab.nl/software/magma[23]). The mapping window from SNP to gene was 5kb. KEGG pathway data (n = 221) was used for gene-set analysis. Significance levels were set as

0.05/18,225=2.74×10^{-6} for gene and 0.05/221=2.26×10^{-4} for gene-set separately.

To investigate the association of the significant locus in chr12 *ANKS1B* with addiction characteristics, we calculated a polygenic risk score (PRS) for all SNPs with P<10^{-6} from the combined GWAS in this locus. They all had high LD ($r^2 > 0.75$) with rs2133896 being the top SNP. PRS was calculated using the --score command in PLINK, which used the BETA value of the combined GWAS as the SNP score. The PRS association with addiction characteristics, (age-onset, dose, frequency for heroin and methamphetamine, MAST for alcoholism) were analyzed using a linear regression model, adjusted for age, gender (only for heroin and methamphetamine) and 10 PCs for the GWAS association analysis.

2.3 MRI acquisition and analysis

MRI data were acquired using a GE Signa Twin speed MRI 1.5T scanner (General Electric Medical System, Milwaukee, WI, USA) and a standard 8-channel head coil at 2011. A voxel-based morphometry (VBM) analysis was used to compare the whole-brain gray matter volume of all subjects using the VBM8 software package (http://dbm.neuro.uni-jena.de/vbm/). Diffusion image preprocessing and analysis were performed using the FMRIB's Diffusion Toolbox (FDT) (FSL 4.1.4; www.fmrib.ox.ac.uk/fsl). For further details, please see supplemental methods.

2.4 Animal models

2.4.1 Daily drug administration

Male Sprague Dawley rats (240-260g) were purchased from the Beijing Vital River Laboratories. Methamphetamine or heroin was dissolved in 0.9% saline and injected intraperitoneally (i.p.). Rats were injected daily with methamphetamine or heroin for 16 consecutive days. Details of the procedure are presented in the supplemental methods (Timeline is presented in Figure 4A).

2.4.2 Adenovirus injection

rAAV2/9-CMV-ANKS1B-flag-pA (5.00 x 10^{12}vg/mL) and rAAV2/9-CMV-flag-WPRE-pA (5.00 x 10^{12}vg/mL) were generated by BrainVTA.Co. All viral vectors were stored in aliquots

at -80°C until use. Approximately 1 μL of virus was injected into the ventral tegmental area (VTA) (−5.2 mm from bregma, +1.8 mm lateral to the midline (10° angles), and −7.5 mm from the skull surface) per side. The syringe was left for ten minutes after the end of the injection to allow for virus diffusion. After virus administration, rats were returned to their cages for postsurgical recovery for 1 week before intravenous drug self-administration. Rats without proper virus targeting were excluded from the study.

2.4.3 Surgery, Training and Testing

Rats were between 300–320g in weight when intravenous drug self-administration was initiated. For self-administration methamphetamine training, there were ten rats in the *anks1b* over expression group and eight rats in the control group. For self-administration heroine training, there were ten rats in the *anks1b* over-expression group and nine rats in the control group. The self-administration training protocol has been described in our previous publication[24], and is additionally presented in Figure 4C and the supplemental methods.

Rats were subjected to a between-session dose-response test with each of 5 injection doses (0, 0.03, 0.1, 0.3, 1 mg/kg per infusion). Each dose was tested for 1 hour and a total of 6 hours for each of the 5 injection doses. The number of injections was recorded during test sessions. After decapitation, rat brains were harvested promptly for western blot assays. The detailed procedures for western blot assays were described in our previous studies[25] as well as in the supplemental methods.

3. Results

3.1 Sample characteristics

The three cohorts for substance dependence (alcohol, methamphetamine and heroin) were used for the GWAS. The demographics and addiction characteristics of the discovery and replication cohorts are presented in Table 1. After quality control, the final GWAS discovery cohort consisted of 1,026 heroin abusers, 1,749 MA abusers, 521 alcoholics and 2,859 healthy controls. The replication cohort included 1,045 heroin abusers, 763 MA abusers, 146

alcoholics and 1,904 healthy controls.

3.2 GWAS analysis and replication

We found four genome-wide significant loci (lead SNPs: *ANKS1B* rs2133896 (P=4.09×10^{-8}), *AGBL4* rs147247472 (P=4.30×10^{-8}), *CTNNA2* rs10196867 (P=4.67×10^{-8}) and *ADH1B* rs1229984 (P=6.45×10^{-10})) in the association analysis for the combined SD vs controls comparison (Table 2). The Manhattan and Q-Q plots for the GWAS results are shown in Figure 1A and 1B respectively. The regional plots for each significant locus are shown in Figure 1C-1F.

We then analyzed the association of these significant SNPs for each of the three addictions vs controls. *ANKS1B* rs2133896, *AGBL4* rs147247472, and *CTNNA2* rs10196867 were synchronously associated with alcoholism, heroin and methamphetamine dependence respectively (P<0.05, Table 2), while *ADH1B* rs1159918 was only significant for alcoholism (P=1.03×10^{-12} for alcohol, P=0.986 for MA and P=0.942 for heroin).

In the replication cohort, the association between *ANKS1B* rs2133896 and SD was validated for both the combined cohort (P$_{meta}$ = 6.23×10^{-9}) and the single drug dependence groups (P$_{meta}$ for heroin= 2.37×10^{-5} and P$_{meta}$ for methamphetamine = 2.77×10^{-7}). The significant results were also replicated for *AGBL4* rs147247472 and *CTNNA2* rs10196867. *ADH1B* rs1159918 was only tested in the alcoholic cohort and was replicated to be associated with alcohol dependence (Table 2).

Since *ADH1B* rs1159918 was mainly associated with alcohol dependence, and *AGBL4* rs147247472 and *CTNNA2* rs10196867 had low MAF (0.0174 and 0.0297 respectively) in our sample cohort, e-QTL analysis, addiction characteristics association analysis, brain imaging analysis and the animal model study focused only on *ANKS1B* rs2133896 (MAF=0.222, protective allele (G), risk allele (T)).

3.3 Gene pathway convergence

Gene-level analysis for the combined GWAS data identified three significant genes (*ADH1A*, *ADH1B* and *ADH6*). The top 10 genes are shown in Supplementary Table 1. Additional three

significant gene loci had low p-values (P=5.59×10^{-6} for *ANKS1B*, P=5.19×10^{-5} for *CTNNA2*, P=5.33×10^{-4} for *AGBL4*) but did not pass multiple corrections.

Gene-set analysis found two significant pathways "Wnt signaling pathway" and "Basal cell carcinoma" that passed multiple corrections. The top 10 KEGG pathways are shown in Supplementary Table 2.

3.4 Effect of *ANKS1B* rs2133896 on gene expression

We assessed the effect of *ANKS1B* rs2133896 on gene expression in the brain by BRAINEAC. The top SNP rs2133896 was not found in BRAINEAC, hence we used its LD-proxy rs10860447 ($r^2 = 0.902$, calculated using --ld in PLINK). C allele (linked with rs2133896 T allele) had high *ANKS1B* expression in white matter and low expression in gray matter of the cerebellum (Figure 2A). Using GTEx, we found that *ANKS1B* was specifically expressed in the brain (Figure 2B).

3.5 Association of *ANKS1B* locus with addiction characteristics

PRS association analysis for chr12 *ANKS1B* locus with age-onset, dose, frequency of heroin and MA, and the MAST of alcoholism showed significant association with heroin frequency (BETA = 5.78, P = 0.035). The higher polygenic risk for combined addiction was associated with higher heroin addiction frequency (Supplementary Table 3).

3.6 Association of *ANKS1B* rs2133896 with imaging characteristics

We first assessed the effect of rs2133896 on gray matter and found a significant interaction effect of *ANKS1B* rs2133896 (TT genotype and G carriers) × group (heroin and healthy) on the gray matter volume of the left calcarine (CAL.L) (x=-27 mm, y=-64 mm, z=7 mm, voxel size=491, $P_{AlphaSim-corrected}$<0.001, Figure 3). *Post hoc* test results showed that the gray matter volume of risk TT allele carriers was significantly decreased in CAL.L.

We then assessed the effect of rs2133896 on white matter. After Bonferroni correction, significant interaction effect of *ANKS1B* rs2133896 × group was found in the right superior longitudinal fasciculus (SLF). *Post hoc* test results showed that, compared with G allele carriers, the mean FA of the right SLF was significantly increased in risk TT allele carriers of

heroin abusers (Figure 3, Table S1).

3.7 *anks1b* modulates gene expression in animal models for drug addiction

We investigated the role of *anks1b* in drug addiction using addiction models. First, we examined the expression of *anks1b* in VTA (one of the most critical brain areas for addiction) in rats that received daily administrations of methamphetamine or heroin for 16 days. We found that VTA *anks1b* expression was significantly decreased in both methamphetamine (t_8=3.411, P=0.09) and heroin treated rats (t_{10}=3.085, P=0.02) (Figure 4A and B).

We then designed a virus vector (rAAV2/9-CMV-ANKS1B-flag-pA) to increase the expression of *anks1b* in VTA (t_{12}=-2.244, P=0.045). After virus infection, rats were trained to self-administer methamphetamine or heroin. The dose effect test was then administered after training (Figure 4C and D). The total number of activated nose pokes and self-injections were no different between the over-expression group and control group (for both methamphetamine and heroin training) (Figure 4E, F, H and I). The repeated measures ANOVA (2 groups × 12 days) revealed no interaction between the groups or doses. Following acquisition, the dose responses for methamphetamine and heroin lowered in the *anks1b* over expression group. The number of injections significantly decreased for the over expression group for methamphetamine 0.03mg/kg ($t_{(1,16)}$=3.147, P=0.006) and heroin 0.1mg/kg ($t_{(1,17)}$=2.750, P=0.014) (Figure 4G and J). Repeated measures ANOVA (2 groups × 5 different doses) revealed the main effect of group on response rates: for methamphetamine group ($F_{(1,16)}$=13.965, P < 0.000) and for heroin group ($F_{(1,17)}$=5.530, P=0.021).

4. Discussion

This is the first GWAS analyses performed using a very large cohort of combined alcohol, heroin and methamphetamine dependent users. This large cohort resulted in a substantial increase in power to identify critical genetic etiology associated with SD. We found three specific SNPs (*ANKS1B* rs2133896, *AGBL4* rs147247472, *CTNNA2* rs10196867) that were synchronously associated with alcoholism, heroin and methamphetamine dependence. Their

significant associations with heroin and methamphetamine dependence were validated. In addition, we found that rs2133896 on *ANKS1B* affected gene expression, frequency of heroin abuse and changes in brain images of heroin patients. Moreover, the role of *ANKS1B* in addiction vulnerability was underscored in rat models for drug dependence.

Addictive substances have several different active mechanisms but converge on the reward pathways[2]. Genetics contribute to about 40%–60% of SD vulnerability [26]. Hence genetics is an important way to assess common neural and molecular mechanisms for addictive disorders. This can be accomplished using combined GWAS of the most commonly used addictive substances. We identified several novel genetic variants that may contribute to the generalized vulnerability to addiction. Gene-set convergence results suggest that the "Wnt signaling pathway", which is involved in the occurrence and development of psychiatry disorders[27], could play a critical role in the common risk of SD. Several distinctive genetic factors contribute to predisposing individuals to use or misuse specific classes of psychoactive substances[28]. Examples of which include the strong association of *ADH1B* rs1229984 with alcoholism. Compared to illegal drug dependence, alcohol has specific metabolic enzymes and relatively weaker effects on reinforcing properties[29,30]. The key genetic components in alcohol metabolizing enzyme genes (e.g., *ADH* and *ALDH* variants) were reliably and specifically correlated with alcoholism[9,31]. As our validation results suggested, the cross-addiction related genetic factors may have relatively uncertain association with alcoholism and needs further mechanistic studies to substantiate their association.

ANKS1B gene (aliases: EB-1, AIDA-1) encodes an activity-dependence effector of post-synaptic signaling[32,33], and may act as a synapse-to-nucleus messenger in the regulation of synaptic plasticity and control of protein biosynthetic capacity (GO Source: InterPro)[34]. *ANKS1B* is predominantly expressed in the brain (Figure 2B). Homozygous *anks1b* KO genotype is partially lethal and exhibits locomotor hyperactivity and increased stereotypy[35]. *ANKS1B* polymorphisms are associated with common genetic risks that share effects on five major psychiatric disorders[36]. In addition, genetic variants of *ANKS1B* were among the top hits in several antipsychotic drug response GWAS[37-39]. These variants also modulate

monoamine metabolite levels in the cerebrospinal fluid [40] that play a role in psychiatry disorders. As our results suggest, *anks1b* modulates the sensibility to drug addiction, which suggests its involvement in resisting addiction. We found that rs2133896 affected *ANKS1B* gene expression, both in the gray and white matter. Phenotype-genetics analysis showed the effect of rs2133896 on heroin usage frequency and further anchored CAL and SLF as its effect-areas for heroin dependence. The visual cortical, which center in CAL, is the main sensory associations that contribute to cue-induced drug craving[41]. SLF is the critical association fiber tract that links the frontal, occipital, parietal and temporal lobes[42]. Hence, we speculate patients who carry the risk rs2133896 TT genotype may facilitate the transition of drug-related sensory stimulus though abnormal white matter connectivity in SLF to aggravate addiction development.

We also identified two significant loci with low MAF, i.e., *AGBL4* rs147247472 and *CTNNA2* rs10196867. *AGBL4* is an ATP/GTP binding protein like 4 gene and is expressed in brain (supplemental Figure 2A). *AGBL4* variants have been reported in several GWAS of SD[43,44]. The age-related methylation changes in *AGBL4* are associated with cognitive function[45]. Whereas rs147247472 located in the *AGBL4* intron region has been rarely reported. *CTNNA2* encodes a cell-adhesion protein-catenin alpha 2 and is strongly expressed in the central nervous system, predominately in the prefrontal cortex (Supplementary Figure 2B). *CTNNA2* is engaged in regulating synaptic plasticity[46], brain morphogenesis and behavior performance[47]. *CTNNA2* variants show genome-wide significance in GWAS of ADHD[48] and excitement-seeking personality[49]. *CTNNA2* has been constantly identified in addiction GWAS[50], which suggest the important role of *CTNNA2* variants in addiction-related behaviors. Given the low MAF of the loci and our limited sample size, additional replication and biological mechanistic studies are necessary for bridging the association between these novel variants and addiction.

There were several limitations to our study. First, the dynamic interaction between genetic, environmental factors and pharmacologic effects is complex; hence the definition of healthy controls and patients with specific addictions would be time-specific. Second, the study was limited to a single ethnic group, and hence needs to be replicated in other types of addiction

that includes additional ethnic groups. In addition, the majority of our patients were concurrent nicotine users (Table 1); hence the confounding effects of smoking were difficult to discriminate.

In summary, our findings revealed several novel genome-wide significant SNPs and genes associated with common susceptibility and phenotypic changes for alcoholism, heroin and MA dependence in a large combined cohort. These findings shed light to the root cause and categorical distinctions of different SD and will help develop early prevention measures for all SD.

Acknowledgements

This work was supported by grants from the National Basic Research Program of China (No. 2015CB553503), the National Natural Science Foundation of China (No. U180220091, 81821092, 81601165), the National Key Research and Development Program of China (No. 2017YFC0803608 and No. 2017YFC0803609), Beijing Municipal Science & Technology Commission (Z181100001518005), and Youth Elite Scientists Sponsorship Program by CASR (No. CSTQT2017002).

Author contributions

J.S. Y.S. and S-H.C. designed the study and obtained financial support; Y.S., F.W., W.-H.Y., H.-Q.S., Z.-J.N. and X.-W.Z. conducted cohort recruitment, collected biological samples and phenotypic data. S-H.C and Y.S. performed genetic data processing, statistical and bioinformatics analysis. Z.L. performed the imaging. L.-B.Z., Y.-B.Z. and Y.C performed the animal and in vitro experiments. Y.S and S.-H.C. drafted the manuscript. J.S and L.L. supervised the experiments and data analysis. All authors critically reviewed the manuscript and approved the final version.

Competing interests

The authors declare no competing financial interests.

References

1. UNODC. World Drug Report 2018. (United Nations publication, 2018).

2. Nestler, E.J. Is there a common molecular pathway for addiction? *Nature Neuroscience* **8**, 1445-1449 (2005).

3. Vanyukov, M.M. *et al.* Liability to substance use disorders: 1. Common mechanisms and manifestations. *Neuroscience and Biobehavioral Reviews* **27**, 507-515 (2003).

4. Krueger, R.F. *et al.* Etiologic connections among substance dependence, antisocial behavior, and personality: Modeling the externalizing spectrum. *Journal of Abnormal Psychology* **111**, 411-424 (2002).

5. Kendler, K.S., Karkowski, L.M., Neale, M.C. & Prescott, C.A. Illicit psychoactive substance use, heavy use, abuse, and dependence in a US population-based sample of male twins. *Archives of General Psychiatry* **57**, 261-269 (2000).

6. Buhler, K.M. *et al.* Common single nucleotide variants underlying drug addiction: more than a decade of research. *Addiction Biology* **20**, 845-871 (2015).

7. Palmer, R.H.C. *et al.* Examining the role of common genetic variants on alcohol, tobacco, cannabis and illicit drug dependence: genetics of vulnerability to drug dependence. *Addiction* **110**, 530-537 (2015).

8. Agrawal, A., Neale, M.C., Prescott, C.A. & Kendler, K.S. Cannabis and other illicit drugs: Comorbid use and abuse/dependence in males and females (vol 34, pg 217, 2004). *Behavior Genetics* **34**, 557-557 (2004).

9. Kendler, K.S., Jacobson, K.C., Prescott, C.A. & Neale, M.C. Specificity of genetic and environmental risk factors for use and abuse/dependence of cannabis, cocaine, hallucinogens, sedatives, stimulants, and opiates in male twins. *American Journal of Psychiatry* **160**, 687-695 (2003).

10. Uhl, G.R. *et al.* "Higher order" addiction molecular genetics: Convergent data from genome-wide association in humans and mice. *Biochemical Pharmacology* **75**, 98-111 (2008).

11. Li, M.D. & Burmeister, M. New insights into the genetics of addiction. *Nat Rev Genet* **10**, 225-31 (2009).

12. Reyes-Gibby, C.C., Yuan, C., Wang, J., Yeung, S.C.J. & Shete, S. Gene network analysis shows immune-signaling and ERK1/2 as novel genetic markers for multiple addiction phenotypes: alcohol, smoking and opioid addiction. *Bmc Systems Biology* **9**(2015).

13. Shields, A.L., Howell, R.T., Potter, J.S. & Weiss, R.D. The Michigan Alcoholism Screening Test and its shortened form: a meta-analytic inquiry into score reliability. *Subst Use Misuse* **42**, 1783-800 (2007).

14. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3 (Bethesda)* **1**, 457-70 (2011).

15. Delaneau, O., Marchini, J. & Zagury, J.F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-81 (2012).

16. Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes.

Nature **491**, 56-65 (2012).

17. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-13 (2007).

18. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-7 (2010).

19. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).

20. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).

21. Consortium, G.T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).

22. Trabzuni, D. *et al.* Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies. *J Neurochem* **119**, 275-82 (2011).

23. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219 (2015).

24. Yan-Xue Xue, Y.-X.L., Ping Wu, Hai-Shui Shi, Li-Fen Xue, Chen Chen, Wei-Li Zhu, Zeng-Bo Ding, Yan-ping Bao, Jie Shi, David H. Epstein, Yavin Shaham, Lin Lu. A Memory Retrieval-Extinction Procedure to Prevent Drug Craving and Relapse. *Science* **336**, 241-245 (2012).

25. Zhang, Y. *et al.* Inhibition of Lactate Transport Erases Drug Memory and Prevents Drug Relapse. *Biol Psychiatry* **79**, 928-39 (2016).

26. Uhl, G.R. Molecular genetic underpinnings of human substance abuse vulnerability: likely contributions to understanding addiction as a mnemonic process. *Neuropharmacology* **47 Suppl 1**, 140-7 (2004).

27. Hoseth, E.Z. *et al.* Exploring the Wnt signaling pathway in schizophrenia and bipolar disorder. *Translational Psychiatry* **8**(2018).

28. Kendler, K.S., Myers, J. & Prescott, C.A. Specificity of genetic and environmental risk factors for symptoms of cannabis, cocaine, alcohol, caffeine, and nicotine dependence. *Archives of General Psychiatry* **64**, 1313-1320 (2007).

29. Grupp, L.A. An Investigation of Intravenous Ethanol Self-Administration in Rats Using a Fixed-Ratio Schedule of Reinforcement. *Physiological Psychology* **9**, 359-363 (1981).

30. Crowley, T.J. The reinforcers for drug abuse: why people take drugs. *Compr Psychiatry* **13**, 51-62 (1972).

31. Tsuang, M.T. *et al.* Co-occurrence of abuse of different drugs in men: the role of drug-specific and shared vulnerabilities. *Arch Gen Psychiatry* **55**, 967-72 (1998).

32. Ghersi, E., Vito, P., Lopez, P., Abdallah, M. & D'Adamio, L. The intracellular localization of Amyloid beta Protein Precursor (A beta PP) intracellular domain associated protein-1 (AIDA-1) is regulated by A beta PP and alternative splicing. *Journal of Alzheimers Disease* **6**, 67-78 (2004).

33. Jordan, B.A. *et al.* Identification and verification of novel rodent postsynaptic density proteins.

Mol Cell Proteomics **3**, 857-71 (2004).

34.   Jordan, B.A., Fernholz, B.D., Khatri, L. & Ziff, E.B. Activity-dependent AIDA-1 nuclear signaling regulates nucleolar numbers and protein synthesis in neurons. *Nat Neurosci* **10**, 427-35 (2007).

35.   Enga, R.M. *et al.* Initial characterization of behavior and ketamine response in a mouse knockout of the post-synaptic effector gene Anks1b. *Neuroscience Letters* **641**, 26-32 (2017).

36.   Cross-Disorder Group of the Psychiatric Genomics, C. *et al.* Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371-9 (2013).

37.   McClay, J.L. *et al.* Genome-wide pharmacogenomic analysis of response to treatment with antipsychotics. *Mol Psychiatry* **16**, 76-85 (2011).

38.   McClay, J.L. *et al.* Genome-Wide Pharmacogenomic Study of Neurocognition As an Indicator of Antipsychotic Treatment Response in Schizophrenia. *Neuropsychopharmacology* **36**, 616-626 (2011).

39.   Garriock, H.A. *et al.* A Genomewide Association Study of Citalopram Response in Major Depressive Disorder. *Biological Psychiatry* **67**, 133-138 (2010).

40.   Luykx, J.J. *et al.* Genome-wide association study of monoamine metabolite levels in human cerebrospinal fluid. *Mol Psychiatry* **19**, 228-34 (2014).

41.   Yalachkov, Y., Kaiser, J. & Naumer, M.J. Functional neuroimaging studies in addiction: multisensory drug stimuli and neural cue reactivity. *Neurosci Biobehav Rev* **36**, 825-35 (2012).

42.   Wang, X. *et al.* Subcomponents and connectivity of the superior longitudinal fasciculus in the human brain. *Brain Struct Funct* **221**, 2075-92 (2016).

43.   Johnson, C. *et al.* Genome wide association for substance dependence: convergent results from epidemiologic and research volunteer samples. *BMC Med Genet* **9**, 113 (2008).

44.   Drgon, T. *et al.* Genome wide association for addiction: replicated results and comparisons of two analytic approaches. *PLoS One* **5**, e8832 (2010).

45.   Starnawska, A. *et al.* Epigenome-Wide Association Study of Cognitive Functioning in Middle-Aged Monozygotic Twins. *Front Aging Neurosci* **9**, 413 (2017).

46.   Abe, K., Chisaka, O., van Roy, F. & Takeichi, M. Stability of dendritic spines and synaptic contacts is controlled by alpha N-catenin. *Nature Neuroscience* **7**, 357-363 (2004).

47.   Park, C., Falls, W., Finger, J.H., Longo-Guess, C.M. & Ackerman, S.L. Deletion in Catna2, encoding alpha N-catenin, causes cerebellar and hippocampal lamination defects and impaired startle modulation. *Nature Genetics* **31**, 279-284 (2002).

48.   Lesch, K.P. *et al.* Molecular genetics of adult ADHD: converging evidence from genome-wide association and extended pedigree linkage studies. *J Neural Transm (Vienna)* **115**, 1573-85 (2008).

49.   Terracciano, A. *et al.* Meta-analysis of genome-wide association studies identifies common variants in CTNNA2 associated with excitement-seeking. *Transl Psychiatry* **1**, e49 (2011).

50.   Hall, F.S., Drgonova, J., Jain, S. & Uhl, G.R. Implications of genome wide association studies for addiction: are our a priori assumptions all wrong? *Pharmacol Ther* **140**, 267-79 (2013).

Figure Legends

Figure 1 GWAS association analysis. (A) Manhattan plot for the GWAS of the combined substance dependence cohort. (B) Q-Q plot for the combined substance dependence GWAS p-values. Regional plots for the four genome-wide significant loci: *ANKS1B* rs2133896 (C), *AGBL4* rs147247472 (D), *CTNNA2* rs10196867 (E) and *ADH1A* rs1229984 (F).

Figure 2. **Stratified expression of *ANKS1B* by rs10860447 and *ANKS1B* gene expression profile.** （A）Effect of *ANKS1B* rs10860447 (LD-proxy with rs2133896 $r^2 = 0.902$) on gene expression in brain using BRAINEAC data. C allele (linked with rs2133896 T allele) had higher *ANKS1B* expression in white matter and lower expression in gray matter of the cerebellum. (B) *ANKS1B* is specifically expressed in brain (GTEx data).

Figure 3. **Interaction effects of ANKS1B rs2133896 by heroin on brain imaging.** (A) Significant interaction effect of *ANKS1B* rs2133896 (TT genotype and G carriers) × group (heroin and healthy) on the gray matter volume of the left calcarine (CAL.L) (B) *Post hoc* test results for CAL.L. (C) After Bonferroni correction, significant interaction effect of genotype × group was found in the right superior longitudinal fasciculus (SLF). (D) *Post hoc* test results for right SLF.

Figure 4. Role of *anks1b* in drug addiction animal models. (A) Timeline of drug administration procedure. (B) Western blot analysis of *anks1b* in VTA after drug treatment. (C) Timeline of the self-administration procedure. (D) Western blot analysis of *anks1b* in VTA from rats injected with AAV-ANKS1B or AAV-CMV. (E) Number of responses (mean ± SEM) on the active and inactive nosepoke devices during methamphetamine training sessions. (F) Number of methamphetamine self-injections in the over-expression and control groups. (G) Following acquisition, the dose response for methamphetamine was lower in the *anks1b* over expression group. (H) Number of responses on the active and inactive nosepoke devices during the heroin training sessions. (I) Number of heroin self-injections in the *anks1b* over expression and control groups. (J) Dose response for heroin was lower in the *anks1b* over-expression group. *P<0.05 compared to the control.

Table 1. Demographic and addiction characteristics of the discovery and validation cohorts.

	Discovery cohort				Replication cohort			
	MA	Heroin	Alcohol	Control	MA	Heroin	Alcohol	Control
N	1749	1026	521	2859	763	1045	146	1904
Gender (male/female)	1255/494	737/289	521/0	1851/1008	763/0	1045/0	146/0	1904/0
Age (year)	30.51 ± 7.40	35.6 ± 6.95	45.17 ± 9.27	34.17 ± 6.41	36.36 ± 8.52	40.06 ± 7.93	41.79 ± 8.90	33.71 ± 11.80
Education level (year)	9.34 ± 2.35	8.31 ± 2.69	10.37 ± 2.80	-	8.25 ± 2.54	7.66 ± 2.33	11.08 ± 3.20	-
Smoker (%)	88.1%	97.1%	94.9%	-	99.7%	99.9%	88.4%	-
Onset age of addiction	24.13 ± 7.00	23.13 ± 6.49	22.39 ± 6.77	-	28.82 ± 9.00	26.36 ± 8.00	17.43 ± 4.29	-
Dosage (g/per time)	0.65 ± 0.48	0.66 ± 0.60	-	-	0.53 ± 0.62	0.30 ± 0.19	-	-
Frequency (times per day)	1.51 ± 1.71	3.67 ± 2.36	-	-	2.03 ± 1.53	2.76 ± 1.49	-	-
MAST*	-	-	10.22 ± 5.14	-	-	-	28.53 ± 10.16	-

* Michigan Alcoholism Screening Test (MAST) was measured for the severity of alcoholism.

Table 2. GWAS analysis and validation results.

SNP	Chr	A1	Freq[a]	Gene	Addiction	Discovery			Replication			Meta		
						BETA	SE	P	BETA	SE	P	BETA	SE	P
rs2133896	12	T	0.2215	ANKS1B	All	**0.102**	**0.019**	**4.09×10^{-8}**	0.129	0.058	0.026	0.104	0.018	**3.60×10^{-9}**
					Alcohol	0.233	0.116	0.044	-0.128	0.161	0.426	0.110	0.094	0.242
					MA	0.093	0.020	3.44×10^{-6}	0.204	0.074	5.79×10^{-3}	0.100	0.019	1.97×10^{-7}
					Heroin	0.080	0.019	4.30×10^{-5}	0.092	0.071	0.192	0.081	0.019	1.78×10^{-5}
rs147247472	1	A	0.0174	AGBL4	All	**0.286**	**0.052**	**4.30×10^{-8}**	**1.051**	**0.175**	**2.03×10^{-9}**	0.348	0.050	**3.40×10^{-12}**
					Alcohol	0.742	0.360	0.040	NA	NA	NA	N.A.	N.A.	N.A.
					MA	0.222	0.055	5.89×10^{-5}	1.038	0.199	1.72×10^{-7}	0.281	0.053	1.34×10^{-7}
					Heroin	0.215	0.052	4.05×10^{-5}	**1.078**	**0.197**	**4.66×10^{-8}**	**0.272**	**0.051**	**7.88×10^{-8}**
rs10196867	2	C	0.0297	CTNNA2	All	**0.239**	**0.044**	**4.67×10^{-8}**	0.275	0.129	0.033	0.243	0.042	**4.73×10^{-9}**
					Alcohol	0.610	0.272	0.025	NA	NA	NA	NA	NA	NA
					MA	0.216	0.046	3.30×10^{-6}	0.098	0.171	0.565	0.208	0.045	3.50×10^{-6}
					Heroin	0.181	0.045	4.83×10^{-5}	0.409	0.147	5.46×10^{-3}	0.200	0.043	2.67×10^{-6}
rs1229984	4	C	0.299	ADH1B	All	**0.105**	**0.017**	**6.45×10^{-10}**	NA	NA	NA	NA	NA	NA
					Alcohol	**0.677**	**0.095**	**1.03×10^{-12}**	**2.038**	**0.156**	**4.80×10^{-39}**	**1.045**	**0.081**	**5.32×10^{-38}**
					MA	0.000	0.019	0.986	NA	NA	NA	N.A.	N.A.	N.A.
					Heroin	0.001	0.019	0.942	NA	NA	NA	N.A.	N.A.	N.A.

[a] Frequency of the SNP A1 allele in the discovery stage.

A

Affymetrix ID t3467351 stratified by rs10860447 (chr12:99890240)

Genotype counts: TT=9; TC=43; CC=82   Allele frequency: T:22.8%; C:77.2%

Source: BRAINEAC

| eQTL p-value | OCTX 0.54 | TCTX 0.32 | FCTX 0.45 | PUTM 0.58 | CRBL 0.043 | HIPP 0.75 | THAL 0.87 | SNIG 0.7 | WHMT 0.033 | MEDU 0.92 |

B

***ANKS1B* expression in human of GTEx data**