Supplementary Online Content

Jung Y, Montel RA, Shen P-H, Mash DC, Goldman D. Assessment of the association of D2dopamine receptor gene and reported allele frequencies with alcohol use disorders: a systematic review and meta-analysis. *JAMA Netw Open*. 2019;2(11):e1914940. doi:10.1001/jamanetworkopen.2019.14940

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix 1. Sample, Diagnosis, and Consent Information

General description

- All the SNPs were genotyped using smokescreen array (https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-2495-7). SNPs with MAF>0.005 were included in the plots.

Finnish Caucasians

- The sample from Helsinki, Finland, has been described in detail elsewhere (GABRG1 and GABRA2 as independent predictors for alcoholism in two populations, Enoch et al).[1] All participants gave written informed consent to the study that was approved by the IRBs of the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the National Institute of Mental Health (NIMH), the University of Helsinki, and the University of Helsinki Central Hospital, Helsinki, Finland.
- Amount 641 genotyped subjects in the plots, 533 were males and 108 were females.
- There are 55 Alcohol Abuse cases and 586 controls. There are 289 Alcohol Dependent or Abuse cases, 352 controls. There are 234 Alcohol Dependent cases, 407 controls. Psychiatric diagnoses were based on DSM-III.
- Logistic regression was performed with European AIMs scores as covariates. (Using ancestry-informative markers to define populations and detect population stratification. Enoch et al).[1]

African American

- African-American patients were recruited from the Substance Abuse Treatment Program (SATP) at the Department of Veteran Affairs New Jersey Healthcare System (VANJHCS), East Orange Campus. A detailed description of this sample has previously been provided

(Two HPA axis genes, CRHBP and FKBP5, interact with childhood trauma to increase the risk for suicidal behavior. Roy, 2012). All participants gave written informed consent to the study that was approved by the Institutional Review Boards (IRBs) of the Department of Veteran Affairs New Jersey Healthcare System (VANJHCS) and the University of Medicine and Dentistry, New Jersey Medical School (UMDNJ).

- Amount 583 genotyped subjects in the plots, 356 were males and 227 were females.
- There are 163 Cocaine dependent cases and 420 controls. There are 119 Heroine Dependent cases, 464 controls. There are 161 Alcohol Dependent cases, 422 controls. Psychiatric diagnoses were based on DSM-IV.
- Logistic regression was performed with Africa AIMs scores as covariates. (Using ancestry-informative markers to define populations and detect population stratification. Enoch et al).[1]

American Indian

- From Southwestern American Indian tribe. All subjects gave informed consent under a human research protocol approved by the Institutional Review Board of the National Institute on Alcohol Abuse and Alcoholism and the Tribal Council. Most participants derived from three large interrelated pedigrees; however, the average sharing of descent between any two individuals was 0.012 which is equivalent to the relationship between second cousins once removed and third cousins. For a more detailed description see Relationship of binge drinking to alcohol dependence, other psychiatric disorders, and behavioral problems in an American Indian tribe. Robin et al).[2]
- Amount 501 genotyped subjects in the plots, 217 were males and 283 were females.

- There are 30 Alcohol Abuse cases and 471 controls. There are 353 Alcohol Dependent or Abuse cases, 148 controls. There are 323 Alcohol Dependent cases, 178 controls. Psychiatric diagnoses were based on DSM-III.
- Logistic regression was performed with American Indian AIMs scores as covariates. (Using ancestry-informative markers to define populations and detect population stratification. Enoch et al).[1]

eReferences

1. Enoch M-A, Shen P-H, Xu K, Hodgkinson C, Goldman D. Using ancestry-informative markers to define populations and detect population stratification. *J Psychopharmacol Oxf Engl*. 2006;20(4 Suppl):19-26. doi:10.1177/1359786806066041

2. Robin RW, Long JC, Rasmussen JK, Albaugh B, Goldman D. Relationship of binge drinking to alcohol dependence, other psychiatric disorders, and behavioral problems in an American Indian tribe. *Alcohol Clin Exp Res*. 1998;22(2):518-5236.

eAppendix 2. Methods

Array based genotyping

Array-based SNP genotyping was performed using Illumina GoldenGate protocols on 96-well format BioRealm Smokescreen arrays. Five hundred nanograms of genomic DNA was used per assay. Pre-PCR DNA processing was performed using a TECAN liquid handling robot running Illumina protocols. Arrays were imaged using an Illumina Beadstation GX500 and data analyzed using GenCall v6.2.0.4 and GTS Reports software v5.1.2.0 (Illumina). Genotype clusters were determined for a test dataset and this template was applied to all subsequent datasets. Data were polished by manual adjustment of the clustering for each SNP to correct for differences between datasets arising from sample integrity and concentration. The GenCall score is a value between 0

and 1 giving a confidence score for that genotype call (the higher the score the higher the confidence in the call) and is derived from the tightness of the clusters for a given locus and the position of the sample relative to its cluster. Loci with call rate <90% were determined to have failed and were excluded. At this point deviation from Hardy–Weinberg equilibrium was not used as an exclusion criterion because all datasets contained both case and control samples, potentially driving deviation from HWE.

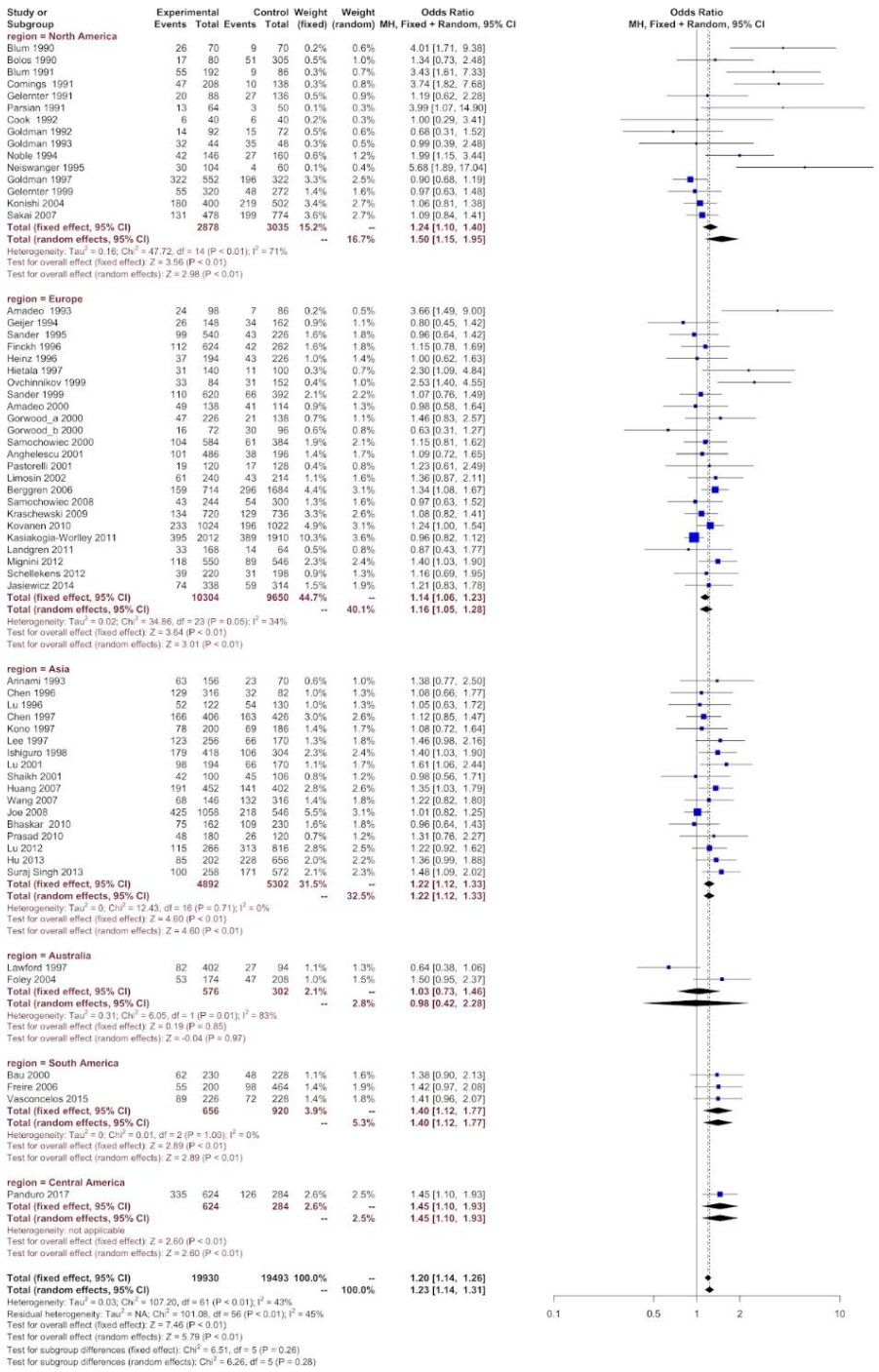
Differential allelic expression (DAE) and SNP genotyping in human brain

Postmortem brain tissue was provided by the University of Miami Brain Bank. Specimens were obtained at autopsy from chronic cocaine and alcohol abusers and age-matched drug- and alcohol-free control subjects. All subjects died suddenly without a prolonged agonal state. Brain pH measures were done as quality control for each case with values > 6.0. Postmortem samples were genotyped for rs1800497 and rs62755, the latter being the reporter locus for DAE. These SNPs were PCR-amplified and genotyped in all subjects using custom 5' exonuclease primers and probes (Life Technologies). Genotypes were resolved by size using an ABI 3730 Capillary Sequencer and GeneMapper Software v4.0.

***DRD2* differential allelic expression**

To determine if there is a cis-acting eQTL affecting expression of *DRD2*, and if rs1800497 is a cis-acting eQTL, the expression of alleles for the reporter SNP rs62755 was compared in 28 heterozygous brain samples identified from a larger (N=82) number of brains. The reporter SNP was selected on the basis that it was exonic with an allele frequency approaching 0.5, thereby ensuring that multiple heterozygous samples would be available for analysis. The messenger RNA levels for both alleles were simultaneously quantified by real-time PCR on hippocampal

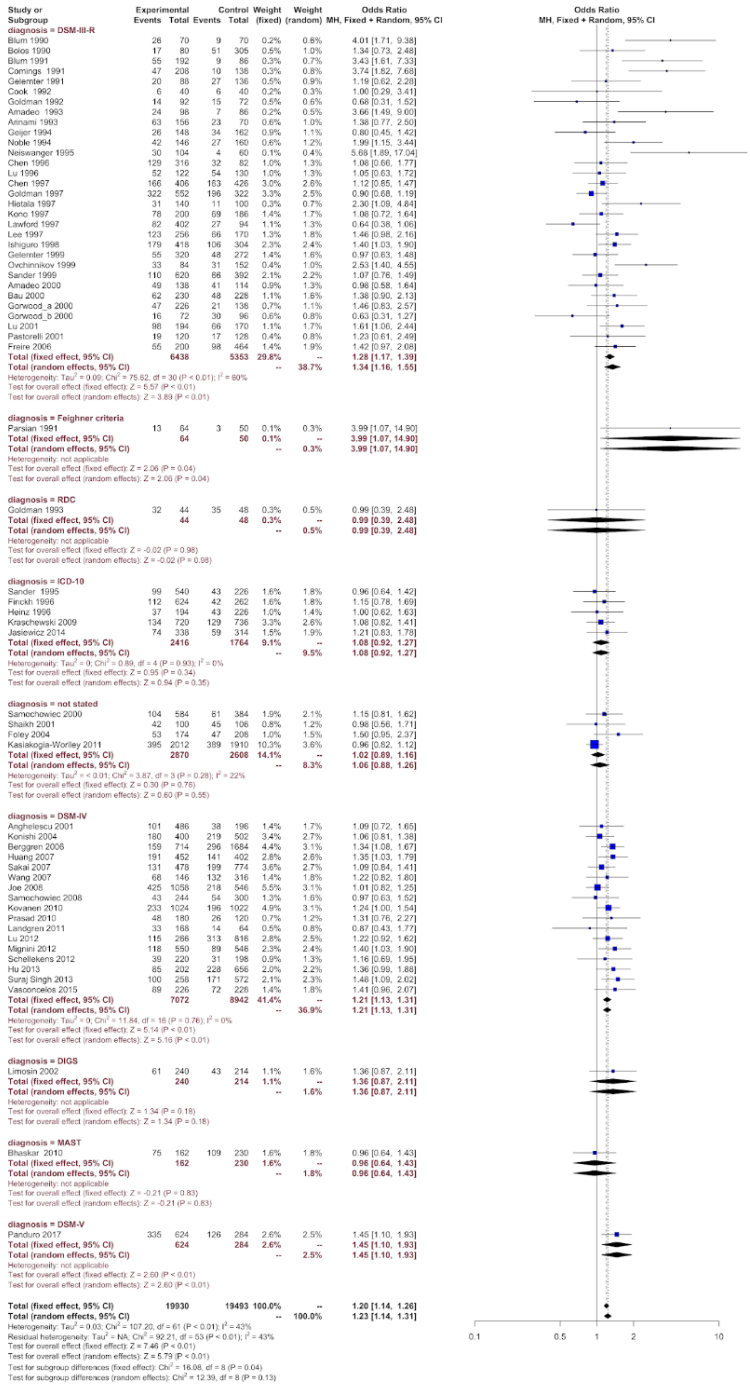
RNA using a Custom Taqman SNP Genotyping Assay (Life Technologies). The complementary DNA (cDNA) was quantified on a QuantStudio 7 Flex Real-Time PCR System (Life Technologies) in a 10 μL qRT-PCR reaction: 2 μL cDNA, 5 μL Amplitaq Gold 360 PCR MasterMix, 0.25 μL custom primer/probe assay, 2.75 μL water. The reaction conditions were as follows: 10 min at 95 °C, and 45 cycles of: 15 s at 95 °C and 1 min at 60 °C. Each sample was analyzed in triplicate. The average difference in Ct (threshold cycle) between the two alleles for cDNA samples was normalized against the average difference in Ct between the two alleles in genomic DNA samples.



eFigure 1. Meta-analysis of *DRD2* rs1800497/AUD Studies Stratified by Region.

Studies were performed in five world regions, Europe, North America, South America, Central

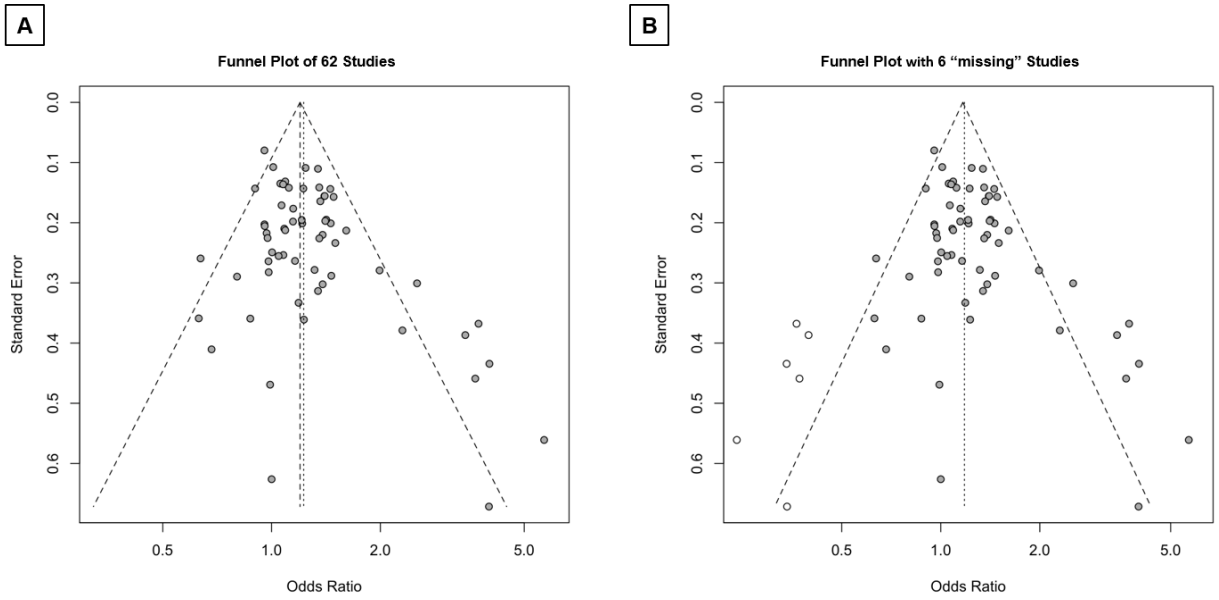
America, Asia and Australia as shown. *Black bars and blue squares* depict 95% confidence intervals (CIs) and odds ratios (ORs) in each study. Pooled effect sizes were estimated under fixed effects and random effects models.



eFigure 2. Meta-analysis of *DRD2* rs1800497/AUD Studies Stratified by Diagnostic Criteria.

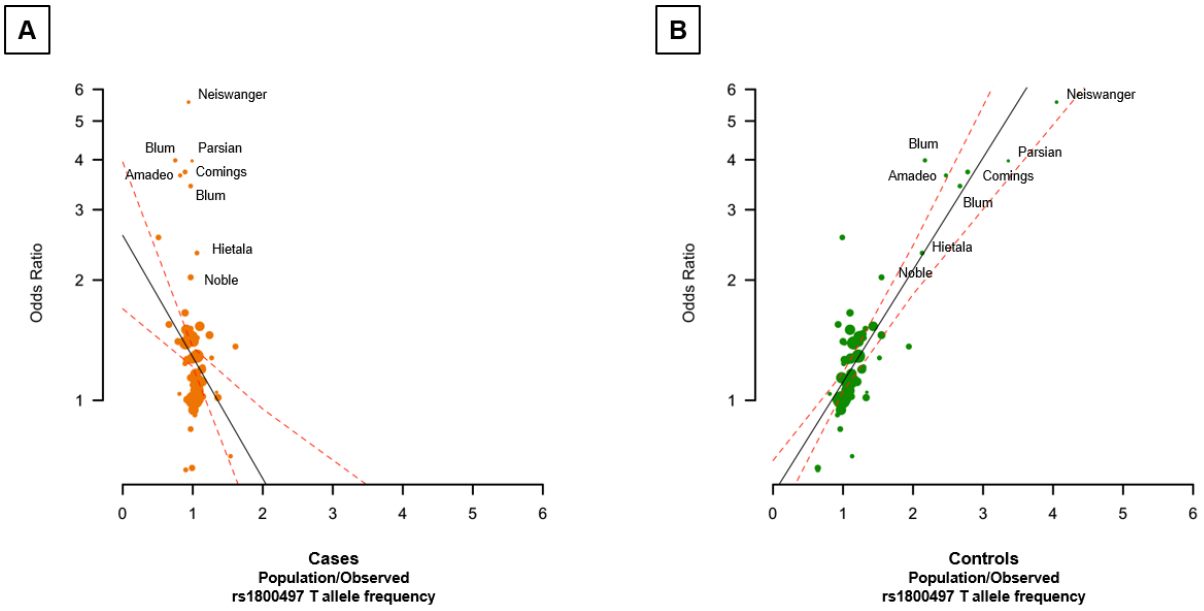
Studies were performed using nine different types of diagnostic criteria as shown. *Black bars*

and blue squares depict 95% confidence intervals (CIs) and odds ratios (ORs) in each study. Pooled effect sizes were estimated under fixed effects and random effects models.



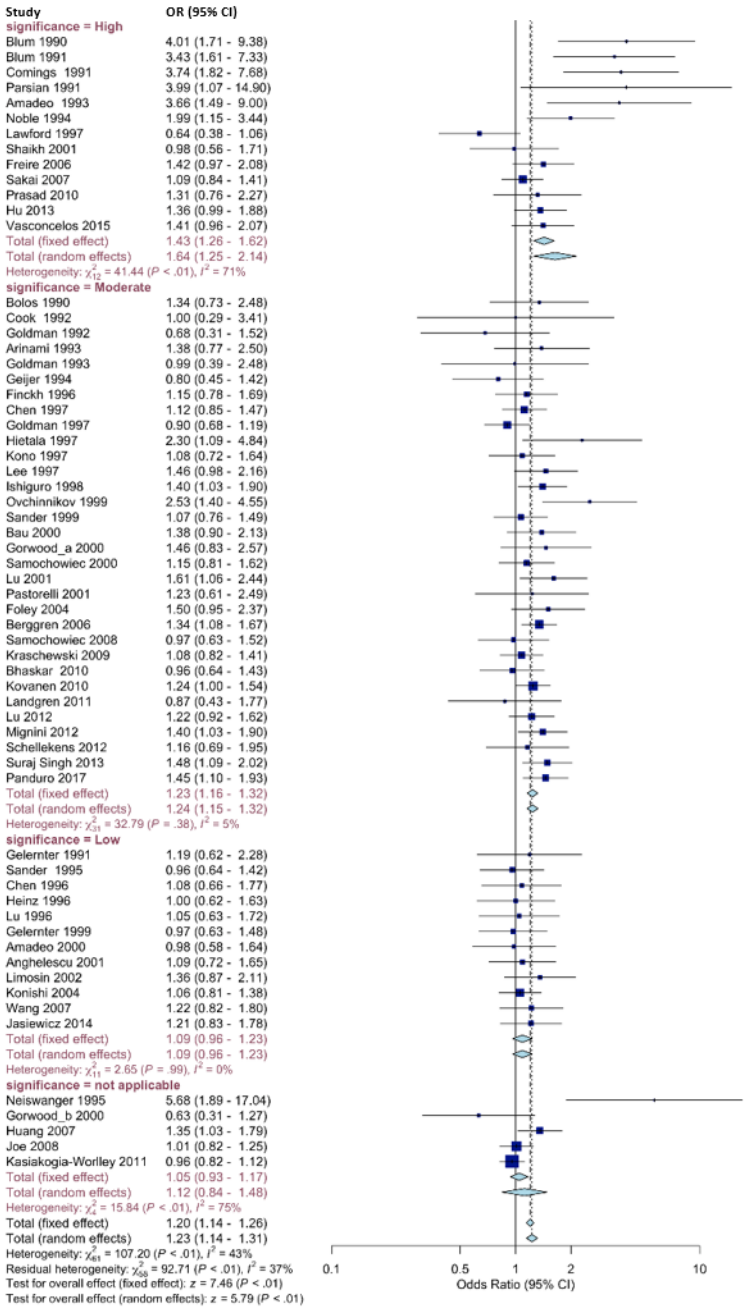
eFigure 3. Publication Bias

(A) Publication bias in *DRD2* (rs1800497)/AUD association studies and (B) prediction of six "missing" studies via the trim and fill model. Solid line at OR=1.23 is the meta-analytic OR derived from 62 published studies. Publication bias was estimated by Begg's rank correlation (p-0.106) and Egger's regression test (p=0.004).



eFigure 4. Meta-Regression Plots

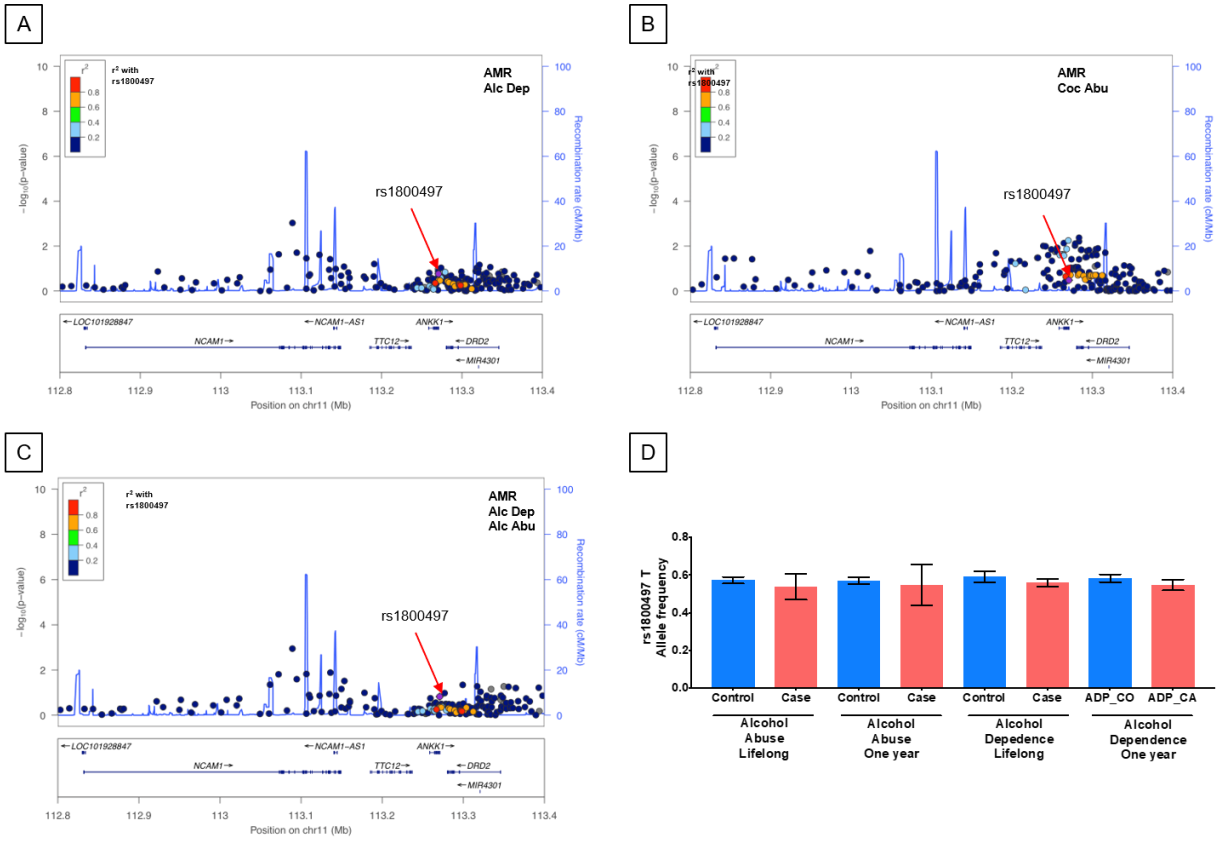
Meta-regression plots of (A) Case allele frequency ratio (B) Control allele frequency ratio (C) Case allele frequency ratio in North American studies (D) Control allele frequency ratio in North American studies. In each case, allele frequency is compared to population allele frequency in ExAC to detect allele frequency deviation. For comparison to 1000 Genomes allele frequencies, see Figure 4. Diameters of circles are proportional to study population size. Solid line represents the meta-regression slope of relationship of odds ratio to allele frequency deviation. (A) z= -3.36, $p<0.0008$ (B) z=7.76, $p<0.0001$.



eFigure 5. Meta-analysis of rs1800497 Association With AUD Stratified by Deviation Significance

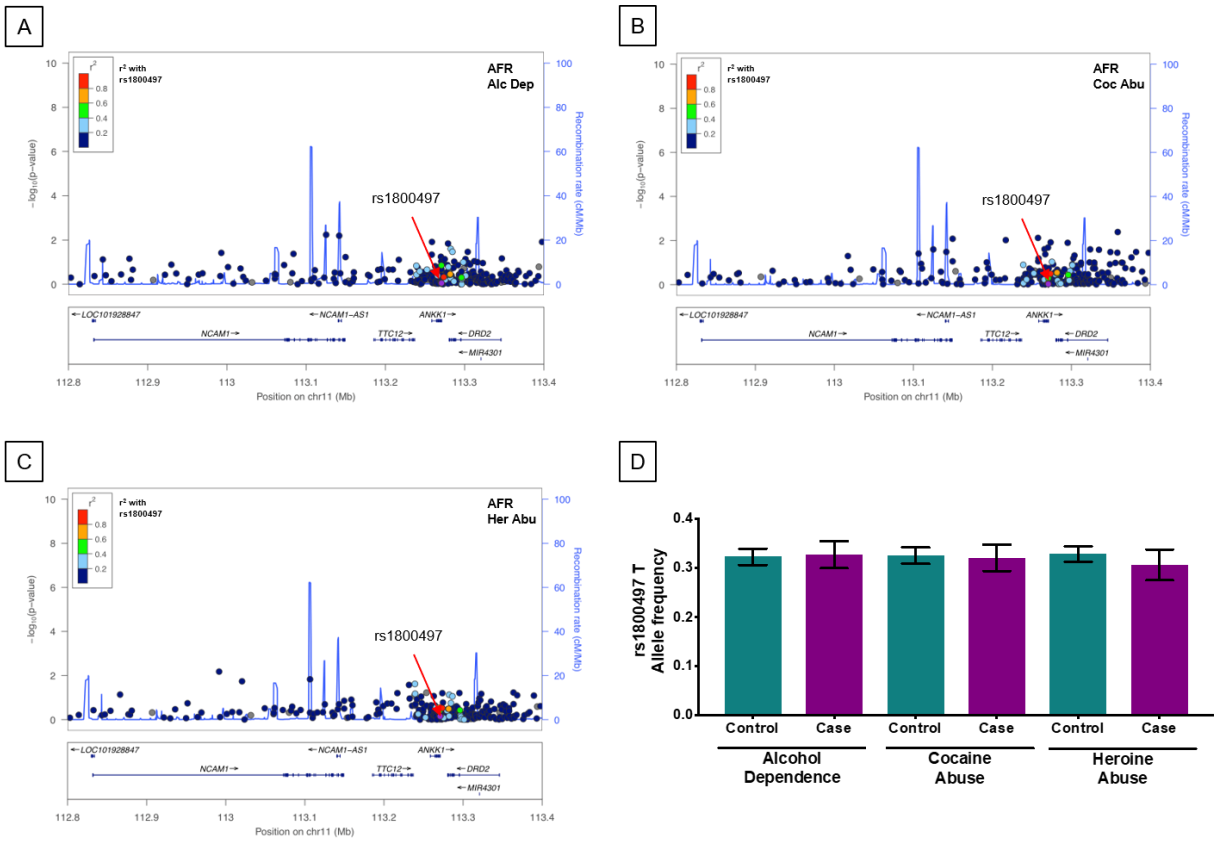
Deviation significance was reported in each study (high, moderate and low). *Black bars and blue squares* depict 95% confidence intervals (CIs) and odds ratios (ORs) in each study. Pooled

effect sizes were estimated under fixed effects and random effects models.



eFigure 6. Association of SNPs in the *DRD2* Region (Native Americans)

Regional association plot of rs1800497 and a total of 201 SNPs in the *DRD2* region to (A) Alcohol dependence, (B) Cocaine abuse, and (C) Alcohol dependence and abuse, in Native Americans (N=501). SNPs are color coded according to linkage disequilibrium (LD) with rs1800497 (purple dot with red arrow) on a scale of r^2 0 to 1. Estimated recombination rates reflect local LD structure in the 600 kb buffer around rs1800497 in this Native American population. rs1800497 allele frequencies (D) did not differ between cases and controls.



eFigure 7. Association of SNPs in the *DRD2* Region (African Americans)

Regional association plot of rs1800497 and a total of 260 SNPs in the *DRD2* region to (A) Alcohol dependence, (B) Cocaine abuse, and (C) Heroin abuse, in African Americans (N=583). SNPs are color coded according to linkage disequilibrium (LD) with rs1800497 (purple dot with red arrow) on a scale of r^2 0 to 1. Estimated recombination rates reflect local LD structure in the 600 kb buffer around rs1800497 in this African American population. rs1800497 allele frequencies (D) did not differ between cases and controls.

eTable. Deviations of Case and Control Allele Frequencies in *DRD2* (rs1800497)/AUD Association Studies (ExAC Allele Frequencies)

Studies	N	Diagnostic Method	Alcohol Use Disorder		Control		Study OR (95% CI)
			Significance of deviation		Significance of deviation		
			E/O	P-value	E/O	P-value	
Blum 1990	140	DSM-III-R	0.89	0.143	2.57	0.017	4.01 (1.71-9.38)
Bolos 1990	385	DSM-III-R	0.89	0.825	0.97	0.872	1.34 (0.73-2.48)
Blum 1991	278	DSM-III-R	1.15	0.845	3.15	0.002	3.43 (1.61-7.33)
Comings 1991	346	DSM-III-R	0.84	0.430	2.62	0.001	3.74 (1.82-7.68)
Gelernter 1991	224	DSM-III-R	0.84	0.589	0.96	0.941	1.19 (0.62-2.28)
Parsian 1991	114	Feighner criteria	0.94	0.975	3.17	0.026	3.99 (1.07-14.90)
Cook 1992	80	DSM-III-R	1.27	0.469	1.27	0.469	1.00 (0.29-3.41)
Goldman 1992	164	DSM-III-R	1.12	0.103	0.82	0.800	0.68 (0.31-1.52)
Amadeo 1993	184	DSM-III-R	0.78	0.337	2.34	0.013	3.66 (1.49-9.00)
Arinami 1993	226	DSM-III-R	1.02	0.658	1.25	0.208	1.38 (0.77-2.50)
Goldman 1993	92	RDC	0.81	0.232	0.81	0.206	0.99 (1.15-3.44)
Geijer 1994	310	DSM-III-R	1.08	0.486	0.91	0.810	0.80 (0.45-1.42)
Noble 1994	306	DSM-III-R	1.15	0.843	1.96	0.008	1.99 (1.15-3.44)
Neiswanger 1995	164	DSM-III-R	0.94	NA	4.05	NA	5.68 (1.89-17.04)
Sander 1995	766	ICD-10	1.04	0.350	1.00	0.709	0.96 (0.64-1.42)
Chen 1996	398	DSM-III-R	1.00	0.610	1.05	0.611	1.08 (0.66-1.77)
Finckh 1996	886	ICD-10	1.06	0.223	1.19	0.138	1.15 (0.78-1.69)
Heinz 1996	420	ICD-10	1.00	0.740	1.00	0.709	1.00 (0.62-1.63)
Lu 1996	252	DSM-III-R	0.91	0.990	0.99	0.839	1.05 (0.63-1.72)
Chen 1997	832	DSM-III-R	1.00	0.576	1.07	0.161	1.12 (0.85-1.47)
Goldman 1997	874	DSM-III-R	1.01	0.862	0.97	0.645	0.90 (0.68-1.19)
Hietala 1997	240	DSM-III-R	0.86	0.751	1.73	0.010	2.30 (1.09-4.84)
Kono 1997	386	DSM-III-R	1.05	0.423	1.11	0.242	1.08 (0.72-1.64)
Lawford 1997	496	DSM-III-R	0.93	0.908	0.66	0.020	0.64 (0.38-1.06)
Lee 1997	426	DSM-III-R	0.85	0.190	1.06	0.439	1.46 (0.98-2.16)
Ishiguro 1998	722	DSM-III-R	0.96	0.969	1.18	0.037	1.40 (1.03-1.90)
Gelernter 1999	592	DSM-III-R	0.99	0.021	0.96	0.048	0.97 (0.63-1.48)
Ovchinnikov 1999	236	DSM-III-R	0.48	0.000	0.93	0.944	2.53 (1.40-4.55)
Sander 1999	1012	DSM-III-R	1.07	0.176	1.13	0.145	1.07 (0.76-1.49)
Amadeo 2000	252	DSM-III-R	0.99	0.920	0.97	0.862	0.98 (0.58-1.64)
Bau 2000	458	DSM-III-R	1.00	0.990	1.28	0.084	1.38 (0.90-2.13)
Gorwood_a 2000	364	DSM-III-R	0.91	0.826	1.25	0.198	1.46 (0.83-2.57)
Gorwood_b 2000	168	DSM-III-R	0.9	NA	0.64	NA	0.63 (0.31-1.27)
Samochowiec 2000	968	not stated	1.07	0.209	1.20	0.063	1.15 (0.81-1.62)
Anghelescu 2001	682	DSM-IV	0.91	0.753	0.98	0.814	1.09 (0.72-1.65)
Lu 2001	364	DSM-III-R	0.81	0.243	1.06	0.439	1.61 (1.06-2.44)
Pastorelli 2001	248	DSM-III-R	1.20	0.293	1.43	0.084	1.23 (0.61-2.49)
Shaikh 2001	206	not stated	0.98	1.000	0.97	0.943	0.98 (0.56-1.71)
Limosin 2002	454	DIGS	0.75	0.069	0.95	0.988	1.36 (0.87-2.11)
Foley 2004	382	not stated	0.62	0.002	0.87	0.640	1.50 (0.95-2.37)
Konishi 2004	902	DSM-IV	0.98	0.262	1.01	0.090	1.06 (0.81-1.38)
Berggren 2006	2398	DSM-IV	0.85	0.205	1.08	0.019	1.34 (1.08-1.67)
Freire 2006	664	DSM-III-R	0.98	0.892	1.28	0.015	1.42 (0.97-2.08)
Huang 2007	854	DSM-IV	1.01	NA	1.22	NA	1.35 (1.03-1.79)
Sakai 2007	1252	DSM-IV	1.20	0.838	1.23	0.822	1.09 (0.84-1.41)
Wang 2007	462	DSM-IV	0.88	0.474	0.98	0.801	1.22 (0.82-1.80)
Joe 2008	1604	DSM-IV	1.06	NA	1.07	NA	1.01 (0.82-1.25)
Samochowiec 2008	544	DSM-IV	1.06	0.381	1.06	0.409	0.97 (0.63-1.52)
Kraschewski 2009	1456	ICD-10	1.02	0.361	1.09	0.114	1.08 (0.82-1.41)
Bhaskar 2010	392	MAST	0.89	0.484	0.87	0.276	0.96 (0.64-1.43)
Kovanen 2010	2046	DSM-IV	0.75	0.343	0.89	0.005	1.24 (1.00-1.54)
Prasad 2010	300	DSM-IV	1.55	0.001	1.86	0.093	1.31 (0.76-2.27)
Kasiakogia-Worlley 2011	3922	not stated	1.02	NA	0.98	NA	0.96 (0.82-1.12)
Landgren 2011	232	DSM-IV	0.97	0.886	0.87	0.757	0.87 (0.43-1.77)
Lu 2012	1082	DSM-IV	0.95	0.894	1.07	0.058	1.22 (0.92-1.62)
Mignini 2012	1096	DSM-IV	0.89	0.492	1.17	0.046	1.40 (1.03-1.90)
Schellekens 2012	418	DSM-IV	1.07	0.425	1.22	0.160	1.16 (0.69-1.95)
Hu 2013	858	DSM-IV	0.97	0.893	1.18	0.002	1.36 (0.99-1.88)
Suraj Singh 2013	830	DSM-IV	1.06	0.333	1.37	0.000	1.48 (1.09-2.02)
Jasiewicz 2014	652	ICD-10	0.87	0.473	1.01	0.594	1.21 (0.83-1.78)
Vasconcelos 2015	454	DSM-IV	1.12	0.040	1.39	0.000	1.41 (0.96-2.07)
Panduro 2017	908	DSM-V	0.89	0.089	1.08	0.273	1.45 (1.10-1.93)