Peer Influence on Binge Drinking Depends on Self Genetic Propensity for Alcohol Use: A Study of Randomly-Assigned Roommates and a Replication

Tianji Cai *and other coauthors*

- Direct correspondence to: Guang Guo, CB#3210, the University of North Carolina, Chapel Hill, NC 27599-3210 (guang_guo@unc.edu).
- Acknowledgement**:** This research uses data from the College Roommate Study funded by a grant from the William T. Grant Foundation and also data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies (www.cpc.unc.edu/addhealth/contract.html). We gratefully acknowledge support from NIH: P01-HD31921 to Harris as PI of Add Health, R03 HD042490-02 to Guang Guo, R03 HD053385-01 to Guang Guo; and support from NSF, SES - 0210389 to Guang Guo. Special acknowledgment is due to Rick Bradley of the Housing Department and the Odum Institute at the University of North Carolina, Chapel Hill.

ABSTRACT

Drawing data from the College Roommate Study (ROOM; N=1,003 roommate pairs) and the National Longitudinal Study of Adolescent Health (Add Health; N>1,600), we investigated gene-environment interaction effects on youth binge drinking. Environmental influence took the form of peer mutual interplay between randomly assigned roommates on a college campus, which removes the possibility that gene-environment correlation biases our estimates of peer influence. On average, having a drinking as opposed to non-drinking peer increased binge drinking in college by 0.5-1.0 episodes per month. However, this peer influence was found only among youth with a medium genetic propensity for alcohol use. Youth with either a low or high propensity were not influenced by peer drinking. This GE interaction is replicated in data drawn from the Add Health study. Also replicated in the two studies is a main genetic effect. Out of the same five SNPs, all five in ROOM and four of the five in Add Health are significantly predictive of alcohol use in a single regression model.

We investigate whether peer influence on binge drinking (five or more drinks in a row for males or four or more drinks in a row for females) depends on self genetic propensity for alcohol use among youth in the United States. Peers are often believed to have a high degree of influence on youth risky behavior. Yet, peer influence remains difficult to investigate because observational data cannot separate peer influence from friend selection (see the review byKandel 1978; 1993; Moffitt 2001). Our study avoids confounds of peer selection by taking advantage of data gathered in the College Roommate Study (ROOM)(Guo, Hardie, Daw et al. 2009) from roommates who were randomly assigned to one another.

Our study tests the gene-environment interaction hypothesis that youth with a medium level of genetic propensity for alcohol use are more vulnerable to influences of drinking peers than those with a low or high level of genetic propensity. Those with low genetic propensities may not be attracted by either alcohol or risk taking associated with binge drinking, even when paired with a drinking peer. Those with very high propensities are likely to engage in binge drinking with or without a drinking peer.

Our experimental study design removes the threat of gene-environment correlation with respect to peer influence. A gene-environment correlation refers to the case when an apparent environmental influence is partially genetic (see the review by Jaffee and Price 2007). Our focus on peers as a single environmental influence alleviates the burden of multiple testing in a gene-environment interaction analysis. We develop a methodological strategy for gene-environment interaction analysis when a large number of genetic variables are involved. And we replicate the gene-environment (GE) interaction findings from ROOM using nonexperimental data on middle- and high-school friends from the National Longitudinal Study of Adolescent Health (Add Health) (Harris, Florey, Tabor et al. 2003).

BACKGROUND

Peer Influence. Several social theories provide frameworks for the interpretation of youth peer influences. Differential association theory highlights the processes by which

individuals are exposed to conflicting norms for certain behaviors (Christakis and Fowler 2007; Matsueda 1982; Sutherland 1947). The theory implies that excessive drinking among youth could be learned through intimate social interactions with drinking peers who act as if excessive drinking is acceptable or normal. Informal social-control theory calls special attention to social bonds between a youth and his or her society, suggesting that an individual is more likely to engage in delinquency when the bonds are weak or broken (Gottfredson and Hirschi 1990; Hirschi 1969; Kornhauser 1978; Sampson and Laub 1993). Adolescence and early adulthood is characterized by markedly weakened social bonds as youth become increasingly independent from parental control. Weakened social bonds leave youth vulnerable to peer influences in this life stage. Peer influences can also originate from the desire to belong to friendship networks. Friendship networks could serve as a source of support and aid the transition to independence from parents by providing role models and social opportunities (Borsari and Carey 2001). Friendships with drinking peers may increase participation in activities that involve alcohol consumption.

The well-known difficulty with observational studies of peer influence is peer self-selection (Manski 1993; Moffitt 2001). Evidence on binge drinking among friends does not establish how much of the similar behavior is caused by friend or peer influence and how much is caused by the possibility that individuals who engage in similar behavior are likely to become friends. In other words, we are unable to determine whether the similarity among friends is because "one takes on the color of one's company" or because "birds of a feather flock together." An unknown portion of the similarity between friends may be due to selection rather than influence. Ignoring selection may overestimate peer influence.

Random assignment of college roommates provides a major opportunity for research on peer influence because peers in this context are assigned rather than selected. Colleges' roommate randomization are often conditioned on housing preferences (e.g., location, number of roommates) expressed by students when they apply for student housing. The conditional

random assignment provided by housing lottery ensures that roommates are no more correlated than by chance in terms of their pre-college drinking behaviors and other characteristics. Data from randomly-assigned roommates in college have been used in studies of academic achievement (Foster 2006; Kremer and Levy 2008; Sacerdote 2001; Zimmerman 2003), fraternity membership (Sacerdote 2001) and drinking (Duncan, Boisjoly, Kremer et al. 2005). None of these roommate studies has considered self genetic propensities.

Genetic Propensity for Alcohol Use. Twin and adoption studies have demonstrated an important genetic base for alcohol-related disorders or alcoholism with large twin studies showing that more than a half of the variation in alcoholism is due to genetic factors (Goldman and Bergen 1998). Advances in genomic studies in recent years have led to an improved understanding of the molecular genetic origin of addictive behavior, including alcoholism. More than 100 mouse gene knockouts and transgenic studies showed that numerous genes were related to addiction-related behaviors (Goldman and Bergen 1998). These animal studies suggest diverse genetic pathways leading to addiction, which in turn suggests that genetic studies of addiction in humans must deal with the possibility of genetic heterogeneity (different cases of alcoholism may be related to different genetic loci) and polygenicity (multiple genes may contribute to an individuals' alcoholism).

Alcoholism is a complex behavior and is likely subject to the influences of a large number of genes in cellular molecular networks. The ADH1B gene and the ALDH2 gene are two exceptionally well-understood alcoholism-related genes encoding for two enzymes catalyzing consecutive steps in alcohol degradation (Crabb, Matsumoto, Chang et al. 2004; Greenfield and Pietruszko 1977). Significant variation in these two genes is confined mostly to East Asian populations. More generally, although many candidate genes have been implicated in alcoholism, understanding of their collective impact is limited. Data from the ROOM and Add Health studies provides us with the opportunity to investigate 101 SNPs in 21 genes that have been implicated in studies of risky behaviors.

Gene-Environment (GE) Interaction refers that the effect of an environmental factor on an individuals' phenotype depends on his or her genotype and vice versa. Ignoring GE interactions forces one to estimate an average environmental effect (averaged over all genotypes) or an average genetic effect (averaged over all environments). Such an analysis may thus underestimate environmental, genetic, or both effects or miss them entirely.

Our analysis examines on whether and the extent to which a peer effect on binge drinking depends on one's genotype. Although individuals with a genetic predisposition for alcohol use could be more easily swayed by peer drinking, this gene-environment interaction may not be linear. Youth with a higher genetic propensity may not be subject to a proportional stronger peer influence. Individuals with a particularly low or high propensity may be less vulnerable to influences from peer drinking for different reasons. Those with a particularly low propensity are likely unswayed because they are inherently unattracted by alcohol or risk taking. Those with a particularly high propensity may not be swayed because their attraction to binge drinking is too strong to be influenced by peers. We conduct empirical tests of this non-linear GE interaction hypothesis. A genetic propensity score ranging from 0 to 1 was estimated for each individual and each individual was assigned to a low (0-0.2 or 0.3), medium, or high (0.8-1.0) genetic propensity group. We hypothesized that individuals with medium propensities are most vulnerable to peer influences.

Analytical Strategy for Gene-Environment Interaction. Although GE interactions are widely considered to be a crucial part of the story that links genetic inheritance and complex human traits, GE interaction analyses involving a large number of genetic variables are rare owing to difficulties of handling false positives. When one or a few genetic variables are available, the estimation of GE interaction effects is straightforward using statistical procedures. But as the genomic 'revolution' had generated information on thousands of SNPs for each individual, it becomes much less clear what analytical strategy is appropriate. Genome-wide association studies (GWAS) estimate main genetic effects by regressing one SNP on a human

phenotype at a time and evaluate the findings by two straightforward criteria: a p-value of $5x10^{-7}$, which is equivalent to Bonferroni correction, and replication in one or more independent datasets (e.g., Frayling, Timpson, Weedon et al. 2007). This GWAS strategy is unlikely to be fruitful for GE interaction analysis when a large number of genetic and environmental variables are involved. A GWAS strategy for GE interaction analysis would require p-values substantially smaller than $5x10^{-7}$.

Although the GWAS strategy relies on the small p-value to identify genetic main effects, the subsequent replication study only needs a conventional p-value of 0.05, the rationale being that the replication study only targets genetic variables identified in the GWAS. Thus, the GWAS-identified genetic main effects may not be highly significant as suggested by the required small p-values. The highly selective strategy of GWAS may select on factors unrelated to the relationship between the genes and the phenotype. This suggests that additional genetic variants that satisfy less stringent p-values may be just as valid predictors of the phenotype as those GWAS-identified variants and that these additional genetic variants may be identified by (1) more moderate p-values than $5x10^{-7}$ and (2) replication. This logic has motivated the development of the analytical strategy for our GE interaction analysis, which search for consistent results in independent studies.

FINDINGS

The FDR screening of the 186 SNPs in 28 genes yielded 73 SNPs in 21 genes that significantly predicted alcohol use at the level of 10% in ROOM (SOM Table S1). More than 30% of tested SNPs (73/186=39%) survived the screening because most of the 28 tested genes were selected for their implication in risky behavior. In the ROOM stepwise regression analysis, 10 of the 73 SNPs remained simultaneously significant at the level of 0.05. The 10 SNPs were from six genes: DRD2, MAOA, LMO3, TPH2, DBH, and DRD4 (SOM Table S2). Five of the 10 SNPs were also genotyped in Add Health; these five were RS4245145 (DRD2), RS2242592 (DRD2),

RS1125394 (DRD2), RS3027405 (MAOA), and RS7975434 (LMO3). In the replication analysis based on Add Health, only these five were used in a logistic regression estimating genetic propensity for drinking. In Add Health, four out of the five SNPs were simultaneously significant when the five were included in a single regression model. These four SNPs' coefficients had the same sign and approximate size as the same four SNPs in ROOM. We estimated the genetic propensity models of ROOM using the same set of five SNPs that were available in Add Health. All the five main effects from the 5-SNP ROOM analysis were very similar to the same 5 main effects in the 10-SNP ROOM analysis (Table 1). A second genetic propensity for drinking based on the same procedure was constructed using the much larger SNP set from Add Health and this propensity score was constructed from 27 SNPs that were simultaneously significant at 0.05 level (Appendix I).

Tables 3 and 4 about here

Table 5 presents initial evidence on peer influence interactions from ROOM based on three levels of genetic propensity. The genetic propensity was based the five SNPs identified in the ROOM Stepwise regression. All peer effects were estimated after adjusting for a full set of controls including bio-ancestry scores. Only college students with genetic propensity scores in the middle range appeared to increase their binge drinking in response to roommate assignment. Pairing these students with a roommate who drank in high school increased binge drinking episodes per month, respectively, in the first semester in college, the past semester and over the past two weeks by 0.95, 0.73, and, 0.88. These amounted to x%, y% and z% relative to the overall average amount of binge drinking reported in the sample. There was no evidence that individuals with either low or high genetic propensity were affected by peer influence.

Table 5 about here

Peer influences in the Add Health data can be estimated by relating Wave III reports of binge drinking of a representative national sample of 18-26 year olds to patterns of binge drinking among the individuals reported in Wave I to be best friends seven years before when

they were in grades 7-12. Regressions in Table 6 repeat the Table 5 analysis using these Add Health and show that key results from the ROOM data are replicated. When genetic propensity for drinking was estimated from the five SNPs, having a drinking friend at Wave I increased 1.1 binge drinking episodes per month over the past two weeks and .055 binge drinking days per month over the past year at Wave III. The results from the 27-SNP genetic propensity were similar to the five-SNP results: nominating a drinking friend at Wave I was associated with 0.76 more binge drinking episodes per month over the past two weeks and 0.72 more binge drinking days per month over the past year at Wave III. All of these increases happened to youth with medium genetic propensity for drinking. Youth in the low or high propensity groups were not influenced by friend drinking.

Table 6 about here

Table 7 shows the GE interaction findings estimated in a single regression model (Equation [2]), in which the low and high propensity individuals were combined into one category to reduce the number of interaction terms. The ROOM analysis revealed that when paired with a roommate with a drinking history in high school, college students with a medium propensity reported 0.76 (p=.036), 0.59 (p=.10), and 0.82 (p=.019) more binge drinking episodes per month for the first semester, the past semester, and the last two weeks, respectively than those whose roommates did not drink in high school. Separate regression models testing the peer influence among individuals with a low or high propensity showed three much smaller coefficients of 0.17, 0.08, and -0.29 and much bigger p-values of 0.69, 0.84 and 0.48, respectively.

Table 7 about here

The findings in ROOM described in Table 7 have been replicated in Add Health (Table 8). For those with a medium genetic propensity, having a drinking friend at Wave I was associated with 1.02 (p=0.0007) and 0.65 (p=0.038) more binge drinking episodes per month for the past two weeks for the 5-SNP propensity and the 27-SNP propensity, respectively than those who did

not report a drinking friend. For those with a low or high propensity, the two estimated peer effects for the 5-SNP propensity and the 27-SNP propensity were much smaller (.22 with p=.42 and .48 with p=.08, respectively). The findings for the second binge drinking measure of days over past year in Add Health were similar, with positive and statistically significant peer effects (.54 with p=.055 and .62 with p=.03) only found among those with a medium genetic propensity. The peer effects for those with a low or high propensity were small and non-significant (.43 with p=.12 and .33 with p=.22). The findings in Tables 7 and 8 are summarized in Figure 1.

Table 8 about here

CONCLUSION AND DISCUSSION

Drawing data from ROOM and Add Health, we investigated the gene-environment interaction effects on youth binge drinking – the interaction between peer influences and self genetic propensity for alcohol use. The gene-environment interaction between genetic propensity for drinking and peer influence was performed in both ROOM and Add Health for replication. Within each data source, the initial GE interaction was obtained by estimating peer influence first at three levels of low, medium and high genetic propensity in three regression models and then in a single regression model that combines those in the low- and high-propensity groups. On average, having a drinking peer increased binge drinking by 0.5-1.0 episode per month as compared to having a non-drinking peer. However, this peer influence was found only among youth with a medium genetic propensity for alcohol use. Youth with low or high propensity were not influenced by peer drinking

This GE interaction finding for youth binge drinking was replicated across the ROOM study and the Add Health study. Within each study, the same GE interaction finding was replicated across more than one binge drinking measure, and across the procedure that estimated peer effect at three levels of propensity separately in three regression models and the procedure that estimated the GE interaction in a single regression model. Within Add Health,

the GE interaction finding was replicated between an analysis based the 5-SNP propensity and an analysis based on the 27-SNP propensity.

Three out of the four SNPs that are replicated between ROOM and Add Health are in the DRD2 gene and the fourth SNP is in the LMO3 gene. The three SNPs in DRD2 are independently associated with alcohol use because the three effects are adjusted for one another in a single regression model. The dopamine D2 receptor (DRD2) gene located on chromosome 11 q22-q23 encodes the dopamine D2 receptor. Because of its key role in the dopaminergic system, DRD2 is a prime suspect in investigations of genetic links with risky behaviors including alcoholism. The DRD2 antagonist haloperidol has long been used to treat aggressive behavior in psychotic patients. Animal models implicated DRD2 in ethanol preference (Crabbe, Phillips, Buck et al. 1999). The well-known polymorphism Taq1A1 (rs1800497) has been studied intensively for its potential link to alcoholism (for a review see Dick, Wang, Plunkett et al. 2007). For a number of years, this genetic variant was thought located in DRD2; but it turns out to be sited in the ANKK1 gene (Neville, Johnstone, and Walton 2004). Unlike this Taq1A1 variant, the three SNPs in our replicated SNP set are in DRD2. Using data from COGA, Dick et al. (2007) showed that alcohol dependence is associated more with ANNKK1 than with DRD2. However, major differences exist between COGA and our data from ROOM and Add Health. While COGA subjects are patients of alcohol dependence, our subjects were are youth who use alcohol to various degrees.

LMO3 belongs to the LIM-only protein family with a function to modulate transcription by using its two tandem LIM domains to bind to DNA-binding proteins (Kadrmas and Beckerle 2004). The relationship between LMO3 and alcoholism has been studied for many years in animals. The fruit fly Drosophila melanogaster has been used to identify novel genes that affect behavioral responses to ethanol. Several studies found that reduced dLMO expression led to increased sensitivity to the sedating effect of ethanol and decreased level of ethanol

consumption, whereas the increased dLMO expression had the opposite effect both in flies and mice, suggesting that LMO3 may play an important role in alcohol preference in invertebrate systems and in mammals (Lasek, Giorgetti, Berger et al. 2011; Tsai, Bainton, Blau et al. 2004). It is speculated that LMO3 may affect behavioral responses to ethanol in humans through its ability to regulate transcription which, in turn, can affect the patterning of certain brain structures such as the cortex or amygdala (Bulchand, Subramanlan, and Tole 2003; Remedios, Subramanian, and Tole 2004). The subtle changes in brain structures may later affect behavior responses to ethanol (Lasek et al. 2011).

The peer effect reported in this analysis represents only a subset of total peer effects. For example, in the ROOM study, potentially important peer influences beyond roommates could come from a girlfriend or boyfriend, friends who do not live in the same dorm, and on-campus student organizations including sororities and fraternities and sports teams. Similarly, this analysis measures only a subset of genetic variants that are related to propensity for alcohol use.

When a GE interaction is based on an observational study, the threat of gene-environment correlation can hardly be eliminated. GE correlation results in an environment factor that is partially genetic, causing difficulties in interpreting GE interaction. Randomized experiments with human subjects protect against GE correlation, but these studies are often financially prohibitive and ethically complex. For these reasons, animal models instead of human models are often used to address the issue (Barr, Newman, Lindell et al. 2004; Barr, Newman, Shannon et al. 2004). The randomly assigned roommates in ROOM amount to a rare solution to the thorny issue of gene-environment correlation (Jaffee and Price 2007). The randomization guarantees that the roommates' pre-college behavior is uncorrelated and that the peer influences are exogenous and uncorrelated with self genetic propensity for alcohol use.

Although peer influences from ROOM and Add Health were based on two very different study designs – ROOM from a randomized experiment and Add Health from a traditional

observational study – the estimated peer effects conditional on genetic propensity from ROOM and Add Health are similar. The research community has generally assumed that peer effects estimated from observational studies overstate peer causation because of peer self-selection. That our findings from self-selected school friend are similar to those obtained from randomly assigned roommates suggest that the biases may be modest, at least in the case of binge drinking. Our evidence on this issue is preliminary. Subsequent efforts should focus on designing observational studies that are as comparable as possible to experimental studies so that the two sets of findings can be compared with more confidence.

In this GE interaction analysis, we focused on the effects of peers. Our question is: at what level of self genetic propensity do peers exert more prominent influences on binge drinking? Another study focusing on the effects of genes may ask a different question of whether the effects of genes on alcohol use are larger when individuals have drinking friends. This latter case illustrates the point that in spite of rapid technological advances in molecular genetics, the understanding of the genetic origins of complex human traits may often require an adequate understanding of environmental circumstances under which the relevant genes are operating.

MEASURES AND METHODS

Data Sources consist of a discovery dataset, the College Roommate Study or ROOM and a replication dataset, the National Longitudinal Study of Adolescent Health or Add Health. In the case of ROOM, subjects were freshmen, sophomores, and juniors in a large US public university in the spring semester of 2008. All had been randomly assigned roommates when they first entered the university. Students who requested a specific roommate or who participated in a themed housing program (e.g., foreign languages, health sciences, substance free, etc.) were excluded. In randomly assigning roommates, the university housing office placed data from applications into a large database, which was loaded into the software program RMS for random matching. Every student was then randomly assigned a unique RMS-ID number. After the first student had been placed in a room, the RMS program assigned his or her

roommate as the next student in the chronological RMS-ID order who had compatible gender, smoking status, and type of requested room. In the procedure, roommates were essentially randomly assigned to each other within each gender/smoking/room type cell.

The information on alcohol use and socio-economic background in ROOM was obtained via a web survey, which was completed by 2,664 (79.5%) of the eligible students. Of those who completed the web survey, 2,080 (78.7%) provided a saliva sample. Students who did not live on campus, who were too young (under 18) to be included in the alcohol study, and who were in a study-abroad program in a foreign country for the semester were considered ineligible. Our final analysis sample included 1,003 pairs of randomly assigned roommates where both roommates participated in the study. Of the 1,003 pairs, 694 had genotype data for both roommates and 309 had genotype data for only one of the two roommates. These 309 pairs of roommates were included in the analysis since only self genotype data were necessary.

Data were also drawn from the Cooperative Institutional Research Program (CIRP). Each year, a large number of universities administer this Freshman Survey to entering students during orientation or registration. The survey gathers information from a range of student characteristics including a small number of health behaviors. This part of our analysis only includes individuals in CIRP who are also roommates in ROOM and who have explicitly consented to our using their CIRP responses. These two independent studies targeted to recruit the same student body.

To verify random assignment of roommates, we calculated within-dorm Gamma correlation coefficients (Goodman and Kruskal 1954) for fourteen pre-college responses obtained by CIRP, which was designed and carried out independently from the ROOM. The individuals in the calculations based on CIRP were also a subset of the roommates from ROOM. None of the CIRP responses between roommates is correlated at $p<.05$ or less (Table 1). We also calculated within-dorm Gamma correlation for two measures on alcohol use in high school obtained from ROOM. Neither drinking nor binge drinking shows a within-dorm correlation.

The findings from CIRP are consistent with random assignment of roommates. The findings from ROOM indicate that the pre-college drinking measures (which are used as peer influence in GE interaction analysis) are, indeed, uncorrelated.

Table 1 about here

Add Health is longitudinal and started as a school-based study of the health-related behaviors of adolescents in grades 7-12 in 1994-5 in the United States (Harris et al. 2003). Eighty high schools were randomly selected from a stratified nationally representative sample of all public and private high schools in the United States. These strata were based on region of the country, urbanicity, school type (public, private, and parochial), and racial composition. For each of the 80 high school, the largest feeder school (typically a middle school or junior high) was targeted for recruiting. The final sample consisted of 134 schools. In Wave I, a self-administered in-school questionnaire was given to all seventh- through 12th-graders attending these schools on a chosen day in 1994-5. About 90,000 students or 77% responded. The questionnaire included questions on students' risky behavior. Students were asked to nominate up to five male and five female friends from rosters of the high school and the feeder school. From this friend information, students' friends' behavior became known from friends' self report. In addition to the in-school survey, an in-home interview was conducted in 1994-5, 1995-6 and 2002 (Waves I-III). The binge drinking outcomes in Add Health in this analysis were from the in-home interview in Wave III. Our Add Health analysis data consist of 2,270 individuals whose saliva was gathered at Add Health Wave III in 2002 and who have valid genotype and survey data. These 2,281 individuals represent 87% of 2,612 individuals whose saliva DNA were collected at Wave III.

Measures of Genotype. In ROOM, DNA was extracted according to the manufacturer's instructions from 2mls of saliva, containing buccal epithelial and white blood cells, collected from participants in an Oragene DNA collection kit. Our median DNA yield was 27.3 ug, with a minimum of zero ug for six individuals and a maximum of 71.3 ug. DNA was

plated for Illumina genotyping at 30 ul at >50 ng/ul. For ROOM, we designed an Illumina GoldenGate assay for 384 candidate SNPs, including a set of 186 ancestral informative markers (Enoch, Shen, Xu et al. 2006). Besides the 162 AIMs, which were successfully genotyped out of the 186 targeted, another 186 SNPs in 28 genes were successfully genotyped and these SNPs were selected mostly because of their implications in risky behavior.

In Add Health, genomic DNA gathered at Wave III in 2002 was isolated from buccal cells with an average yield of DNA of 58±1 μg. The genotype data used in this analysis were based on an Illumina GoldenGate assay for 1,536 candidate SNPs including the same 186 AIMs (Enoch et al. 2006) targeted in ROOM. Excluding the successfully genotyped 121 of the 186 targeted AIMs, a total of 1019 SNPs in 130 genes became available for analysis. The GoldenGate array of 1,536 was designed to include primarily genetic variants related to risky behavior such as aggression, alcohol use, smoking and illegal drug use. These variants include those in the 55 genes assembled by Steve Maxson and colleagues to keep track of the genes that have been shown to have an effect on aggression in mice studies (Maxson and Canastar 2003; Maxson In press). Between ROOM and Add Health, 101 SNPs are common after excluding AIMs. In both ROOM and Add Health, the bio-ancestry scores of Africans, Europeans and East Asians were estimated using the Structure procedure (Pritchard, Stephens, and Donnelly 2000); the three scores for each individual summed to one.

Measures of Binge Drinking and Controls. In the ROOM GE interaction analysis, three binge drinking measures were used as the outcome variables: a monthly count of binge drinking episodes (1) during the fall semester of the first year of college, (2) during the past fall semester, and (3) during the past two weeks. The responses of "never, less than once a month, once or twice a month, about once a week, 2-4 times a week, and every day or almost every day" were coded as 0, 0.5, 1.5, 4.3, 12.9, and 25, respectively. The two binge drinking measures used as outcome variables in the Add Health GE interaction analysis were from Wave-III: monthly

binge drinking episodes in the past two weeks and monthly count of binge drinking days over the past year. These Add Health outcomes were coded in the same fashion as those in ROOM.

Table 2 about here

In ROOM, "roommate drank" was coded as one for individuals who used alcohol in high school and zero otherwise. Pre-college drinking rather than college drinking was used to measure peer drinking in order to avoid the simultaneity of the roommates' drinking in college. Peer drinking measures were coded as an indicator or 0-1 variable to simplify GE interaction analysis. In Add Health, peer drinking was based on primarily reported drinking by nominated friends themselves at Wave I in 1994 rather than on reported drinking by egos where egos were individuals whose binge drinking was predicted in regression models. If peer drinking from nominated friends was missing, ego-reported friend drinking at wave I was used instead. To make peer drinking data in Add Health more similar to those in ROOM in age, we excluded those who were younger than 15 years old at Wave I when they nominated their friends. "Friend drank" at Wave I was measured by an indicator variable coded as 1 for those who drank beer, wine, or liquor or who got drunk over the past 12 months, and zero otherwise.

In the ROOM analysis, the response variable in the false-discovery-rate (FDR) procedure and the stepwise regression was constructed from the mean of three drinking measures: the frequency of alcohol beverage use (1) during the fall semester of the first year of college; (2) during the past fall semester, and (3) over the past two weeks. The response variable in the model of genetic propensity in Add Health was constructed from the mean of two binge drinking measures at Wave III: monthly binge drinking episodes in the past two weeks and monthly count of binge drinking days over the past year.

The GE interaction analysis for ROOM controlled for gender, father's education, mother's education, roommate's father's education, roommate's mother's education, family income, roommate's family income, weekly church attendance or religiosity, roommate's religiosity, total SAT score, roommate's total SAT score, GPA, roommate's GPA, having

nonwhite roommate and fixed effects constructed from the housing preferences on the housing form. The GE interaction models for Add Health controlled for gender, age at Wave III, PVT test score, parental education, family income, parental unemployment status, presence of two biological parents, household size, and religiosity. Both ROOM and Add Health controlled for bio-ancestry scores.

Analytical Strategy. We developed an approach for gene-environment interaction analysis that involves a large number of genetic variables. Our overall strategy consisted of two stages. Stage 1 yielded a genetic propensity score ranging 0 to 1. In Stage 2, the score was used in a GE interaction analysis, in which the propensity score was interacted with the pre-college drinking behavior of the roommate to predict self binge drinking. The data for Stages 1 and 2 are only from ROOM, the discovery dataset. The Add Health skipped the FDR procedure and the stepwise regression and the data from Add Health were then used to replicate both the main genetic effects and GE interaction effects from ROOM.

Stage 1 used a procedure of false discovery rate (FDR) that selected SNPs out of the 186 SNPs available in ROOM. The selected SNPs were individually predictive of binge drinking at the significance level of 0.10. Then, all FDR-selected SNPs were simultaneously entered in a stepwise regression with the same binary drinking outcome as in the FDR procedure. Only genetic variants with a p-value of 0.05 or smaller in the step-wise regression were retained and used to calculate the genetic propensity score for each individual, which is the predicted probability based on the final step-wise logistic regression.

Selection criteria were much more stringent in the step-wise regression than in the FDR procedure. While the FDR procedure estimated the effect of each genetic variant independently of the other genetic variants, the stepwise regression only chose a set of genetic variants that remained statistically significant after removing all redundant genetic influences. For this reason, the number of genetic variants that survived the step-wise regression tended to be much smaller than those that survived the FDR procedure.

In Stage 2, the gene-environment interaction analysis consisted of two sets of regression. The first set of regression compared the effect of peer drinking across groups with low, medium and high levels of genetic propensity for alcohol use in three separate regression models. Then the GE interaction was estimated in a single regression.

Equation (1) describes the GE interaction analysis that estimated peer effect separately in the low, medium, and high propensity groups:

$$(1)\quad \text{selfbingedrinking}_{ij} = \beta_0 + \beta_1 \text{peerdrank}_i + \text{Controls}_i \boldsymbol{\beta}_2 + v_j + e_{ij},$$

where "peerdrank" represents pairing with a roommate who drank in high school and v_j are fixed effects of the cells. Equation (1) explores the genetic propensity by peer interaction or how self genetic propensity for alcohol use conditions peer influences. Equation (2) describes the GE interaction analysis in a single regression model:

$$(2)\quad \begin{aligned} &\text{selfbingedrinking}_{ij} = \beta_0 + \beta_1 (\text{selfmedium/peerdrank})_i + \\ &\beta_2 (\text{selflowhigh/peernondrank})_i + \beta_3 (\text{selflowhigh/peerdrank})_i + \text{Controls}_i \boldsymbol{\beta}_4 + v_j + e_{ij}, \end{aligned}$$

where "peernondrank" represents pairing with a roommate who did not drink in high school; "selfmedium" and "selflowhigh" stand for "self in the medium genetic propensity group" and "self in the low or high genetic propensity group," respectively; the combination of "selfmedium/peernondrank" was the omitted reference group; subscripts *i* and *j* are indexed for individuals and housing preference cells, respectively; and v_j are fixed effects of the cells.

The GE interaction model (2) includes dummy variables for three of the four combinations of self genetic propensity and roommate's precollege drinking behavior. Drawing on the exploratory GE interaction findings from Equation (1), individuals in the low and high propensity groups were combined into a single category in Equation (2). The central hypothesis tests β_1 or whether a college student with a medium genetic propensity binge-drank more when paired with a roommate who drank in high school than those also with a medium genetic propensity but paired with a roommate who did not drink in high school.

The coefficient β_3 for "selflowhigh/peerdrank" in Equation (2) can be estimated after omitting the combination of "selflowhigh/peernondrank." This model tests the hypothesis that a college student with a low or high genetic propensity would not increase binge drinking when paired with a roommate who drank in high school. Evidence for the test lends additional support for our GE interaction thesis.

The ROOM dataset consists of pairs of dorm roommates. Either member of a pair can be used to construct the response variable in a regression analysis. To avoid arbitrariness, we performed 500 analyses with each analysis randomly selecting one of the two members in a roommate pair to construct the response variable. This randomizing procedure was applied to 694 of the 1,003 roommate pairs where both roommates have DNA measures. The final regression coefficients and t statistics were averages over the coefficients and t statistics from the 500 analyses.

In the analysis of the ROOM data, to further ensure that estimated peer influences are based solely on variation induced by random assignment, we added fixed-effect controls for preference "cells" in our regression analysis. In the analysis of the Add Health data, the correlation within sibling clusters was addressed by a generalized estimation equation model (GEE) (Liang, Zeger, and Qaqish 1992). The missing values of non-genetic variables were imputed by the multiple imputation technique (Rubin 1987). The multiple completed datasets were then analyzed separately by SAS before the results were combined to produce the overall inference. Missing values in genotype data were imputed via Math (Li, Willer, Sanna et al. 2009; Marchini and Howie 2010). To address population stratification, all stepwise regressions and GE interaction models controlled for ancestry scores of Africa and Europe (Pritchard et al. 2000).

Figure 1. The gene-environment interaction effects on binge drinking– interaction between peer effects and self genetic propensity for alcohol use, each bar representing a peer effect from a separate regression model. Error bars represent ±1 SEM. All models are adjusted for bio-ancestry and SES controls. #:*P*<0.1, *:*P*<0.05 and ***:*P*<0.001. (A) the ROOM Study where peers were randomly assigned roommates and the genetic propensity was based on the 5 replicated SNPs (N=1,003 roommate pairs). (B) the Add Health Study where peers were ego-nominated friends and the genetic propensity was based on the five replicated SNPs (N=1,612). (C) the Add Health Study where peers were ego-nominated friends and the genetic propensity was based on the 27 SNPs (N=1,604).

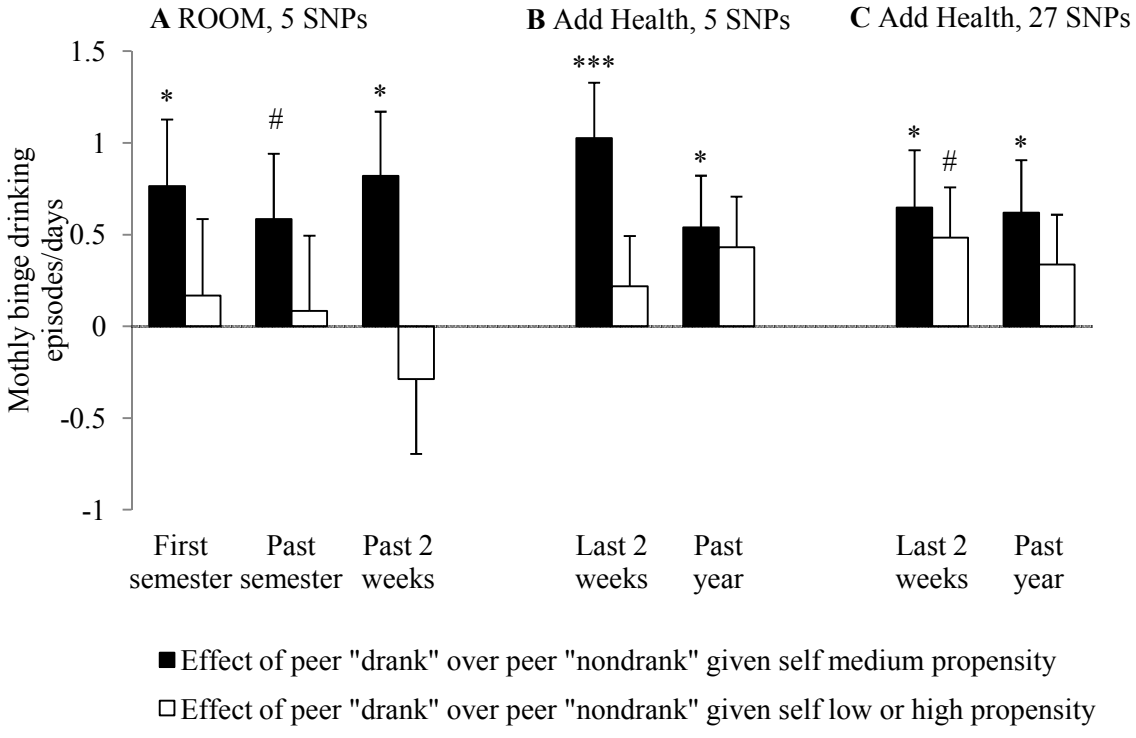

Table 1. Within-dorm correlation Gama correlation coefficient for 15 responses in CIRP and for 2 high school measures on alcohol use for checking random assignment of roommates in ROOM

Variables	Intraclass correlation coefficient	P >|t|	df	Model
Pre-college behaviors from CIRP :within-dorm correlation based on pre-college CIRP responses; these CIRP subjects also roommates in ROOM				
Had drank beer	0.088	0.322	894	Clog
Had drank wine/liquor	0.020	0.800	892	Clog
Smoked cigarettes	0.000	.	898	Clog
Physical exercise	0.087	0.185	878	Lin
Partying	0.000	.	878	Lin
Religious Service Attendance	0.000	1.000	897	Clog
Felt Depressed	0.090	0.301	897	Clog
Frequency of Volunteering	0.000	1.000	895	Clog
Freq of Community service	0.013	0.859	891	Clog
Political view	0.044	0.525	870	Clog
Hours Socializing with Friends	0.000	.	879	Lin
Hours Volunteering	0.025	0.746	871	Lin
Hours Watching TV	0.022	0.724	876	Lin
Hours Reading	0.039	0.580	877	Lin
Hours Playing Video Games	0.000	.	877	Lin
Pre-college alcohol use from ROOM: within-dorm correlation based on ROOM high school responses; these subjects are in ROOM				
High school drinking	0.010	0.866	1,019	Clog
High school binge drinking	0.000	.	1,042	Bin

Table 1. Within-dorm correlation Gama correlation coefficient for 14 responses in CIRP and for 2 high-school measures on alcohol use for checking random assignment of roommates in ROOM

Variables	Gamma correlation coefficient	P of χ^2	N
Pre-college behaviors from CIRP:within-dorm correlation based on pre-college CIRP responses; these CIRP subjects also roommates in ROOM			
Had drank	0.16	0.09	434
Smoked cigarettes	-0.41	0.23	438
Physical exercise	0.31	0.41	418
Partying	-0.05	0.63	416
Religious Service Attendance	0.00	0.49	440
Frequency of Depression	0.07	0.45	436
Frequency of Volunteering	-0.02	0.54	434
Freq of Community service	0.02	0.57	438
Political view	0.09	0.95	420
Hours Socializing with Friends	-0.10	0.49	422
Hours Volunteering	0.08	0.58	414
Hours Watching TV	-0.01	0.99	416
Hours Reading	0.07	0.77	410
Hours Playing Video Games	0.25	0.20	418
Pre-college alcohol use from ROOM: within-dorm correlation based on ROOM high school responses; these subjects are in ROOM			
High school drinking	0.047	0.90	1,690
High school binge drinking	0.056	0.12	1,685

Table 2. Description of the variables in ROOM and Add Health

	ROOM			Add Health		
Variables	Mean	SD	N	Mean	SD	N
Outcomes and Measures of Peer Influences						
Binge drinking first semester monthly episodes	2.28	4.01	1696			
Binge drinking past semester monthly episodes	2.25	3.94	1697			
Binge drinking past 2 weeks monthly episodes	2.33	3.91	1696			
Roommate drank in high school (0 or 1)	0.39		1693			
Binge drinking past 2 weeks monthly episodes				2.03	4.18	1621
Bing drinking days past year monthly count				1.56	3.92	1613
Friend drinking at Wave I (0 or 1)				0.66		1635
Controls						
European ancestry score	.78	.36	1702	.69	.40	1644
African ancestry score	.15	.32	1702	.17	.34	1644
Male	.39		1703	.48		1644
Age at Wave III				22.7	1.22	1644
Parent unemployed				.045		1409
Father's education	16.06	2.17	1660			
Roommate's father's education	16.04	2.15	1657			
Mother's education	15.91	1.98	1695			
Roommate's mother's education	15.88	2.0	1690			
Parent education, higher of the two parents						1566
No high school				0.126		197
High school				0.284		445
More than high school				0.590		924
Family income in $10,000	13.81	11.48	1635			
Roommate's family income in $10,000	13.84	11.65	1624			
Family income						1275
0-20k				0.180		229
20k-60k				0.550		701
>60k				0.271		345
2 bio-parent presence				0.611		1644
Household size						1644
<=2				0.013		22
2-7				0.841		1382
>7				0.146		240
Church attendance weekly	.193		1685	.408		1644
Church attendance weekly roommate	.206		1683			
SAT	1322	157	1457			
Roommate's SAT score	1328	147	1443			
GPA	4.21	.48	1671			
Roommate's GPA	4.21	.48	1664			
Having none-white roommate	.33		1703			
PVT test score						1581
<90				0.238		377
90-110				0.471		744
110-150				0.291		460

SOM Table S1. Genes, their chromosome, number of SNPs in each gene tested, and SNPs selected by the FDR procedure in the discovery dataset ROOM

Gene and chromosome	Number of SNPs tested	SNPs selected by FDR
ADH1A, 4	2	Rs182609 rs4147531 (2)
ADH1B, 4	10	Rs1159918 rs1229982 rs7673353 (3)
ALDH2, 12	11	rs10849970 rs2158029 rs2238151 rs671 rs7296651 rs7311852 (6)
ANKK1, 11	1	0
ARVCF, 22	1	Rs5993891 (1)
BDNFOS, 11	1	0
CHRM2, 7	5	rs1455858 rs7357341 (2)
CHRNA4, 20	1	Rs2236196 (1)
CHRNB2, 1	1	0
CNR1, 6	1	0
COMT, 22	9	rs174696 rs739368 (2)
DBH, 9	7	rs1541332 rs3025410 rs77905 (3)
DDC, 7	3	rs1451371 rs1470750 rs998850 (3)
DEAF1, 11	1	0
DRD2, 11	26	rs1076563 rs1079596 rs11214605 rs1125394 rs12283680 rs12364283 rs2242592 rs2471857 rs2587548 rs2734833 rs4245145 rs4581480 rs7109897 (13)
DRD4, 11	5	rs11604855 rs1800443 rs3758653 rs916457 (4)
FTO, 16	8	rs10521303 rs6499640 (2)
GABRA2, 4	15	rs16859292 rs16859325 rs16859348 rs6857343 rs7678520 (5)
HTR1B, 6	17	rs1213366 rs13212041 (2)
HTR2A, 13	6	Rs6304 (1)
LMO3, 12	13	rs11057005 rs16912030 rs16912043 rs7975434 (4)
MAOA, X	11	rs2072744 rs3027405 rs5905859 rs5906729 rs5906883 (5)
MAOB, X	8	rs1040399 rs12394221 rs17462 rs1799836 rs2239441 rs3027459 rs6520902 rs9887047 (8)
SLC18A2, 10	2	Rs363333 (1)
SLC6A4, 17	17	rs2054848 rs9903602 (2)
TPH21, 7	10	rs1386483 rs2171363 rs7967586 (3)
TTC12k, 11	1	0
TXNRD2, 22	3	0
Total = 28	186	73

SOM Table S2. Regression coefficients and p-values of the SNPs selected by step-wise regression (a single regression model) from the discovery dataset ROOM and the replication dataset Add Health

SNP	Gene	ROOM (10 SNPs)	Add Health (5 SNPs)	ROOM (5 SNPs)
rs4245145	DRD2	.348 (.023)	0.70(<.0001)	.416(.0048)
rs3027405	MAOA	.350 (.0029)	0.058(.54)	.369(.0013)
rs2242592	DRD2	.281(.0018)	0.20(.0023)	.340(<.0001)
rs7975434	LMO3	-.241(.013)	-0.32(<.0001)	-.269(.0048)
rs1125394	DRD2	.443 (<.0001)	0.24(.0012)	.470(<.0001)
RS7967586	TPH2	.841 (.012)		
RS12283680	DRD2	1.59 (.0043)		
RS1541332	DBH	.215 (.0059)		
RS3758653	DRD4	.238 (.014)		
RS12364283	DRD2	-.404 (.032)		
N		2060	2249	2060

1. Out of the 10 SNPs selected from the discovery dataset ROOM, only 5 were genotyped in the replication dataset Add Health; replication was attempted only on these 5 SNPs. These 5 SNPs were tested again in ROOM.
2. All FDR selected SNPs were entered into the stepwise regression.
3. The stepwise regression controlled for bio-ancestry scores to address population admixture.
4. The larger sample size in this ROOM analysis than that in the ROOM GE interaction analysis is because the latter analysis requires subjects in pairs of roommates. The larger sample size in this Add Health analysis than the Add Health GE interaction analysis is because the Add Health analysis excluded those who dominated friends at Wave III when they were younger than 15. Neither restriction is present in the estimation of main genetic effects.

Table 5. Full models of peer influence by genetic propensity interaction on binge drinking with peer effect estimated separately at low, medium and high genetic propensity (ranging 0 to 1) for alcohol use: ROOM

College binge drinking →	**First semester binge drinking**			**Past semester binge drinking**			**Past 2 week binge drinking**		
Genetic propensity	low (0-.2)	medium (.2-.8)	high (.8-1)	low (0-.2)	medium (.2-.8)	high (.8-1)	low (0-.2)	medium (.2-.8)	high (.8-1)
N of SNPs		5 SNPS			5 SNPs			5 SNPs	
Roommate drank in high school	.415 (.43)	**.952*** **(.021)**	-.296 (.771)	.271 (.63)	**.734#** **(.069)**	-.514 (.508)	-.540 (.399)	**.882*** **.022)**	-.629 (.42)
Respondent characteristics									
Female	-.052	-1.38**	-1.42#	-.233	-1.52***	-.956	-.936	-1.56***	-1.19
Father's education	-.109	.048	-.047	-.094	.058	-.126	-.070	0.056	-.157
Mother's education	.010	.030	.004	.003	-.004	-.061	.079	.034	.016
Family income $10,000	.015	.037*	.020	.038	.042*	.046	.009	.045*	.043
GPA	-.502	-.494	-.152	-.324	-.557	-.239	-.154	-.524	.216
SAT/100	.170	-.038	-.027	.194	-.061	-.007	.146	-.034	-.132
Nonwhite roommate	-.244	.107	-.552	-.127	.091	-1.09	.843	.317	-.776
Church attendance weekly	-1.21 *	-.1.64**	-1.36	-1.38*	-1.16**	-1.05	-.552	-1.44**	-1.44
Roommate characteristics									
Fathers' education	-.153	-.007	-.046	-.099	-.009	-.163	-.119	-.041	-.153
Mother's education	.012	-.061	-.125	.005	.069	-.033	-.006	.044	.129
Family income $10,000	-.012	.001	.074*	-.003	-.015	.063*	-.002	-.006	.046
GPA	-.135	.362	.366	-.370	.439	.567	-.428	.432	.510
SAT/100	.101	.010	.212	.043	-.097	.163	.186	.051	.306
Church attendance weekly	-.067	.353	-.026	-.076	.091	-.459	-.147	-.054	-.560
Bioancestry(African)	-.245	.542	-2.05	-.014	.485	1.32	-.365	.126	-.977
Bioancestry(European)	1.09	1.41#	1.65	1.16	1.48#	5.12	1.48	1.35#	1.85
N	202	674	127	202	674	127	202	674	127

1 Genetic propensity for binge drinking was measured by 5 SNPs from the results from a stepwise logistic regression (Table 4). The predicted genetic propensity score ranging 0 to 1 is used to divide the entire sample into three groups with 2 cutoff points at 0.2 and 0.8.

2 ***= p-value<.001; **= p-value<.01; *= p-value<.05; #= p-value<.10.

3 The boldface type is used only to highlight the effects of primary interest. p-values are provided only for these coefficients.

Table 6. Full models of peer influence by genetic propensity interaction on binge drinking with peer effect estimated separately at low, medium, and high genetic propensity (ranging 0 to 1) for alcohol use: Add Health

Wave III →	**Binge drinking past 2 weeks monthly episodes**						**Bring drinking past year monthly days**					
Genetic Propensity	low (0-.3)	low (0-.3)	med (.3-.8)	med (.3-.8)	high (.8-1)	high (.8-1)	low (0-.3)	Low (0-.3)	med (.3-.8)	med (.3-.8)	high (.8-1)	high (.8-1)
N of SNPs	5	27	5	27	5	27	5	27	5	27	5	27
Friend drank	-0.12	0.52	**1.1*** **	**0.76***	0.41	0.35	0.21	0.06	**0.55***	**.72***	0.71	0.65
(p value)	(.75)	(.074)	**(.0007)**	**(.017)**	(.30)	(.53)	(.55)	(.84)	**(.059)**	**(.015)**	(.13)	(.24)
Age	-.032	.203	-.135	-.38***	-.49**	-.29	.067	-.026	-0.22*	-0.21*	-0.51**	-0.42*
Male	1.8***	1.4***	1.8***	1.8***	1.6***	2.0***	1.6***	1.2***	1.4***	1.8***	2.2***	2.2***
European Ancestry	.180**	.068	.037	.088	.131	-.062	.155**	.131**	.092*	.063	.002	-.081
African Ancestry	.009	-.008	.056	.015	-.2440	.217	.074	.025	.073	.114	.149	-.224
Cognitive score 90-110	-	-	-	-	-	-	-	-	-	-	-	-
<90	0.150	-0.251	0.295	0.853	0.383	-0.631	-0.215	-0.533	0.465	0.553	-0.594	-0.498
110-150	-0.716	-0.036	0.443	-0.376	-0.360	0.373	-0.116	0.272	0.207	-0.312	-0.460	-0.016
Parent unemployment	-0.230	0.165	-0.452	-0.288	1.070	0.560	0.368	0.516	-0.695*	-0.380	0.769	0.209
Parent education high school	-	-	-	-	-	-	-	-	-	-	-	-
<High school	-0.100	-0.556	-1.017*	0.351	1.424	-0.702	0.753	-0.215	-0.883*	0.162	0.771	0.119
>Highs school	0.316	0.127	-0.110	0.101	0.568	0.436	0.409	-0.241	-0.307	0.139	-0.173	0.166
Family income 20k-60k	-	-	-	-	-	-	-	-	-	-	-	-
0-20k	-0.602	-0.559	-0.594	-0.481	-0.079	-0.271	-0.608	-0.48	-0.685	-0.607	-0.785	-1.220
>60k	0.151	0.015	0.014	0.143	0.060	-0.375	-0.058	-0.059	0.132	0.028	-0.370	-0.536
2 biological parents	-0.514	-1.14**	-0.154	0.072	0.140	0.652	-0.049	-0.60	-0.298	0.134	-0.157	-0.489
Household size 3-6	-	-	-	-	-	-	-	-	-	-	-	-
Household size 1-2	0.940	-0.879	1.114	1.081	0.050	1.327	1.546	0.124	-1.098	0.202	-0.959	-2.98**
Household size >7	-0.022	0.390	0.461	-0.515	-0.951	-0.23	-0.188	-0.02	0.244	0.093	-0.459	-1.12**
Church weekly	0.215	-0.259	-0.392	-0.358	-0.96*	-0.547	0.329	0.271	-0.252	-0.268	-0.330	-0.503
N	485	478	706	830	421	304	482	474	704	827	418	303

1 Genetic propensity for binge drinking was measured by the same 5 SNPs from the ROOM stepwise regression (Table 4).
The predicted genetic propensity score ranging 0 to 1 is used to divide the entire sample into three groups with 2 cutoff points at 0.3 and 0.8.
2 ***= p-value<.001; **= p-value<.01; *= p-value<.05; #= p-value<.10.
3 The boldface type is used only to highlight the effects of primary interest. P-values are provided only for these coefficients.

Table 7. Full models of peer influence by genetic propensity interaction on binge drinking with peer effect estimated in a single regression model: ROOM

College binge drinking →	**First semester binge drinking**		**Past semester binge drinking**		**Past 2 week binge drinking**	
Genetic propensity	medium	low & high	medium	low & high	medium	low & high
N of SNPs	5	5	5	5	5	5
Roommate drank/self medium	**.764* (.036)**	.746#	**.585# (.101)**	0.641#	**0.821* (.019)**	0.445
Roommate nondrank /self low or high	0.018	-	-0.057	-	0.376	-
Roommate nondrank/self medium	-	-0.018	-	0.057	-	-0.376
Roommate drank/self low or high	0.184	**0.167 (.69)**	0.026	**0.083 (.84)**	0.089	**-0.288 (.48)**
Respondent characteristics						
Female	-1.09***	-1.09***	-1.02***	-1.02***	-1.29***	-1.29***
Father's education	0.005	0.005	0.011	0.011	-0.017	-0.017
Mother's education	0.025	0.025	-0.014	-0.014	0.010	0.010
Family income $10,000	0.029*	0.029*	0.039**	0.039**	0.040***	0.040***
GPA	-0.328	-0.328	-0.385	-0.385	-0.305	-0.305
SAT/100	0.018	0.018	0.023	0.023	0.001	0.001
Nonwhite roommate	-0.175	-0.175	-0.211	-0.211	0.078	0.078
Church attendance weekly	-1.39***	-1.39***	-1.34***	-1.34***	-1.22***	-1.22***
Roommates characteristics						
Fathers' education	-0.032	-0.032	-0.033	-0.033	-0.099	-0.099
Mother's education	-0.052	-0.052	0.033	0.033	0.071	0.071
Family income $10,000	0.011	0.011	0.003	0.003	0.005	0.005
GPA	0.205	0.205	0.266	0.266	0.306	0.306
SAT/100	0.059	0.059	-0.033	-0.033	0.095	0.095
Church attendance weekly	0.161	0.161	-0.038	-0.038	-0.128	-0.128
Bioancestry(African)	-.042	-.042	0.071	0.071	-0.432	-0.432
Bioancestry(European)	1.39*	1.39*	1.42*	1.42*	1.31	1.31
N	1,003	1,003	1,003	1,003	1,003	1,003

1 Each column presents the coefficients from a single regression model. The three "medium" models test the effect of pairing with a roommate who drank in high school relative to pairing with a roommate who did not drink in high school given self medium genetic propensity. In contrast, the three "low or high" models test the same effect given self low or high genetic propensity. The two bolded coefficients for each binge drinking outcome provide the GE interaction estimates.
2 Genetic propensity for alcohol use was measured by 5 SNPs from the results for a stepwise logistic regression (Table 4)
3 ***= p-value<.001; **= p-value<.01; *= p-value<.05; #= p-value<.10.
4 The boldface type is used only to highlight the effects of primary interest. P-values are provided only for these coefficients.

Table 8. Full models of peer influence by genetic propensity interaction on binge drinking with peer effect estimated in a single regression model: Add Health

Wave III →	**Binge drinking past 2 weeks monthly episodes**				**Bring drinking past year monthly days**			
Genetic propensity	medium		low and high		Medium		low and high	
N of SNPs	5	27	5	27	5	27	5	27
Friend drank/self medium	**1.026*** (.0007)**	**0.648* (.038)**	0.405	0.833**	**0.540* (.055)**	**0.619* (.03)**	0.352	0.670*
Friend nondrank /self low or high	0.621*	-0.184	-	-	0.188	-0.052	-	-
Friend nondrank/self medium	-	-	-0.621*	0.184	-	-	-0.188	0.052
Friend drank/self low or high	.840**	.299	**.219 (.42)**	**.484 (.08)**	.619*	.285	**.431 (.12)**	**.337 (.22)**
Age	-0.191*	-0.193**	-0.191*	-0.193**	-0.206*	-0.210**	-0.206*	-0.210**
Male	1.79***	1.78***	1.79***	1.78***	1.69***	1.69***	1.69***	1.69***
European Ancestry	.102**	.096**	.102**	.096**	.094**	.088**	.094**	.088**
African Ancestry	-.007	.002	-.007	.002	.020	.027	.020	.027
Cognitive score (90-110)								
<90	0.243	0.266	0.243	0.266	-0.001	0.012	-0.001	0.012
110-150	-0.120	-0.100	-0.120	-0.100	-0.086	-0.078	-0.086	-0.078
Parent unemployment	-0.080	-0.055	-0.080	-0.055	-0.026	-0.011	-0.026	-0.011
Parent education high school								
<High school	-0.145	-0.168	-0.145	-0.168	0.040	0.031	0.040	0.031
>Highs school	0.179	0.174	0.179	0.174	-0.007	-0.006	-0.007	-0.006
Family income 20k-60k								
0-20k	-0.456	-0.438	-0.456	-0.438	-0.672	-0.659*	-0.672	-0.659*
>60k	0.066	0.071	0.066	0.071	-0.050	-0.056	-0.050	-0.056
Having 2 biological parents	-0.208	-0.218	-0.208	-0.218	-0.189	-0.208	-0.189	-0.208
Household size 1-2	0.303	0.281	0.303	0.281	-0.228	-0.207	-0.228	-0.207
Household size >7	-0.084	-0.066	-0.084	-0.066	-0.095	-0.094	-0.095	-0.094
Church weekly	-0.347	-0.364	-0.347	-0.364	-0.148	-0.153	-0.148	-0.153
N	1,612	1,612	1,612	1,612	1,604	1,604	1,604	1,604

1 Each column presents the coefficients from a single regression model. The three "medium" models test the effect of pairing with a friend who drank at Wave I relative to pairing with a roommate who did not drink at Wave I given self medium genetic propensity. In contrast, the three "low or high" models test the same effect given self low or high genetic propensity. The two bolded coefficients for each binge drinking outcome provide the GE interaction estimates.
2 Genetic propensity for alcohol use was measured by 5 SNPs at first to replicate the findings from ROOM and then by 27 SNPs
3 ***= p-value<.001; **= p-value<.01; *= p-value<.05; #= p-value<.10.
4 The boldface type is used only to highlight the effects of primary interest. P-values are provided only for these coefficients.

Binge drinking	Interaction with mean effects Friend drank × self medium				Two propensity group difference Friend drank: medium- low and high		
	Averaged effect (std.err)	Averaged t (p value)	t adjusted (p value)	t (p value)	Wilcoxon signed rank Test		
					Averaged diff (s.d)	Lower 95% CL	Upper 95% CL
past 2 weeks	0.650(0.507)	1.285(0.199)	1.283(0.200)	1.285(0.199)	0.650(0.311)***	0.627	0.656
past year	0.094(0.513)	0.182(0.856)	0.183(0.855)	0.183(0.855)	0.094(0.300)***	0.077	0.104
1st semester	0.335(0.525)	0.639(0.523)	0.638(0.524)	0.638(0.524)	0.335(0.305)***	0.330	0.355

We conducted two analyses to test the effect of roommate drank conditional on the genetic binge drinking propensity. In the 1st analysis, we test the interaction effect between friend drank and self-medium propensity. The value reported under "Averaged effect" is the mean of the interaction term across 5 imputation and 500 replicates. The pooled standard error was estimated according to Rubin (1987) work. There are 3 t values and their corresponding p values reported. The first one—averaged t, is the mean of t statistics among 5 imputation and 500 replicates. The second one—t adjusted, is the t statistic using Rubin (1987) formula. The last one is directly calculated by the averaged effect over the pooled standard error.

We can clearly see that none of these tests is significant. The Rubin (1987) formula which combines within imputation and between imputation variances together might not be appropriate in our case. Usually the within imputation variance is the dominate factor in the pooled standard error (Steyerberg, 2009). However, this might not be true in our case. For example, using the past 2week binge drinking as the dependent variable, the between variance and within variance are .24 and .0004, respectively.

In the 2nd analysis, we calculated the difference between the medium and low/high groups for the estimated effect of roommate drank. And then we conducted a Wilcoxon signed rank test to see if the difference equals to zero. The Wilcoxon signed rank test is a nonparametric test, which does not require the distribution assumption for paired samples. The mean difference with standard deviation, and the confidence interval based on median are reported. The significance is indicated by asterisks.

Appendix I. Regression coefficients and p-values of the 27 SNPs selected by step-wise regression from the entire panel of SNPs excluding AIMs available from Add Health. Two of the three bioancestry scores are included in the regression as controls. All 27 SNPs are simultaneously statistically significant at the 0.05 level in a single regression.

SNP	Gene	Coefficient	p-value
rs1008098	OPCML	0.209	0.0095
rs10456876	FYN	0.1393	0.0326
rs10865408	TACR1	0.1747	0.0254
rs10894669	OPCML	-0.2566	0.0036
rs11015015	GAD2	-0.1721	0.0273
rs11609535	LMO3	0.1916	0.0263
rs12514354	CAMK2A	0.2075	0.0043
rs13245899	MUC3B	-0.2622	0.0029
rs1952586	ESR2	0.2424	0.0083
rs2000589	OPCML	-0.203	0.0026
rs2158029	ALDH2	0.3572	0.008
rs2161382	TRPC7	0.2027	0.0056
rs238300	CTNNBL1	0.1634	0.0147
rs324576	CHRM2	0.2341	0.02
rs376063	APP	0.2204	0.0179
rs4578395	OPCML	0.1663	0.0325
rs5911570	GRIA3	-0.1519	0.0112
rs6869634	CAMK2A	0.2869	0.0002
rs7135281	LMO3	0.2179	0.0016
rs7195954	FAM86A	-0.1581	0.0307
rs759588	TACR1	-0.1807	0.0119
rs762513	FAM50A	-0.3051	0.0423
rs7805828	IL6	0.1784	0.0078
rs7885398	MAOA	0.5508	0.001
rs806368	CNR1	-0.1646	0.0329
rs827419	ESR1	-0.1981	0.0082
rs985933	HTR2A	0.136	0.0381

References

Barr, C. S., T. K. Newman, S. Lindell, C. Shannon, M. Champoux, K. P. Lesch, S. J. Suomi, D. Goldman, and J. D. Higley. 2004. "Interaction between serotonin transporter gene variation and rearing condition in alcohol preference and consumption in female primates." *Archives of General Psychiatry* 61:1146-1152.

Barr, C. S., T. K. Newman, C. Shannon, C. Parker, R. L. Dvoskin, M. L. Becker, M. Schwandt, M. Champoux, K. P. Lesch, D. Goldman, S. J. Suomi, and J. D. Higley. 2004. "Rearing condition and rh5-HTTLPR interact to influence limbic-hypothalamic-pituitary-adrenal axis response to stress in infant macaques." *Biological Psychiatry* 55:733-738.

Borsari, B. and K. B. Carey. 2001. "Peer influences on college drinking: A review of the research." *Journal of Substance Abuse* 13:391-424.

Bulchand, S., L. Subramanlan, and S. Tole. 2003. "Dynamic spatiotemporal expression of LIM genes and cofactors in the embryonic and postnatal cerebral cortex." *Developmental Dynamics* 226:460-469.

Christakis, Nicholas A. and James H. Fowler. 2007. "The Spread of Obesity in a Large Social Network over 32 Years." *N Engl J Med*:370-379.

Crabb, David W, Michinaga Matsumoto, Chang Chang, and Min You. 2004. "Overview of the role of alcohol dehydrogenase and aldehyde dehydrogenase and their variants in the genesis of alcohol-related pathology." *Proceedings of the Nutrition Society* 63:49-63.

Crabbe, J. C., T. J. Phillips, K. J. Buck, C. L. Cunningham, and J. K. Belknap. 1999. "Identifying genes for alcohol and drug sensitivity: recent progress and future directions." *Trends in Neurosciences* 22:173-179.

Dick, D. M., J. C. Wang, J. Plunkett, F. Aliev, A. Hinrichs, S. Bertelsen, J. P. Budde, E. L. Goldstein, D. Kaplan, H. J. Edenberg, J. Nurnberger, V. Hesselbrock, M. Schuckit, S. Kuperman, J. Tischfield, B. Porjesz, H. Begleiter, L. J. Bierut, and A. Goate. 2007. "Family-based association analyses of alcohol dependence phenotypes across DRD2 and neighboring gene ANKK1." *Alcoholism-Clinical and Experimental Research* 31:1645-1653.

Duncan, G. J., J. Boisjoly, M. Kremer, D. M. Levy, and J. Eccles. 2005. "Peer effects in drug use and sex among college students." *Journal of Abnormal Child Psychology* 33:375-385.

Enoch, M., P. Shen, K. Xu, C. Hodgkinson, and D. Goldman. 2006. "Using ancestry-informative markers to define populations and detect population stratification." *J Psychopharmacol* 20:19-26.

Foster, G. 2006. "It's not your peers, and it's not your friends: Some progress toward understanding the educational peer effect mechanism." *Journal of Public Economics* 90:1455-1475.

Frayling, T. M., N. J. Timpson, M. N. Weedon, E. Zeggini, R. M. Freathy, C. M. Lindgren, J. R. B. Perry, K. S. Elliott, H. Lango, N. W. Rayner, B. Shields, L. W. Harries, J. C. Barrett, S. Ellard, C. J. Groves, B. Knight, A. M. Patch, A. R. Ness, S. Ebrahim, D. A. Lawlor, S. M. Ring, Y. Ben-Shlomo, M. R. Jarvelin, U. Sovio, A. J. Bennett, D. Melzer, L. Ferrucci, R. J. F. Loos, I. Barroso, N. J. Wareham, F. Karpe, K. R. Owen, L. R. Cardon, M. Walker, G. A. Hitman, C. N. A. Palmer, A. S. F. Doney, A. D. Morris, G. D. Smith, A. T. Hattersley, and M. I. McCarthy. 2007. "A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity." *Science* 316:889-894.

Goldman, D. and A. Bergen. 1998. "General and specific inheritance of substance abuse and alcoholism." *Archives of General Psychiatry* 55:964-965.
Goodman, L. A. and W. H Kruskal. 1954. "Measures of association for cross classifications." *Journal of the American Statistical Association*:732-764.
Gottfredson, M.R. and Travis. Hirschi. 1990. *A general theory of crime*. Stanford, CA:: Stanford University Press.
Greenfield, Norma J and Regina Pietruszko. 1977. "Two aldehyde dehydrogenases from human liver. Isolation via affinity chromatography and characterization of the isozymes." *Biochimica et Biophysica Acta* 483:35-45.
Guo, Guang, Jessica Halliday Hardie, Jonathan K. Daw, Yilan Fu, Hedwig Lee Lee, Amy Lucas, Craig Owen, Emily McKendry-Smith, and Greg J. Duncan. 2009. "DNA COLLECTION IN A RANDOMIZED SOCIAL SCIENCE STUDY OF COLLEGE PEER
EFFECTS." *Sociological Methodology* 39:1-29.
Harris, K.M., F. Florey, J. Tabor, P.S. Bearman, J. Jones, and J.R. Udry. 2003. "The National Longitudinal Study of Adolescent Health: Research design." Available online at: http://www.cpc.unc.edu/projects/addhealth/design.", vol. 2005.
Hirschi, Travis. 1969. *Causes of Delinquency*. Los Angeles, CA: University of California Press.
Jaffee, S. R. and T. S. Price. 2007. "Gene-environment correlations: a review of the evidence and implications for prevention of mental illness." *Molecular Psychiatry* 12:432-442.
Kadrmas, J. L. and M. C. Beckerle. 2004. "The LIM domain: From the cytoskeleton to the nucleus." *Nature Reviews Molecular Cell Biology* 5:920-931.
Kandel, D. B. 1978. "HOMOPHILY, SELECTION, AND SOCIALIZATION IN ADOLESCENT FRIENDSHIPS." *American Journal of Sociology* 84:427-436.
Kornhauser, Ruth. 1978. *Social Sources of Delinquency*. Chicago: University of Chicago Press.
Kremer, M. and D. Levy. 2008. "Peer effects and alcohol use among college students." *Journal of Economic Perspectives* 22:189-206.
Lasek, A. W., F. Giorgetti, K. H. Berger, S. Tayor, and U. Heberlein. 2011. "Lmo Genes Regulate Behavioral Responses to Ethanol in Drosophila melanogaster and the Mouse." *Alcoholism-Clinical and Experimental Research* 35:1600-1606.
Li, Yun, Cristen Willer, Serena Sanna, and Goncalo Abecasis. 2009. "Genotype Imputation." *Annual Review of Genomics and Human Genetics* 387-406.
Liang, K. Y., S. L. Zeger, and B. Qaqish. 1992. "MULTIVARIATE REGRESSION-ANALYSES FOR CATEGORICAL-DATA." *Journal of the Royal Statistical Society Series B-Methodological* 54:3-40.
Manski, C. . 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*:60: 531-542.
.
Marchini, Jonathan and Bryan Howie. 2010. "Genotype imputation for genome-wide association studies." *Nature Reviews Genetics*:499-511.
Matsueda, R. L. 1982. "TESTING CONTROL-THEORY AND DIFFERENTIAL ASSOCIATION - A CAUSAL-MODELING APPROACH." *American Sociological Review* 47:489-504.
Maxson, S. C. and A. Canastar. 2003. "Conceptual and methodological issues in the genetics of mouse agonistic behavior." *Hormones and Behavior* 44:258-262.
Maxson, Stephen. In press. "The genetics of offensive aggression in mice." in *Handbook of Behavior Genetics*, edited by K. Y. Kim. New York: Springer.

Moffitt, R.A. 2001. "Policy interventions, low-level equilibria and social interactions." edited by S. Durlauf and P. Young. Cambridge. MA: MIT Press.
Neville, M. J., E. C. Johnstone, and R. T. Walton. 2004. "Identification and characterization of ANKK1: A novel kinase gene closely linked to DRD2 on chromosome band 11q23.1." *Human Mutation* 23:540-545.
Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. "Inference of population structure using multilocus genotype data." *Genetics* 155:945-959.
Remedios, R., L. Subramanian, and S. Tole. 2004. "LIM genes parcellate the embryonic amygdala and regulate its development." *Journal of Neuroscience* 24:6986-6990.
Rubin, D.B. . 1987. *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley and Sons
Sacerdote, B. 2001. "Peer effects with random assignment: Results for Dartmouth roommates." *Quarterly Journal of Economics* 116:681-704.
Sampson, Robert J. and John H. Laub. 1993. *Crime in the making: Pathways and turning points through life*. Cambridge, MA: Harvard University Press.
Sutherland, E. H. 1947. *Principles of Criminology, 4th ed.* . Philadelphia: J. B. Lippincott.
Tsai, L. T. Y., R. J. Bainton, J. Blau, and U. Heberlein. 2004. "Lmo mutants reveal a novel role for circadian pacemaker neurons in cocaine-induced behaviors." *Plos Biology* 2:2122-2134.
Zimmerman, D. J. 2003. "Peer effects in academic outcomes: Evidence from a natural experiment." *Review of Economics and Statistics* 85:9-23.