

Genotypes over-represented among college students are linked to better cognitive abilities and socioemotional adjustment

Chuansheng Chen · Chunhui Chen ·
Robert K. Moyzis · Qinghua He · Xuemei Lei ·
Jin Li · Bi Zhu · Gui Xue · Qi Dong

Accepted: 2 February 2013 / Published online: 12 March 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract Although much is known about college students as a special sample in terms of their behavioral traits such as intelligence and academic motivation, no studies have examined whether college students represent a “biased” sample in terms of their genotype frequencies. The present study investigated this issue by examining the Hardy–Weinberg equilibrium of genotype frequencies of 284 SNPs covering major neurotransmitter genes in a sample of 478 Chinese college students, comparing these frequencies with those of a community sample (the 1000 Genomes dataset), and examining behavioral correlates of the SNPs in Hardy–Weinberg disequilibrium. Results showed that 24 loci showed Hardy–Weinberg disequilibrium among college students, but only two of these were in disequilibrium in the 1000 Genomes sample. These loci were found to be associated with mathematical abilities, executive functions, motivation, and adjustment-related behaviors such as alcohol use and emotion recognition. Generally, genotypes overrepresented in the

Electronic supplementary material The online version of this article (doi:
[10.1007/s40167-013-0003-3](https://doi.org/10.1007/s40167-013-0003-3)) contains supplementary material, which is available to authorized users.

C. Chen (✉)
Department of Psychology and Social Behavior, University of California, Irvine,
CA 92697-7085, USA
e-mail: cschen@uci.edu

C. Chen · X. Lei · J. Li · B. Zhu · G. Xue · Q. Dong (✉)
State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University,
Beijing 100875, China
e-mail: dongqi@bnu.edu.cn

R. K. Moyzis
Department of Biological Chemistry and Institute of Genomics and Bioinformatics,
University of California, Irvine, CA, USA

Q. He · G. Xue
Brain and Creativity Institute, University of Southern California, Los Angeles, CA, USA

college sample showed better performance and adjustment than under-represented or non-biased genotypes. This study illustrates a new approach to studying genetic correlates of traits associated with a socially-selected group—college students—and presents the first evidence of genetic stratification in terms of education attainment.

Keywords Hardy–Weinberg disequilibrium · Adjustment · College students · HapMap · 1000 Genomes

Introduction

College students are probably by far the most often studied human subjects, because they are easily accessible to researchers, most of whom work at universities. This convenience, however, comes with a cost, namely, this group's uncertain representativeness of the populations at large. For decades now, critics (Sears 1986; Henry 2008) have leveled serious complaints against relying on the “narrow data base” of “college sophomores”. Indeed, college students are expected to differ from others in many aspects. By definition, they have different educational experience. They are also typically within a narrow developmental range from late adolescence to early adulthood. They are likely to be healthy. Due to the nature of the college admission process, college students on average also score higher on intelligence tests, display better academic performance, and have higher achievement motivation. Because of close associations between school performance and other behaviors such problem behaviors, college students are also likely to have displayed fewer problem behaviors (at least pre-college). Similarly, college students may also differ from others in their personality traits, social attitudes, and family backgrounds. The latter's importance in school achievement and educational attainment has been clearly demonstrated by sociologists and psychologists for half a century since the Coleman Report (Coleman et al. 1966).

Decades of human behavioral genetics research have shown that individual differences in most human traits such as those mentioned above (intelligence, motivation, personality, social cognition, and psychopathology) are, at least to a moderate extent, due to genetic factors (Greven et al. 2009; Haworth et al. 2009; Petrill et al. 2009). Putting together the non-representativeness of college students and the genetic bases of traits relevant to being college students, an obvious question arises as to whether college students would show biases in their allele frequencies. The current study is the first attempt to examine whether college students lack representativeness in their genetic makeup. It uses two approaches: the Hardy–Weinberg equilibrium test and a comparison with a community sample. Finally, when biased phenotypes were discovered, this study further examined whether the relevant genotypes were linked to behaviors relevant to being college students.

More than a century ago, English mathematician Hardy (1908) and German physician Weinberg (1908) independently came to the conclusion that gene pool frequencies should be stable or in equilibrium from generation to generation in absence of disturbances. Mathematically, if alleles A and a of a locus have

frequencies of p and q , respectively, then the expected frequencies of genotypes AA , Aa , and aa can be calculated as p^2 , $2pq$ and q^2 . When genotype frequencies of a given sample show deviations from this equilibrium (thus Hardy–Weinberg disequilibrium or HWD), it is likely that this sample is different from the larger population from which it was drawn or the large population had experienced (within the generation of the sample) disturbances such as inbreeding or severe non-random mating, mutation, and selection. Because most research samples are diverse and large enough, inbreeding and non-random mating should not be a major problem. Mutation rates are typically too low to account for biases in HWD in humans. Similarly, with rare examples such as alleles linked to sickle cell anemia, natural selection within one generation is an unlikely explanation of HWD in humans. Therefore, HWD has been used in recent decades for three purposes. First, it may indicate genotyping errors (Gomes et al. 1999; Hosking et al. 2004). Second, it suggests that a biased nonrandom sample was selected from the larger population. Third, it may indicate that the sample was not biased because of a sampling problem, but “biased” because it is a special sample (e.g., patients). The last possibility has been explored recently to discover disease genes (Feder et al. 1996; Nielsen et al. 1998; Jiang et al. 2001; Lee 2003; Li and Li 2008).

The current study used a large sample of Chinese college students ($N = 478$) and examined the frequencies of their genotypes based on 284 SNPs that cover major neurotransmitter genes (those in the dopamine, serotonin, glutamate, and GABA systems). We first investigated whether any of the SNPs showed HWD. The SNPs with HWD were then compared with a community sample (the 1000 Genomes dataset) to see if the HWD was population-wide. Finally, we examined behavioral correlates of the SNPs with HWD in our college sample.

Materials and methods

Participants

Participants were 480 healthy Chinese college students (average age = 19.9 years, standard deviation = 0.9; 208 males and 272 females) from Beijing Normal University, Beijing, China. All were Han Chinese and in good health (i.e., no history of neurological and psychiatric disorders based on self-reports). This experiment was approved by the IRB of the State Key Laboratory of Cognitive Neuroscience and Learning at Beijing Normal University, China. Written consent form was obtained from each subject after a full explanation of the study procedure. Two participants were excluded from final analysis because of poor genotyping results (see below).

Genetic analysis

We initially selected 332 single nucleotide polymorphisms (SNPs) of well-characterized genes related to neurotransmitters (dopamine, serotonin, GABA, glutamine, VMAT, as well as related enzymes including COMT, MAOA, and

MAOB) because these genes are mostly expressed in the human brain and are good candidates for associations with behaviors of interest. In order to cover as much genetic diversity as possible and to include genes shown to be under recent selection (Wang et al. 2006), we selected SNPs with two approaches: first we selected the tag SNPs (tSNPs) defined by the HapMap project (<http://hapmap.ncbi.nlm.nih.gov> (phase 3), The International HapMap Consortium 2007), which are the minimum set of SNPs needed to sample most genetic diversity through linkage disequilibrium (LD). The tSNPs were defined by HapMap in 2007 using the four populations investigated at that time (Europeans, African-Yorubans, Chinese, and Japanese), and used a general r^2 value of 0.8 for identification. Additional SNPs were added for some genes in regions of high LD uncovered in genomic searches for recent adaptive selection (Wang et al. 2006; Hawks et al. 2007). These SNPs covered both coding and regulatory regions (for the latter up to 10 kb beyond the coding region).

Subjects donated a blood sample through venipuncture. DNA was extracted and purified following the standard procedure. The SNPs were genotyped using the standard Illumina Golden Gate Genotyping protocol (see Illumina Golden Gate Assay Protocol for details, www.southgene.com.cn, Shanghai South Gene Technology Co., Ltd, Shanghai, China).

Gene data preprocessing

In addition to automatic calling of genotypes, Illumina genotyping platform supplied a quantitative quality measure known as the GenCall score. It measures how close a genotype is to the center of the cluster of other samples assigned to the same genotypes, compared with the centers of the clusters of the other genotypes. This measure ranges from 0 to 1, with a higher score indicating a more reliable result. The conventional cutoff point is 0.25 (Guan et al. 2009). Two subjects (1 male and 1 female) with more than 10 % null genotyping were excluded. Eighteen SNPs had no genotype calls and were thus excluded. Additional data cleaning included the treatment of low-frequency alleles. Seventeen SNPs that showed no polymorphisms and 12 SNPs that showed too low minor allele frequencies (10 or fewer participants with the minor genotype) were also deleted. In the end, 478 subjects with genotype data on 284 SNPs were retained for this study. Of the remaining SNPs, 779 (0.57 %) genotypes across all subjects had low Gencall <0.25 and were assigned as missing data. Details of genotyped SNPs can be found in Online Supplementary Table S1.

Behavioral measures

To detect potential behavioral correlates of alleles that might have been over-represented among college students, we used a battery of behavior measures covering 12 major aspects of human behaviors: (1) perception, (2) memory, (3) executive function, (4) intelligence, (5) emotion, (6) personality, (7) social relationships, (8) coping style, (9) problem behavior, (10) mathematical abilities, (11) language abilities, and (12) others (e.g., simple reaction times, Iowa Gambling Task). In general each construct was measured with several cognitive tasks or

psychological inventories, with a total of 49 tasks. The specific instruments were selected because they have been widely used in previous research and proved to have good psychometric properties. A detailed list of the measures (domain, tool name, a brief description, measurement index, number of subjects who completed the test, and references) is in the online supplementary materials (Table S2). To avoid potential biases due to a few outliers (values outside of the mean \pm 3 SD), these values were converted to the nearest values within 3 SD.

Data analysis

First, Hardy–Weinberg equilibrium (HWE) index was calculated using the χ^2 test and setting *df* to 1 on all 284 SNPs to identify those loci that showed HWD. For SNPs on autosomes, HWE was calculated with all subjects as well as with males or females separately. For the SNPs on the X chromosome, HWE was calculated for females only. For SNPs that showed HWD, the original genotype results were visually examined to confirm the automatic calling by the machine to rule out obvious genotyping errors. Next we compared our data with those of a community sample (the 1000 Genomes Project, <http://www.1000genomes.org>, whose Chinese sample came from the HapMap Project <http://hapmap.ncbi.nlm.nih.gov>). Finally, we examined behavioral correlates of the HWD loci using ANOVA.

Results

Of 284 SNPs, 24 autosomal loci showed HWD ($P < 0.05$), and none of the X chromosome SNPs showed HWD for females. Half of them (12 SNPs) showed consistent HWD for both males and females. Of the remaining SNPs, six showed HWD for females only and two for males only. Finally four SNPs were significant for the total sample but not when males and females were analyzed separately (although there was one marginal effect for each gender), mainly because the effects for the whole sample were barely significant and the splitting by gender reduced the statistical power.

As Table 1 shows, ten of the 24 SNPs are of genes related to the serotonin system, followed by six related to the dopamine system and two additional SNPs of the dopamine-modulating neurotensin system, and the remaining six are related to the other systems such as GABA, glutamine, monoamine, and VMAT.

If the HWD of the above SNPs indeed reflect the specialness of the college sample, we would expect the same SNPs to show HWE, not HWD, in a community sample. Table 2 shows the genotype frequency distribution of the Chinese sample in the 1000 Genomes database. As expected, only two of the SNPs that showed HWD in our sample showed significant HWD (and an additional marginal one) in the 1000 Genomes Projects' Chinese-Beijing (CHB) sample (Table 2).

To investigate what behavioral correlates were linked to the overrepresented genotypes among college students, we ran ANOVA on behavioral measures by the three genotypes. This analysis was conducted for 19 of the above 24 loci because five of the loci had too few cases (0–3) for the under-represented genotype to allow

Table 1 SNPs that showed HWD in our sample

SNP	Chr.	Gene	System	Major	N	Heter.	N	Minor	N	HWE <i>P</i>	Male HWE <i>P</i>	Female HWE <i>P</i>
rs6962356	7	DDC	Dopamine	AA	94	AG	359	GG	24	0.000	0.000	0.000
rs10033951	4	DRD5	Dopamine	GG	92	AG	328	AA	58	0.000	0.000	0.000
rs1850744	4	DRD5	Dopamine	GG	303	AG	174	AA	0	0.000	0.005	0.000
rs4102942	4	DRD5	Dopamine	GG	105	AG	325	AA	47	0.000	0.000	0.000
rs7685396	4	DRD5	Dopamine	CC	319	AC	150	AA	2	0.000	0.006	0.002
rs10053602	5	SLC6A3	Dopamine	AA	373	AG	93	GG	12	0.039	0.964	0.004
rs4930046	11	TH	Dopamine	AA	368	AG	88	GG	21	0.000	0.002	0.000
rs12502104	4	GABRG1	GABA	GG	108	AG	281	AA	75	0.000	0.003	0.000
rs2832411	21	GRIK1	Glutamate	AA	120	AC	262	CC	96	0.030	0.983	0.004
rs1309822	5	NLN	Neurotensin	AA	134	AG	260	GG	83	0.025	0.044	0.217
rs2405254	12	NTS	Neurotensin	TT	138	AT	216	AA	124	0.037	0.366	0.049
rs655888	13	HTR2A	Serotonin	GG	150	AG	214	AA	113	0.034	0.144	0.163
rs6589400	11	HTR3B	Serotonin	AA	305	AC	77	CC	45	0.000	0.000	0.000
rs10937159	3	HTR3D	Serotonin	AA	240	AC	183	CC	55	0.030	0.378	0.039
rs7733401	5	HTR4	Serotonin	CC	123	AC	263	AA	92	0.022	0.504	0.013
rs10917509+	1	HTR6	Serotonin	AA	309	AG	2	GG	167	0.000	0.000	0.000
rs4912138	1	HTR6	Serotonin	AA	138	AG	216	GG	124	0.037	0.309	0.060
rs140701	17	SLC6A4	Serotonin	AA	308	AG	142	GG	28	0.037	0.057	0.340
rs4583306	17	SLC6A4	Serotonin	GG	310	AG	141	AA	27	0.045	0.107	0.263
rs10488683	11	TPHI	Serotonin	AA	168	AG	208	GG	102	0.014	0.019	0.223
rs11179000+	12	TPH2	Serotonin	AA	349	AT	0	TT	129	0.000	0.000	0.000
rs3779671	8	SLC18A1	VMAT	GG	219	CG	189	CC	70	0.007	0.549	0.002

Table 1 continued

SNP	Chr.	Gene	System	Major	Heter.	Minor	N	HWE <i>P</i>	Male HWE <i>P</i>	Female HWE <i>P</i>
rs952858+	8	SLC18A1	VMAT	GG	CG	CC	3	0.000	0.000	0.000
rs4752045	10	SLC18A2	VMAT	GG	CG	CC	281	0.000	0.007	0.000

Statistically significant values are shown in bold. + These three SNPs indicated by the + sign appeared to show few or no heterozygotes. An alternative (and likely) explanation for their HWD is a common genotyping error of having one enzyme not working as well thus grouping all heterozygotes with the minor alleles. This suspicion was supported by comparing the allele frequencies of our sample with the 1000 Genomes Project's sample (assuming genotyping error, we found the same allele frequencies for the two groups). Furthermore, these SNPs did not have any significant cognitive correlates. Therefore their inclusion or exclusion did not change our main conclusions

Table 2 Genotype frequency and HWE in the 1000 Genomes Project's Chinese sample (CHB) for the SNPs included in Table 1

SNP	Chr.	Gene	System	Major	<i>N</i>	Heter.	<i>N</i>	Minor	<i>N</i>	HWE <i>P</i>
rs6962356	7	DDC	Dopamine	AA	25	AG	39	GG	33	0.06
rs10033951	4	DRD5	Dopamine	GG	13	AG	52	AA	32	0.26
rs1850744	4	DRD5	Dopamine	GG	61	AG	31	AA	5	0.68
rs4102942	4	DRD5	Dopamine	GG	27	AG	41	AA	29	0.13
rs7685396	4	DRD5	Dopamine	CC	67	AC	29	AA	1	0.26
rs10053602	5	SLC6A3	Dopamine	AA	69	AG	25	GG	3	0.69
rs4930046	11	TH	Dopamine	AA	65	AG	28	GG	4	0.17
rs12502104	4	GABRG1	GABA	GG	45	AG	43	AA	9	0.78
rs2832411	21	GRIK1	Glutamate	AA	24	AC	55	CC	18	0.17
rs1309822	5	NLN	Neurotensin	AA	28	AG	42	GG	27	0.19
rs2405254	12	NTS	Neurotensin	TT	16	AT	55	AA	26	0.15
rs655888	13	HTR2A	Serotonin	GG	22	AG	53	AA	22	0.36
rs6589400	11	HTR3B	Serotonin	AA	69	AC	25	CC	3	0.69
rs10937159	3	HTR3D	Serotonin	AA	50	AC	36	CC	11	0.26
rs7733401	5	HTR4	Serotonin	CC	24	AC	50	AA	23	0.76
rs10917509	1	HTR6	Serotonin	AA	68	AG	29	GG	0	0.08
rs4912138	1	HTR6	Serotonin	AA	27	AG	49	GG	21	0.89
rs140701	17	SLC6A4	Serotonin	AA	70	AG	21	GG	6	0.02
rs4583306	17	SLC6A4	Serotonin	GG	70	AG	21	AA	6	0.02
rs10488683	11	TPH1	Serotonin	AA	30	AG	45	GG	22	0.36
rs11179000	12	TPH2	Serotonin	AA	29	AT	47	TT	21	0.74
rs3779671	8	SLC18A1	VMAT	GG	49	CG	40	CC	8	0.97
rs952858	8	SLC18A1	VMAT	GG	53	CG	37	CC	7	0.88
rs4752045	10	SLC18A2	VMAT	GG	38	CG	40	CC	19	0.16

(1) Because the CHB sample size was relatively small, we did not calculate HWE for males and females separately. (2) There were discrepancies in labeling the alleles between the 1000 Genome Project and our study because we used the Illumina's TOP/BOT rule for allele coding, while the 1000 Genome Projects used the forward strand rule. To allow for easy comparison, we changed the labels of the 1000 Genomes Project to the Illumina labels as in Table 1 according to the A/T and C/G complementary rule. (3) Statistically significant values are shown in bold

for ANOVA. As in HWE analysis, we ran ANOVA with the total sample as well as by gender because some loci showed HWD only for one gender and because a number of studies (Harrison and Tunbridge 2008; Chen et al. 2011b) have recently reported gender specificity in gene-behavior associations. Table 3 shows the significant ($P \leq 0.01$) associations between SNPs and behaviors. Significant effects were found for mathematics performance, executive functions, motivation, and other behaviors such as alcohol use. We detail the results in the following paragraphs.

First, genotypes overrepresented in our college sample were associated with better mathematical abilities than were their underrepresented counterparts. Specifically, heterozygotes of two SNPs (rs2832411, rs12502104) were overrepresented and they performed better in two basic mathematical cognition tasks (dot estimation and parity

Table 3 Significant behavioral correlates ($P < 0.01$) of HWD Loci

Behavior	SNP	Mean1 (sd1)	N1 (exp)	Mean2 (sd2)	N2 (exp)	Mean3 (sd3)	N3 (exp)	F	P	Post-hoc	Sample
Cognitive abilities											
Math (dots est. (deviation))	rs2832411	8.92 (1.62)	59 (67)	8.23 (1.50)	133 (119)	9.10 (1.91)	47 (53)	7.00	0.001	2 < 1, 2 < 3	Female
Parity judgment (RT)	rs12502104	494.81 (59.77)	59 (109)	467.49 (47.44)	140 (100)	485.27 (61.27)	33 (23)	6.02	0.003	2 < 1	Female
Stroop (effect size)	rs2832411	136.36 (81.13)	117 (129)	120.73 (74.04)	247 (228)	147.50 (71.67)	93 (101)	5.51	0.004	2 < 3	Total
Stroop (effect size)	rs2832411	124.01 (76.98)	61 (74)	116.62 (66.88)	151 (130)	156.48 (75.87)	49 (57)	5.84	0.003	1 < 3, 2 < 3	Female
Stop Signal (RT)	rs6589400	186.87 (47.34)	115 (109)	204.84 (52.68)	24 (41)	146.20 (39.12)	15 (4)	7.17	0.001	3 < 1, 3 < 2	Male
WCST (ER)	rs1309822	0.88 (1.13)	129 (116)	0.76 (1.05)	251 (228)	0.34 (0.62)	76 (112)	6.89	0.001	3 < 1, 3 < 2	Total
WCST (ER)	rs4752045	0.60 (0.89)	145 (163)	0.68 (0.93)	266 (220)	1.09 (1.17)	46 (74)	4.75	0.009	1 < 3, 2 < 3	Total
Reversal learning (ER)	rs655888	19.92 (4.10)	150 (119)	19.61 (4.23)	214 (239)	21.10 (4.17)	113 (119)	4.80	0.009	1 < 3, 2 < 3	Total
Auditory discrimination (CR)	rs1309822	6.93 (4.19)	59 (55)	5.76 (3.79)	112 (100)	8.62 (6.39)	29 (45)	5.32	0.006	2 < 3	Male
Socioemotional characteristics											
Alcohol use	rs4752045	2.99 (3.35)	142 (165)	2.51 (2.90)	271 (222)	4.12 (4.21)	49 (75)	5.54	0.004	1 < 3, 2 < 3	Total
Alcohol use	rs4752045	1.46 (2.10)	76 (94)	1.41 (2.04)	160 (127)	2.82 (3.74)	28 (43)	4.67	0.010	1 < 3, 2 < 3	Female
Alcohol use	rs4102942	3.02 (3.03)	42 (47)	4.89 (3.75)	141 (99)	5.87 (3.07)	15 (52)	5.50	0.005	1 < 2, 1 < 3	Male
WPI-IM	rs10033951	19.46 (8.84)	52 (44)	21.26 (9.42)	189 (130)	15.07 (8.79)	30 (97)	6.04	0.003	3 < 1, 3 < 2	Female

Table 3 continued

Behavior	SNP	Mean1 (sd1)	NI (exp)	Mean2 (sd2)	N2 (exp)	Mean3 (sd3)	N3 (exp)	F	P	Post-hoc	Sample
Temper	rs10033951	36.43 (7.91)	92 (77)	36.19 (8.43)	328 (230)	40.16 (8.68)	58 (171)	5.61	0.003	1 < 3, 2 < 3	Total
Emotion recognition	rs2405254	59.64 (5.49)	138 (96)	58.64 (4.75)	216 (236)	57.40 (4.96)	124 (145)	6.46	0.002	3 < 1, 3 < 2	Total
Emotion recognition	rs2405254	61.40 (4.36)	83 (55)	58.82 (4.57)	119 (134)	58.16 (4.60)	69 (82)	11.68	0.000	2 < 1, 3 < 1	Female
HSP	rs12502104	117.10 (15.21)	63 (124)	125.14 (15.90)	160 (113)	123.4 (13.08)	40 (26)	6.25	0.002	1 < 2, 1 < 3	Female

Cells in bold are overrepresented in our sample of college students

WCST Wisconsin Card Sorting Task, WPI-IM Work Preference Inventory-Intrinsic Motivation, HSP Highly Sensitive Person, RT reaction time, ER error rates, CR correct rates, Mean# mean score; sd# standard deviation; N# sample size; and N#exp expected sample size based on the allele frequencies of the 1000 Genomes data

judgment, respectively) than did homozygotes. In the numerosity estimation task (Krueger 1982), subjects were asked to estimate dot arrays, and their performance was indexed as the mean deviation of their estimates from true values. A low score indicates better number sense. Overrepresented heterozygotes of rs2832411 scored lower than both groups of homozygotes. In the parity judgment task (Dehaene et al. 1993), subjects were asked to judge whether a number was odd or even. Faster reaction times indicate better number sense. Overrepresented heterozygotes of rs12502104 had the shortest mean reaction times (the differences were significant when compared to major allele homozygotes, but not when compared to minor allele homozygotes).

Scores on the four widely used executive functions tasks (i.e., the Color Word Stroop, Stop Signal, Wisconsin Card Sorting, and Reversal Learning tasks) were also associated with HWD SNPs. The first two tasks reflect response inhibition and the last two reversal learning (i.e., mental flexibility or breaking away from learnt rules). In the Color Word Stroop task (Stroop 1935), subjects were asked to name the color of the words while ignoring their meaning. Smaller reaction time differences between incongruent (e.g., the word “red” printed in green) and congruent conditions (e.g., “red” printed in red) reflect better ability to inhibit interference in incongruent trials and thus better executive function. We found that overrepresented heterozygotes of rs2832411 had the best performance (i.e., smallest differences in RT between incongruent and congruent conditions). In the Stop Signal task (Aron and Poldrack 2006), subjects had to inhibit a response that was already under way. The Stop Signal Reaction Time (SSRT) was used to index inhibition performance, with shorter SSRT showing better inhibition. Minor allele homozygotes of rs6589400, which were overrepresented in our sample, showed a better performance than the underrepresented heterozygotes. Overrepresented major allele homozygotes also showed shorter SSRT than heterozygotes, but the difference did not reach statistical significance.

In the WCST (Berg 1948), subjects were asked to sort cards according to changing rules. Similarly, in the Reversal Learning task (Cools et al. 2002), subjects had to find out the changing rules based on feedback. High error rates indicate that subjects could not unlearn the rules they already learnt and adapt quickly to the changing rules. Overrepresented heterozygotes of rs4752045 showed lower error rates than minor allele homozygotes, but overrepresented major homozygotes and heterozygotes of rs1309822 showed higher error rates in WCST than the minor allele homozygotes. Finally, overrepresented major allele homozygotes of rs655888 showed lower errors in the Reversal Learning task than minor allele homozygotes, but did not differ from the underrepresented heterozygotes.

In terms of perceptual abilities, only one significant difference was found. In the auditory discrimination task adapted from Zatorre (2003), subjects were asked to differentiate and name musical notes. Overrepresented heterozygotes of rs1309822 performed more poorly than their minor allele homozygotic counterparts.

In terms of socio-emotional measures, over-represented heterozygotes of rs4752045 (for both the total sample and females) showed lower scores on alcohol abuse [as measured by the alcohol use disorder identification test (AUDIT, Saunders et al. 1993)] than one or both underrepresented genotypes. For males, overrepresented heterozygotes of rs4102942 showed a medium level of alcohol use between

the two types of homozygotes. Overrepresented major homozygotes and heterozygotes of rs10033951 showed higher intrinsic motivation (measured with the Work Preference Inventory, Amabile et al. 1994) and lower temper scores (a brief scale adapted from the Aggression Questionnaire, Buss and Warren 2000) than the underrepresented minor allele homozygotes.

Recognition of emotions (measured with scales from Wang and Markham 1999, and Matsumoto and Ekman 1988) was also associated with one SNP (rs2405254), for the total sample and females: Overrepresented major allele homozygotes showed better emotion recognition than did minor allele homozygotes for the total sample and heterozygotes as well for females.

Finally, highly sensitive personality is characterized as having high levels of sensitivity and arousal to external stimuli (Aron and Aron 1997). We found that overrepresented heterozygotes and minor allele homozygotes of rs12502104 were more sensitive than underrepresented major allele homozygotes.

As mentioned earlier, there was evidence of greater HWD among females than males (six female-specific vs. two male-specific HWD SNPs). Similarly, genetic effects were more obvious for females than for males (seven for females and three for males, see Table 3).

Discussion

This study aimed to examine whether college students, who as a group represent a biased sample in phenotypes, would also show biases in genotypes (or genetic stratification in terms of educational attainment). We found that 24 out of 284 SNPs in a college sample were not in Hardy–Weinberg equilibrium. In contrast, only two of these SNPs showed HWD (and an additional marginal one) in the independent community sample used in the 1000 Genomes Project (CHB). Further gene–behavior association analysis showed that several cognitive and socioemotional measures were significantly related with the HWD SNPs. In general, overrepresented genotypes among college students showed better performance in these cognitive and behavioral measures as compared to underrepresented and nonbiased genotype groups. The biases found appeared to have some gender specificity in both genotypes and their associations with phenotypes. This finding adds to the growing literature on gender specificity in gene–behavior associations (e.g., Jazin and Cahill 2010; Chen et al. 2011b; Dmitrieva et al. 2011; Harrison and Tunbridge 2008).

Although neurotransmitter systems have garnered much interest in molecular behavior genetics, most research has focused on a few well-studied genetic variations such as the DRD4 VNTR, 5-HTTLPR, and MAOA. Few studies have examined a broad array of SNPs. A survey of the online PubMed database has revealed that only a few studies have examined some of the SNPs we identified as in HWD for college students. Specifically, no prior research was found for 12 of the 24 SNPs included in our Table 1 (i.e., rs6962356, rs4930046, rs12502104, rs2832411, rs1309822, rs2405254, rs6589400, rs10937159, rs7733401, rs10917509, rs3779671, and rs952858). In addition, the association between SNP rs4102942 and alcohol use included in the PubMed database (Chen et al. 2011a) was from the same data

as the current study. Interestingly, seven of these SNPs (rs2832411, rs12502104, rs2832411, rs6589400, rs1309822, rs4102942, and rs2405254) showed one or more significant (at $P < 0.01$) behavioral correlates in our study (Table 3). Intuitively these behavioral correlates seem to be of great importance for college enrollment, but they have not been studied much in terms of their genetic bases.

For the SNPs that have been examined in previous research, some were studied as candidate genes for physical illnesses that are of limited relevance to our interest in normal variations related to college students; and only two found positive associations: e.g., rs7685396's link to serum urate (Yang et al. 2010) and rs140701's link to breast cancer (Sangrajang et al. 2010). Among the five alleles that have been linked to mental health issues/disorders, the results are generally consistent with our hypothesis. First, the minor allele (G) of rs140701 of *SLC6A4* was linked to major depression disorder (Dong et al. 2009), and the major allele (A) was linked to panic disorder (Strug et al. 2008). In our college sample, the homozygotes were underrepresented while the heterozygotes were overrepresented. Second, the G allele of SNP rs4583306 of *SLC6A4* (in an LD block) has been linked to panic disorder (Strug et al. 2008) and mood disorder in subjects who experienced childhood physical abuse (Brezo et al. 2009), whereas the heterozygotes were overrepresented in our college sample. Third, it was reported that the G allele of SNP rs10488683 of *TPHI* was linked to the diathesis for suicide attempts (Brezo et al. 2009), whereas our study showed a slight overrepresentation of the A allele homozygotes among college students. Finally, for *SLC18A2* rs4752045, the CC genotype showed a lower level of depressive symptoms in an elderly population (Christiansen et al. 2007). In our college sample, as mentioned earlier, the overrepresented heterozygotes were associated with better executive functions (WCST) and low or moderate alcohol use. The only real exception to our observation that better behavioral outcomes are linked to genotypes overrepresented among college students appears to be SNP rs1850744 of *DRD5*. Pal et al. (2009) found that the A allele was underrepresented among schizophrenia patients and it was also underrepresented in our college sample. In sum, based on the limited evidence available from previous studies on mental disorders and our own findings among a college sample, alleles overrepresented among college students are likely to be associated with better behavioral outcomes. Our results seem to demonstrate the utility of using HWD to search for gene variants linked to special groups beyond patients.

Although we aimed to identify significant correlates of HWD SNPs, it is equally informative to examine the negative findings. For example, we did not find that HWD SNPs were associated with general intelligence or general personality traits, but only to more specific cognitive measures and behaviors. There are two possible explanations. First, because our sample of college students was from a tier 1 university in China, it was thus a highly selected group and had attenuated variations in traits such as intelligence. For example, the average IQ of our sample was 125.6 and the standard deviation was 7.9, as compared to the presumed population mean of about 100 and standard deviation of 15. Second, the effects of genes are more likely to be on specific facets of a behavior, rather than general traits such as intelligence and personality. This idea is consistent with the failed attempts

to identify specific genes for general intelligence (Davis et al. 2010). Of course more research is needed to confirm this, especially by meeting the challenge of identifying the levels of measurements for phenotypes that are most directly affected by genetic factors.

Our findings have two major implications for the current discussion on culture and genes. First, as a progressive cultural institution, higher education is supposed to provide opportunities to all who are qualified and motivated enough to pursue it. This is the first study to show that, by selecting individuals who have strong achievement motivation and academic competence, the competitive process of gaining a higher education (inadvertently) selects genes that are associated with those behavioral patterns. Through this process of genetic stratification, culture is acting as a potentially powerful “natural selection” agent. Researchers should begin to discuss the implications of this finding for our understanding of presumably social, but likely to be partially genetic, stratification and social policies that deal with social/genetic stratification.

Second, if culture is a strong selection agent and culture varies greatly from one nation to another or even within a nation, much more work is needed to delineate how culture shapes genes in different groups. Our findings show one way (i.e., social stratification) in which culture can play an active role in the evolution of genes (Richerson et al. 2010, see also Chiao and Blizinsky 2010). Social stratification can also take different forms in different societies, which might eventually result in different genetic structures in different societies. Closer to the topic of the current study, college admission processes vary across societies. It is thus likely that different societies (even different higher education institutions) select different behavioral patterns for their college students, which consequently create different gene pools in their students. Much more research is needed to examine both culture-general and culture-specific processes in genetic stratification. Closely related to this point, it should also be noted that there is growing evidence that genetic effects on behaviors may vary by ethnic/racial groups (Li et al. 2008; Kimura 2011; Wang et al. 2013). Much more research is needed to ascertain cultural generality or specificity of conclusions on human behaviors and their genetic and neural bases (Chiao and Cheon 2010; Henrich et al. 2010).

There are several limitations to the current study. First of all, although we tried to cover all the major neurotransmitter genes and major aspects of human behaviors, we were only able to select 332 polymorphism loci (with 284 yielding informative data) and 12 aspects of behavior (49 tasks in total). It is extensive, but by no means exhaustive. There are likely other genes outside of the neurotransmitter systems that may show “biases” (HWD) for college students and are linked to behavioral traits not measured in the study. Second, little information was available about our comparison sample—the 1000 Genomes Projects’ Chinese sample from Beijing. All we know is that it came from the neighborhood around Beijing Normal University. It was not clear how many if any college students or graduates were included. Future research should directly contrast college and non-college samples. Third, we analyzed all significant main effects at the $P < 0.01$ level, without using more stringent corrections for multiple comparisons. We deemed this as an exploratory study to see if there were any behavioral or cognitive correlates of the SNPs in HWD. These results should provide bases for future confirmatory hypothesis-testing research.

Acknowledgments This study was supported by the 111 Project of the Ministry of Education of China (B07008) and National Natural Science Foundation of China (31100807 and 31200850). We thank all graduate research assistants who helped with data collection.

References

- Amabile, T. M., Hill, K. G., Hennessey, B. A., & Tighe, E. M. (1994). The Work Preference Inventory: Assessing intrinsic and extrinsic motivational orientations. *Journal of Personality and Social Psychology, 66*(5), 950.
- Aron, E., & Aron, A. (1997). Sensory-processing sensitivity and its relation to introversion and emotionality. *Journal of Personality and Social Psychology, 73*(2), 345–368.
- Aron, A. R., & Poldrack, R. A. (2006). Cortical and subcortical contributions to stop signal response inhibition: Role of the subthalamic nucleus. *Journal of Neuroscience, 26*(9), 2424–2433. doi: [10.1523/JNEUROSCI.4682-05.2006](https://doi.org/10.1523/JNEUROSCI.4682-05.2006).
- Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of General Psychology, 39*(1), 15–22.
- Brezo, J., Bureau, A., Merette, C., Jomphe, V., Barker, E. D., & Vitaro, F. (2009). Differences and similarities in the serotonergic diathesis for suicide attempts and mood disorders: A 22-year longitudinal gene–environment study. *Molecular Psychiatry, 15*(8), 831–843.
- Buss, A. H., & Warren, W. L. (2000). *The aggression questionnaire manual*. Los Angeles: Western Psychological Services.
- Chen, C., Chen, C., Moyzis, R., He, Q., Li, H., & Li, J. (2011a). Genetic variations in the dopaminergic system and alcohol use: A system-level analysis. *Addiction Biology, 17*(2), 479–489.
- Chen, C., Chen, C., Moyzis, R. K., Dong, Q., He, Q., & Zhu, B. (2011b). Sex modulates the association between the COMT gene and personality traits. *Neuropsychopharmacology, 36*(8), 1593–1598.
- Chiao, J. Y., & Blizinsky, K. D. (2010). Culture–gene coevolution of individualism–collectivism and the serotonin transporter gene. *Proceedings of the Royal Society. Biological Sciences, 277*, 529–537.
- Chiao, J. Y., & Cheon, B. K. (2010). The weirdest brains in the world. *The Behavioral and Brain Sciences, 33*(2–3), 88–90. doi: [10.1017/S0140525X10000282](https://doi.org/10.1017/S0140525X10000282).
- Christiansen, L., Tan, Q., Iachina, M., Bathum, L., Kruse, T. A., & McGue, M. (2007). Candidate gene polymorphisms in the serotonergic pathway: Influence on depression symptomatology in an elderly population. *Biological Psychiatry, 61*(2), 223–230.
- Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, A., & Weinfeld, F. (1966). *Equality of educational opportunity* (No. OE-38001). Washington, DC: National Center for Education Statistics, US Department of Health, Education, and Welfare, Office of Education.
- Cools, R., Clark, L., Owen, A. M., & Robbins, T. W. (2002). Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *The Journal of Neuroscience, 22*(11), 4563–4567.
- Davis, O. S. P., Butcher, L. M., Docherty, S. J., Meaburn, E. L., Curtis, C. J. C., & Simpson, M. A. (2010). A three-stage genome-wide association study of general cognitive ability: Hunting the small effects. *Behavior Genetics, 40*(6), 759–767.
- Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122*(3), 371.
- Dmitrieva, J., Chen, C., Greenberger, E., Ogunseitan, O., & Ding, Y. C. (2011). Gender-specific expression of the DRD4 gene on adolescent delinquency, anger and thrill seeking. *Social Cognitive and Affective Neuroscience, 6*(1), 82–89. doi: [10.1093/scan/nsq020](https://doi.org/10.1093/scan/nsq020).
- Dong, C., Wong, M., & Licinio, J. (2009). Sequence variations of ABCB1, SLC6A2, SLC6A3, SLC6A4, CREB1, CRHR1 and NTRK2: Association with major depression and antidepressant response in Mexican-Americans. *Molecular Psychiatry, 14*(12), 1105–1118.
- Feder, J. N., Gnirke, A., Thomas, W., Tsuchihashi, Z., Ruddy, D. A., & Basava, A. (1996). A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nature Genetics, 13*(4), 399–408. doi: [10.1038/ng0896-399](https://doi.org/10.1038/ng0896-399).
- Gomes, I., Collins, A., Lonjou, C., Thomas, N. S., Wilkinson, J., & Watson, M. (1999). Hardy–Weinberg quality control. *Annals of Human Genetics, 63*(Pt 6), 535–538. doi: [10.1017/S0003480099007824](https://doi.org/10.1017/S0003480099007824)

- Greven, C. U., Harlaar, N., Kovas, Y., Chamorro-Premuzic, T., & Plomin, R. (2009). More than just IQ: School achievement is predicted by self-perceived abilities—but for genetic rather than environmental reasons. *Psychological Science*, *20*(6), 753–762. doi:[10.1111/j.1467-9280.2009.02366.x](https://doi.org/10.1111/j.1467-9280.2009.02366.x).
- Guan, L., Wang, B., Chen, Y., Yang, L., Li, J., & Qian, Q. (2009). A high-density single-nucleotide polymorphism screen of 23 candidate genes in attention deficit hyperactivity disorder: Suggesting multiple susceptibility genes among Chinese Han population. *Molecular Psychiatry*, *14*(5), 546–554. doi:[10.1038/sj.mp.4002139](https://doi.org/10.1038/sj.mp.4002139).
- Hardy, G. H. (1908). Mendelian proportions in a mixed population. *Science*, *18*, 49–50.
- Harrison, P. J., & Tunbridge, E. M. (2008). Catechol-O-methyltransferase (COMT): A gene contributing to sex differences in brain function, and to sexual dimorphism in the predisposition to psychiatric disorders. *Neuropsychopharmacology*, *33*(13), 3037–3045.
- Hawks, J., Wang, E. T., Cochran, G. M., Harpending, H. C., & Moyzis, R. K. (2007). Recent acceleration of human adaptive evolution. *Proceedings of the National Academy of Science of the United States of America*, *104*(52), 20753–20758.
- Haworth, C. M., Wright, M. J., Martin, N. W., Martin, N. G., Boomsma, D. I., & Bartels, M. (2009). A twin study of the genetics of high cognitive ability selected from 11,000 twin pairs in six studies from four countries. *Behavior Genetics*, *39*(4), 359–370. doi:[10.1007/s10519-009-9262-3](https://doi.org/10.1007/s10519-009-9262-3).
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*(2–3), 61–83; discussion 83–135. doi:[10.1017/S0140525X0999152X](https://doi.org/10.1017/S0140525X0999152X).
- Henry, P. J. (2008). College sophomores in the laboratory redux: Influences of a narrow data base on social psychology's view of the nature of prejudice. *Psychological Inquiry*, *19*(2), 49–71.
- Hosking, L., Lumsden, S., Lewis, K., Yeo, A., McCarthy, L., & Bansal, A. (2004). Detection of genotyping errors by Hardy–Weinberg equilibrium testing. *European Journal of Human Genetics*, *12*(5), 395–399. doi:[10.1038/sj.ejhg.5201164](https://doi.org/10.1038/sj.ejhg.5201164).
- Jazin, E., & Cahill, L. (2010). Sex differences in molecular neuroscience: From fruit flies to humans. *Nature Reviews Neuroscience*, *11*(1), 9–17. doi:[10.1038/nrn2754](https://doi.org/10.1038/nrn2754).
- Jiang, R., Dong, J., Wang, D., & Sun, F. Z. (2001). Fine-scale mapping using Hardy–Weinberg disequilibrium. *Annals of Human Genetics*, *65*(Pt 2), 207–219. doi:[10.1017/S0003480001008570](https://doi.org/10.1017/S0003480001008570).
- Kimura, T. (2011). East meets West: Ethnic differences in prostate cancer epidemiology between East Asians and Caucasians. *Chinese Journal of Cancer*, *31*(9), 421–429.
- Krueger, L. E. (1982). Single judgments of numerosity. *Attention, Perception, & Psychophysics*, *31*(2), 175–182.
- Lee, W. C. (2003). Searching for disease-susceptibility loci by testing for Hardy–Weinberg disequilibrium in a gene bank of affected individuals. *American Journal of Epidemiology*, *158*(5), 397–400.
- Li, M., & Li, C. (2008). Assessing departure from Hardy–Weinberg equilibrium in the presence of disease association. *Genetic Epidemiology*, *32*(7), 589–599. doi:[10.1002/gepi.20335](https://doi.org/10.1002/gepi.20335).
- Li, H., Gu, S., Cai, X., Speed, W. C., Pakstis, A. J., & Golub, E. I. (2008). Ethnic related selection for an ADH class I variant within East Asia. *PLoS One*, *3*(4), e1881.
- Matsumoto, D., & Ekman, P. (1988). *Japanese and Caucasian facial expressions of emotion (JACFEE) [Slides]*. San Francisco: Intercultural and Emotion Research Laboratory, Department of Psychology, San Francisco State University.
- Nielsen, D. M., Ehm, M. G., & Weir, B. S. (1998). Detecting marker-disease association by testing for Hardy–Weinberg disequilibrium at a marker locus. *American Journal of Human Genetics*, *63*(5), 1531–1540. doi:[10.1086/302114](https://doi.org/10.1086/302114).
- Pal, P., Mihanovic, M., Molnar, S., Xi, H., Sun, G., & Guha, S. (2009). Association of tagging single nucleotide polymorphisms on 8 candidate genes in dopaminergic pathway with schizophrenia in Croatian population. *Croatian Medical Journal*, *50*(4), 361–369.
- Petrill, S. A., Kovas, Y., Hart, S. A., Thompson, L. A., & Plomin, R. (2009). The genetic and environmental etiology of high math performance in 10-year-old twins. *Behavior Genetics*, *39*(4), 371–379. doi:[10.1007/s10519-009-9258-z](https://doi.org/10.1007/s10519-009-9258-z).
- Richerson, P. J., Boyd, R., & Henrich, J. (2010). Colloquium paper: Gene–culture coevolution in the age of genomics. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(Suppl 2), 8985–8992.
- Sangrajrang, S., Sato, Y., Sakamoto, H., Ohnami, S., Khuhaprema, T., & Yoshida, T. (2010). Genetic polymorphisms in folate and alcohol metabolism and breast cancer risk: A case–control study in Thai women. *Breast Cancer Research and Treatment*, *123*(3), 885–893.

- Saunders, J. B., Aasland, O. G., Babor, T. F., Fuente, J. R., & Grant, M. (1993). Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II. *Addiction*, *88*(6), 791–804.
- Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, *51*(3), 515.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.
- Strug, L. J., Suresh, R., Fyer, A., Talati, A., Adams, P. B., & Li, W. (2008). Panic disorder is associated with the serotonin transporter gene (SLC6A4) but not the promoter region (5-HTTLPR). *Molecular Psychiatry*, *15*(2), 166–176.
- The International HapMap Consortium. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, *449*, 851–861.
- Wang, L., & Markham, R. (1999). The development of a series of photographs of Chinese facial expressions of emotion. *Journal of Cross-Cultural Psychology*, *30*(4), 397–410.
- Wang, E., Kodama, G., Baldi, P., & Moyzis, R. K. (2006). Global landscape of recent inferred Darwinian selection for *Homo sapiens*. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(1), 135–140.
- Wang, Y., Li, J., Chen, C., Chen, C., Zhu, B., Moyzis, R. K. et al. (2013). COMT rs4680 Met is not always the “smart allele”: Val allele is associated with better working memory and larger hippocampal volume in healthy Chinese. *Genes, Brain, and Behavior*. doi:10.1111/gbb.12022.
- Weinberg, W. (1908). On the demonstration of heredity in man. In S. H. Boyer (Ed.), *Papers on human genetics* (pp. 4–15). Englewood Cliffs, NJ: Prentice-Hall.
- Yang, Q., Köttgen, A., Dehghan, A., Smith, A. V., Glazer, N. L., & Chen, M. H. (2010). Multiple genetic loci influence serum urate levels and their relationship with gout and cardiovascular disease risk factors clinical perspective. *Circulation: Cardiovascular Genetics*, *3*(6), 523–530.
- Zatorre, R. J. (2003). Absolute pitch: A model for understanding the influence of genes and development on neural and cognitive function. *Nature Neuroscience*, *6*(7), 692–695.