



Systematic Review of Polygenic Gene–Environment Interaction in Tobacco, Alcohol, and Cannabis Use

Joëlle A. Pasman¹ · Karin J. H. Verweij^{1,2} · Jacqueline M. Vink¹

Received: 10 April 2018 / Accepted: 5 May 2019 / Published online: 20 May 2019
© The Author(s) 2019

Abstract

Studies testing the effect of single genetic variants on substance use have had modest success. This paper reviewed 39 studies using polygenic measures to test interaction with any type of environmental exposure (G×E) in alcohol, tobacco, and cannabis use. Studies using haplotype combinations, sum scores of candidate-gene risk alleles, and polygenic scores (PS) were included. Overall study quality was moderate, with lower ratings for the polygenic methods in the haplotype and candidate-gene score studies. Heterogeneity in investigated environmental exposures, genetic factors, and outcomes was substantial. Most studies ($N=30$) reported at least one significant G×E interaction, but overall evidence was weak. The majority ($N=26$) found results in line with differential susceptibility and diathesis–stress frameworks. Future studies should pay more attention to methodological and statistical rigor, and focus on replication efforts. Additional work is needed before firm conclusions can be drawn about the importance of G×E in the etiology of substance use.

Keywords Gene–environment interaction · Alcohol · Tobacco · Cannabis · Polygenic risk · Genetic risk

Introduction

The use of tobacco, alcohol, and cannabis continues to be widespread. Global smoking prevalence in individuals above age 15 is around 23% (World Health Organization 2016). On average, people drink a glass of alcohol per day, with higher estimates for America and Europe (World Health Organization 2014). Lifetime prevalence of cannabis use is 26% in the European Union and up to 44% in the United States (European Monitoring Centre for Drugs and Drug Addiction 2017; U.S. Department of Health and Human Services 2016). Risk factors on biological, social, and psychological level have been found to contribute to individual differences in substance use behaviors.

Edited by Hermine Maes.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10519-019-09958-7>) contains supplementary material, which is available to authorized users.

✉ Joëlle A. Pasman
j.pasman@pwo.ru.nl

¹ Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands

² Amsterdam UMC, Amsterdam, The Netherlands

Genetic vulnerability is an important risk factor. Traditionally, this factor has been investigated using family and twin designs to determine how much variance in a trait is explained by genetic factors (Boomsma et al. 2002). Heritability estimates for substance use, abuse and dependence derived from these types of studies are moderate to high (about 30–75%; Ducci and Goldman 2012; Vink 2016).

Genetic molecular studies have tried to identify specific genetic variants underlying this heritability. In early studies the focus was on candidate-genes, selected based on their proposed biological function. In recent years, researchers have tried to identify genetic variants in a hypothesis-free manner in genome-wide association studies (GWASs), thereby focusing on single-nucleotide polymorphisms (SNPs). Although GWASs have had more success than candidate-gene studies, results are modest: only a handful of variants have been identified for substance use.

To increase power and because behavior is highly polygenic, studies have tried to test the effect of multiple genetic variants simultaneously. Some studies have used combinations of variants that are strongly related [i.e., are in high linkage disequilibrium (LD)] and are transmitted to offspring together in so-called LD blocks. The exact combination of alleles a person has on the variants in such a block is called a haplotype. It has been assumed that the effects of

haplotypes are larger and thus easier to detect than those of single variants. Other studies have sought to increase power by combining several (unrelated) candidate-gene variants in a single sum score. A newer method uses summary statistics of GWASs to create weighted sums of the number of risk variants an individual carries, often called polygenic scores (PS). Research using haplotypes, candidate-gene sum scores, or PS has had some success in predicting substance use phenotypes. However, for all approaches, explained variance is still much smaller than expected based on the heritability estimates from twin research.

Possibly, explained variance can be increased by taking into account the interplay between genetic factors and the environment. In gene–environment interaction (G×E), the effect of a genetic factor depends on the presence of an environmental factor. The premise is that genetic factors underlie biological mechanisms (e.g., stress system responsivity) that make a person more or less vulnerable to environmental circumstances (Belsky and Pluess 2009). Indeed, twin research has shown that the extent to which genetic risk contributes to substance use can depend on environmental factors (Dick 2011). Early molecular genetic studies investigated G×E using single candidate-genes. For example, it was found that childhood maltreatment increased chances of early alcohol initiation more in carriers of the s-allele of the 5-HTTLPR polymorphism of the serotonin transporter gene than in non-carriers (Kaufman et al. 2007). This finding is one of many in line with the diathesis-stress model, stating that adverse environmental circumstances enhance the chance that genetic vulnerability comes to expression (Monroe and Simons 1991). Other G×E frameworks include the differential susceptibility model, posing that genetic predisposition might enhance the effect of adverse, but also of positive environmental factors (Belsky and Pluess 2009). Less commonly it has been predicted that more adverse outcomes arise when genetic plasticity is high and environmental risk is *either high or low*, as both might lead to high stress reactivity (Boyce and Ellis 2005).

Studies using single candidate-genes to test G×E in substance use and other complex phenotypes have yielded mixed findings (see e.g. Do and Maes 2016; Milaniak et al. 2015 for recent reviews). Non-replication and contradicting results seem the rule rather than the exception. The merits of the different theoretical G×E models remain unclear. Low powered study designs and publication bias are likely to have contributed to these mixed findings (Duncan and Keller 2011). To increase power, the logical next step has been to use polygenic rather than single-variant measures in G×E designs.

Whereas previous reviews focused on G×E with single (candidate) genes, this review presents a summary of G×E studies that used a polygenic measure, including haplotype-based measures, sum scores of risk alleles

in candidate-genes, and PS based on SNPs identified in GWASs. We focused on (ab)use of and dependence on tobacco, alcohol, and cannabis, as these are the most frequently used substances, and most literature was available for these substances. No previous studies to our knowledge have attempted a review of G×E with polygenic measures, or developed a method to systematically review study quality. Because this field is relatively new, we included all G×E studies, regardless of the type of environmental exposure under investigation, ranging from cohort effects to childhood trauma. Based on our findings, methodological and theoretical recommendations for future research were formulated.

Methods

For this review, PRISMA guidelines were used (Moher et al. 2009). The study method was preregistered in PROSPERO (CRD42017057478).

Search strategy

Literature searches were conducted in Web of Science, PubMed and Google Scholar, and based on title and abstract potentially relevant articles were added. Only articles published in peer-reviewed journals were considered. Keywords included substance use, gene–environment interaction, and polygenic risk. The exact keyword combinations used can be found in supplemental Table SI. Reference lists were checked for additional articles. The last search was conducted February 1st, 2018.

Study eligibility

Inclusion criteria were met if (a) the study included human subjects; (b) the outcome was some form of tobacco, alcohol, or cannabis use, or a combination thereof; (c) the study was an original research report; (d) the measure of genetic risk comprised a combination of multiple risk variants (i.e., no single variant designs); and (e) an interaction with an environmental variable was tested statistically. Criterion d allowed for studies that looked at multiple variants within one gene. Although this kind of study does not meet the strict definition of ‘polygenic’, it might be more powerful than studies looking at only one variant (Oroszi et al. 2009). Earlier reviews of candidate-gene studies have not explicitly investigated the merits of this method. Criterion e allowed for any demographic/environmental factor that has been investigated in this context, including for example birth cohort or something as specific as roommate’s alcohol use levels in high school.

Table 1 Symbol allocation for quality characteristics of the G×E studies

Method	Characteristic	–	–+	+	Not applicable
All	Study type	Correlational	Case control	Randomized	
	Sample size	< 1000	1000–2500	> 2500	
	Power calculation	No	–	Yes	
	Control for age and sex ^a	None	Descriptive	Statistical	Homogenous sample/ age as predictor or outcome
	Control for ethnicity ^b	None	Descriptive	Statistical	Homogenous sample
	Control for rGE ^c	None	Descriptive	Statistical	Interventions/cohort effects
	Phenotype measures	Self-developed short survey	Validated survey/ interview	Biological/combined measures	Interventions/cohort effects
Haplotype	# of blocks ^d	1–4	–	> 4	
	# of genes ^d	1–3	–	> 3	
	# of variants ^d	< 5	5–10	> 10	
	Rationale for risk haplotype ^e	Debatable	–	Solid	
Candidate	# of genes ^d	1–3	–	> 3	
	# of variants ^d	< 5	5–10	> 10	
	Rationale for risk allele	Debatable	–	Solid	
Polygenic score (PS)	Based on	Overlapping sample GWAS	–	Independent GWAS	
	Discovery sample size	< 10,000	10,000–25,000	> 25,000	
	<i>p</i> value threshold ^f	<i>p</i> < .0001	–	<i>p</i> ≥ .0001	
	Correspondence phenotypes ^g	Weak	Moderate	Strong	

^aGenetic associations may vary in different age and sex groups (Kendler et al. 2008; The Wellcome Trust Case Control Consortium 2007)

^bPopulation stratification resulting from ancestry differences can distort genetic association results (Price et al. 2006); statistical control using principal component analysis is preferable to control for these effects

^cIn gene-environment correlation (rGE) genetic make-up influences to what environment an individual is exposed (only possible in non-randomized studies). These effects can muddle G×E findings (Rathouz et al. 2008a, b)

^dInclusion of more genetic factors in the aggregate predictor was considered better. Cut-offs were based on commonly chosen numbers of variants for these studies

^eThe rationale for defining which haplotype or allele was the active (risk/protective) allele was deemed less strong when it was based on the results of the main analyses in the same sample, rather than on theory or results from independent samples

^fThis threshold most commonly concerns the *p* value for the association between the SNPs and the phenotype in the original GWAS. The lower this value, the fewer SNPs are included in the PS. We considered PS including only a few SNPs as less strong than PS including more SNPs, although the exact optimal threshold depends on several other study characteristics (Chatterjee et al. 2013; Dudbridge 2013)

^gThe more similar the outcome variable is to the original GWAS phenotype on which the PS was based, the better the predictive value (Wray et al. 2014)

Assessment of study quality

For each study, quality characteristics were assessed. Important hallmarks included study design, sample size, power (sample sizes necessary to achieve different levels of power are described in Supplementary Table SII), the method for controlling for confounders (sex, age, population stratification/ethnicity, gene–environment correlation), and phenotype measurement. The quality of the operationalization of the polygenic measure was assessed separately for the haplotype, candidate-gene score, and PS studies.

As no scale exists for assessing the characteristics of this specific type of study, study quality was visualized using symbols (–, +–, +). Symbol allocations for study characteristics are summarized in Table 1. Although literature was consulted for handholds (Table 1), quality cut-offs had to be chosen without objective reference points. Assessment of study quality was done in duplicate (JP, KV); any disagreement was solved through discussion with a third assessor (JV).

Data extraction and evaluation of results

The studies were categorized according to (a) the measure of genetic risk (haplotype, candidate-gene score, or PS), and (b) the nature of the environmental exposure (intervention or other, e.g., traumatic experiences). Further categorization could not be realized due to heterogeneity in environmental factors, outcomes, and study designs. No meta-analysis nor formal publication bias assessment could be attempted because of study heterogeneity, inconsistent statistical reporting, and absence of report of (standardized) effect sizes.

As most studies did report p values for the G×E analysis, a p curve analysis could be conducted (Simonsohn et al. 2014) to give an indication of the strength of the evidence and of the probability that p hacking occurred in the included studies. The assumption of the p curve method is that if the investigated effect is real, there should be more small p values than large p values reported in the literature. If there are more large than small p values, this might be seen as evidence for p hacking or selective reporting; it is more likely that investigators have been conducting tests until they reached a p value just below .05. Supplementary Table SIII summarizes what G×E test statistics were selected from each study. The analysis was conducted twice, once using all the reported G×E p values in each study, and once using only the first reported p value in each study. If a study only reported that the p value was smaller than some threshold (e.g., $p < .05$, $p < .001$), we included this threshold minus one decimal value (e.g., $p = .049$, $p = .00099$) as the estimated p value in the analysis.

Results

Selection

The study selection process is summarized in the flow chart in Fig. 1. In total, 34 articles describing 39 studies were left for inclusion in the systematic review. These studies described results from 27 independent samples.

Study description

In Tables 2, 3, and 4 key features and G×E findings are summarized separately for studies using (a) haplotype, (b) candidate-gene score, and (c) PS measures. Symbols are used to annotate what studies used data from overlapping samples.

Sample characteristics for each study are given in Supplementary Table SIV. Samples consisting of only European descent individuals were overrepresented (51%). Studies included clinical ($N=8$), clinically ascertained ($N=5$), and general populations ($N=26$). Eleven studies included

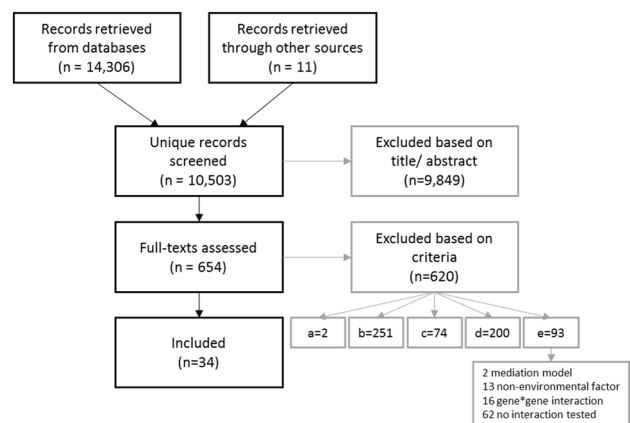


Fig. 1 Flow-chart of study selection for inclusion in the review. Exclusion criteria: **a** non-human subjects; **b** no substance use outcome; **c** no original research; **d** no polygenic risk predictor; **e** no statistical test of interaction with environmental variable

family-related individuals. Studies comprised various age ranges starting from adolescence, with 19 studies specifically focusing on adolescents or young adults and two on older adults. There was an approximately equal representation of female and male subjects within the studies. Sample sizes ranged from $N=81$ to $N=11,423$, with an average of $N=1865$.

Fifteen studies used some form of correlational design (7 longitudinal), 12 were case–control studies, 11 were RCTs and 1 was a randomized longitudinal design. Twenty studies included alcohol and 16 included tobacco outcomes (among others), 4 focused on combined phenotypes (e.g., substance use disorder), and only 3 included cannabis outcomes. There were 11 intervention studies, 12 studies that included measures of trauma-like experiences, and 16 that focused exclusively on typical environmental exposures in for example the family or peer context. Most haplotype studies focused on interventions or psychological trauma as environmental exposures, whereas the candidate-gene score and PS studies more often focused on common environmental factors.

Quality

General quality characteristics per study type are summarized in Table 3. Quality of the implementation of the polygenic method is summarized in Table 4. Full details on quality characteristics per study are given in Supplementary Tables SVa–SVc.

Haplotype method

Haplotype studies ($N=16$) were on average published 7 years before the date of inclusion in this review. They used a strong experimental design (i.e., RCT or case control)

Table 2 Summary of G×E studies using haplotypes as a measure for polygenic risk (G)

ID	Year	1 st author	Genes	Environmental exposure	Outcome	Quality ^a	Finding
INTERVENTION							
1*	2007	Berrettini	COMT	pharmacotherapy vs placebo	7-day smoking abstinence	+/-	3 out of 4 haplotypes predicted abstinence more strongly for individuals randomized to a bupropion intervention. Weakened at 6-month follow-up.
2*	2007	Berrettini	COMT	pharmacotherapy vs placebo	7-day smoking abstinence	+/-	2 out of 4 haplotypes predicted abstinence more strongly for individuals randomized to a bupropion intervention. Not retained at 6-month follow-up.
3	2009	Oroszi	OPRM1	pharmacotherapy vs placebo	alcohol abstinence or moderate drinking	-	Patients treated with naltrexone who had 1 haplotype out of 3 combinations had better outcomes than those treated with placebo or other haplotypes . No interaction effect for haplotypes in other block.
4	2012	Chen	CHRNA5 CHRNA3 CHRN4	pharmacotherapy vs placebo	smoking cessation	+	1 haplotype predicted failed abstinence only for individuals randomized to placebo and not for those treated with nicotine patch, nicotine lozenge, or bupropion
5*	2013	Brody	DRD2 ANKK1 GABRG1 GABRA2	prevention trial	alcohol use	+	Risk haplotypes in 3 out of 5 blocks predicted only predicted increased alcohol use for individuals randomized to control, not for individuals in the prevention group
6	2015	Tyndale	CHRNA5 CHRNA3 CHRN4	pharmacotherapy vs placebo	smoking cessation	+	no interaction
OTHER							
7 ^o	2006	Lerer	HTR6 HTR1B	trauma exposure	smoking initiation & nicotine dependence	-	no interactions
8 ^o	2007	Segman	DAT1	trauma exposure	smoking initiation & nicotine dependence	+/-	in 1 of 2 blocks, 1 of 3 haplotypes protected for the effect of trauma on nicotine dep Not replicated for smok init
9	2008	Ducci	MAOA MAOB	childhood sexual abuse	alcoholism	--	in 1 of 2 blocks, 1 of 5 haplotypes predicted alcoholism, but only in sexually abused individuals
10	2010	Nelson	CRHR1	childhood sexual abuse	alcohol use & alcohol dependence	-	1 of 2 haplotype combinations protected for the effect of childhood sexual abuse on alc consumption and dep
11	2010	Enoch	GABRA2	childhood trauma	alcohol dep & substance dependence	-	in 1 of 2 blocks, 1 of 4 haplotypes was protective for the effects of childhood trauma on subst dep not replicate for alc dep

Table 2 (continued)

ID	Year	1 st author	Genes	Environmental exposure	Outcome	Quality ^a	Finding	
12 ^a	2011	Kranzler	CRHR1	childhood adverse events	alcohol dependence	-	no interaction	
13 ^a	2011	Kranzler	CRHR1	childhood adverse events	alcohol dependence	-	no interaction	
14	2013	Ray	CRHR1	trauma exposure	alcoholism	+/-	in block 1 there was 1 out of 3 and in block 2 there were 2 out of 7 haplotypes that were a risk factor in non-trauma-exposed individuals, but were protective in trauma-exposed individuals	
15	2015	Handley	FKBP5	childhood maltreatment	marijuana dependence	-	childhood maltreatment only predicted marijuana dep in 1 of 2 haplotype combinations	
16	2017	Handley	FKBP5	childhood maltreatment	problem drinking	+/-	the internalizing pathway from maltreatment to alc problems was only present for haplotype carriers	no interaction for externalizing pathway
INTERVENTION								
17	2012	McGeary	ANKK1 DRD4 DAT1 COMT	pharmacotherapy vs placebo	smoking cessation	--	no interaction	
18 [*]	2013	Brody	DRD2 GABRG1 GABRA2	prevention trial	alcohol use	+/-	score predicted alcohol use, but only for the individuals randomized to control and not to prevention trial	
19	2013	David	ANKK1 COMT DRD4 DAT1	pharmacotherapy vs placebo	smoking cessation	+	score predicted time to first lapse only for individuals randomized to placebo	Not replicated for abstinence at end of treatment
OTHER								
20	2015	Guo	DRD2 MAOA LMO3 and others	roommate's binge drinking in high school	binge drinking in college	+/-	medium propensity predicted higher levels of binge drinking when the roommate was a drinker than high/ low propensity. Similar results for larger SNP subsets	
21 [§]	2015	Guo	DRD2 MAOA LMO3 and others	roommate's binge drinking in high school	binge drinking in college	+/-	medium propensity predicted higher levels of binge drinking when the roommate was a drinker than low/ high propensity. Similar results for larger SNP subsets	
22	2016	Bountress	DRD2 ADH4 CNR1 GABRA2 PDYN OPRM1	parental knowledge & peer substance use	substance use disorders	+/-	highest level of score predicted substance use disorder more strongly when parental knowledge was low and peer substance use was high	
23 [§]	2016	Stogner	DAT1 DRD2 DRD4 SERT MAOA	parental rejection	alcohol use initiation	+/-	score was a risk factor for alcohol use when parental rejection was high, but a protective factor when rejection was low	
24 [@]	2017	Pasman	DAT1 DRD4 DRD2 OPRM1	parental education level	polysubstance use	+/-	no interaction	
25	2017	Pasman	DAT1 DRD4 DRD2 OPRM1	(parental) education level	polysubstance use	+/-	score marginally predicted lower polysubstance use for people with a low education level	
26	2017	Coley	DAT1 DRD2 DRD4 DRD5 COMT MAOA	parent and peer alcohol use & stressful events	alcohol use, intoxication & use disorders	+	no interactions	

Table 2 (continued)

ID	Year	1 st author	PS phenotype	Environmental exposure	Outcome	Quality ^a	Finding
INTERVENTION							
27 [#]	2015a	Musci	smoking cessation	prevention trial	age smoking initiation	+-	higher PS predicted stronger effect of the intervention on age of smok init
28 [#]	2016	Musci	smoking cessation	prevention trial	age cannabis initiation	-	higher PS predicted stronger effect of the intervention
OTHER							
29	2012	Vrieze	smoking heaviness	age cohort	smoking heaviness & alcohol use	+	PS predicted smok more strongly in the older cohorts than in the younger cohorts PS predicted less alc use more strongly in the older cohorts than in the younger cohorts
30	2013	Meyers	smoking heaviness	neighborhood cohesion & trauma	smoking heaviness	+	higher PS predicted smok more strongly for high levels of trauma exposure and less so individuals experiencing high neighborhood social cohesion
31 [*]	2014	Salvatore	alcohol problem	parental knowledge & peer deviance	alcohol problems	+-	PS had a stronger effect on alc problems when parental knowledge was low or peer deviance was high Not replicated for analyses using subset of top-SNPs
32 [#]	2015b	Musci	smoking cessation	friend's substance use & parental monitoring	smoking frequency & cannabis frequency	+-	PS in combination with high parental monitoring and low friend's substance use environmental risk predicted lower tobacco and marijuana use
33 [*]	2016	Domingue	smoking initiation	birth cohort	smoking initiation	+	marginal interaction such that influence of PS became more important over time
34 [*]	2016	Schmitz	smoking initiation	veteran status & educational attainment	smoking initiation & smoking heaviness	+	PS predicted smoking initiation and heaviness more strongly in veterans than in non-veterans. Post-war educational attainment buffered for this effect
35	2017	Li	alcohol dependence symptoms	close friend's substance use	heavy episodic drinking	-	no interaction
36 [@]	2017	Treur	smoking heaviness	childhood smoke exposure	smoking heaviness & smoking initiation	+	PS marginally predicted smok heaviness for individuals exposed to smoke in childhood and not for unexposed individuals not replicated for smok init
37 [@]	2018	Mies	alcohol use	stress at home & life satisfaction	alcohol use & alcohol problem		no interaction
38	2018	Polimanti	bipolar disorder, major depressive disorder & schizophrenia	trauma exposure	alcohol misuse & nicotine dependence	+	bipolar disorder PS predicted more alcohol misuse for individuals exposed to trauma and less for unexposed individuals. no interaction for bipolar disorder and schizophrenia PS on nicotine dependence. Inconsistent marginal effects for other GxE analyses
39 [*]	2018	Salvatore	alcohol problems	romantic relationship	alcohol intoxication	+	No interaction

Table 2 (continued)

Top rows for studies testing intervention/prevention as environmental exposure (E). Only first author of the respective papers is mentioned
 Green = reinforcing, dark green = reinforcing such that G only has effect in one E, blue = E has only effect for one level of G, orange = G's effect is reversed by E, gray = no evidence for G×E

*,+,%,&,\$,@,# Studies denoted with the same symbol used data from identical or overlapping samples

^aQuality ratings based on characteristics from Supplementary Tables SVa-SVc

more often than both other study types. Sample sizes were quite low given the expected small effects, with an average of $\hat{N} = 771$ and almost half of the studies using a sample of less than $N = 500$ individuals. In many cases, exact sample sizes were only reported for the main effects analysis and not for the G×E analysis. Power calculations were not reported in 10 out of 16 studies.

Only three studies controlled statistically for both age and sex, although other studies often reported that outcomes and predictors did not vary for different sexes or age groups ('descriptive' control). In a few cases, statistical control was unnecessary as the sample was sufficiently homogeneous (e.g., all female or all within the same 2-year age range). Control for ethnicity was absent or rudimentary in all but one study (16), for example consisting of self-reported 'white/non-white' racial background. In nine studies the sample was reasonably ethnically homogeneous. In non-randomized studies ($N = 10$) gene–environment correlation (rGE) might confound the G×E interaction results. Only four of the ten studies reported on rGE and no studies controlled for these effects.

The quality of the application of the haplotype method was limited. Almost all studies tested haplotypes in one or two LD blocks, with the number of tested variants ranging between 2 and 18. Many studies looked at 1 or 2 blocks in a single gene or a few genes in high LD, so that they are hardly more 'polygenic' than single candidate-gene studies. Most studies did not formulate a literature-based directional prediction. In many cases, only haplotypes that showed a main effect on the outcome were included in the G×E analysis.

For severe outcomes (e.g., clinical diagnosis) or environmental exposures (e.g., traumatic experiences) interviews were used as a measurement instrument. In the smoking cessation trials, biological measures were employed to validate self-reported abstinence. Other outcomes and exposures were mostly measured using validated questionnaires or more crudely using short questionnaires that were developed for the purpose of this or an earlier study.

Most haplotypes investigated were located in genes involved in (dopamine-related) reward and inhibition processes in the brain (e.g. *COMT*, *ANKK1*, *DRD2*, *DAT1*, *OPRM1*, *HTR6*, *HTR1B*, and *GABA*- and *MAO*-related genes). Other candidate-genes included *CRHR1* and *FKBP5*, related to the stress system. Genes such as the nicotine

metabolism, cannabinoid receptor, or the alcohol dehydrogenase genes seem equally suitable candidates, but received much less research attention.

Summarizing, weaknesses of the haplotype studies included small sample sizes, low statistical control, and limitations in the implementation of the polygenic method. Strengths included the use of strong designs and phenotypical measures.

Candidate–gene method

Studies using candidate-gene scores were on average 3 years old at the time of inclusion in this review. The study designs were somewhat less strong than those used in the haplotype studies, with four out of ten using some randomization procedure. Average sample size was larger than for the haplotype studies ($\hat{N} = 2141$), although for 2 studies exact sample size for the G×E analyses were not reported and the average was boosted by one study with $N = 11,423$. Three out of ten studies reported power calculations.

Control for confounders was more stringent than in the haplotype studies, with 8 studies exerting statistical control for both age and sex. Five studies used some control for genetic ancestry and another four used a relatively homogeneous sample. One of the seven studies that did not use a randomization procedure statistically controlled for rGE effects and three studies reported on them.

All but one candidate-gene score study used an unweighted sum score of the number of risk alleles as a predictor. Sum scores were based on risk alleles in on average seven variants. Almost all variants were located in previously investigated candidate-genes related to dopamine-signaling. The rationale for selecting the risk allele was debatable in three cases. Many studies did omitted a description of conflicting literature on the risk allele of the candidate-gene. Outcome and environmental exposure measures were generally of lower quality than those in the haplotype studies and mostly comprised self-developed short questionnaires.

Candidate-gene studies scored slightly better than the haplotype studies on sample sizes, control for confounding, and the implementation of the polygenic method.

Table 3 Summary of general study quality per category, expressed in percentage of studies that met a criterion as specified and assigned with a -, +-, or + in Table 1

	Haplotype	Candidate-gene score	Polygenic score (PS)
Design	% correlational	60.0%	69.2%
	% case control	0%	15.4%
	% randomized	40.0%	15.4%
Sample size	Rating	+ -	-
	M (SD)	N=771 (658)	N=3001 (3439)
Power	% with $N \geq 1000$	60.0%	53.8%
	Rating	-	+ -
	% reported	30%	38.5%
Control for confounders	Rating	-	-
	% age statistically controlled ^a	88.9%	100%
	% sex statistically controlled ^a	100%	100%
	% ethnicity statistically controlled ^a	83.3%	100%
	% rGE reported ^b	40.0%	80.0%
	% rGE statistically controlled ^b	0%	10%
Phenotype measures ^c	Rating	+ -	+
	% self-developed short survey	60.0%	72.7%
	% validated survey/interview	30.0%	27.2%
	Biological/combined measures	10.0%	0%
	Rating	-	-

For details per study, see Supplementary Tables SVa–Vc

^aWhen sample was sufficiently homogeneous, the study was not considered for calculating this percentage

^bWhen rGE could not be an issue (in the case of cohort effects or intervention studies) the study was not considered for calculating this percentage

^cOnly percentages for the outcome measures are mentioned here. For details on environmental measures, see Supplementary Tables SVa–Vc

Table 4 Summary of quality of the implementation of the polygenic method

		Average quality		
		-	+–	+
Haplotype	# of blocks ^a (M = 1.6)	1–4	-	> 4
	# of genes ^a (M = 1.6)	1–3	-	> 3
	# of variants ^a (M = 6.2)	< 5	5–10	> 10
	Rationale for risk haplotype ^b (50% solid)	Debatable	-	Solid
	Overall quality			
Candidate-gene score	# of genes ^a (M = 8.8)	1–3	-	> 3
	# of variants ^a (M = 6.2)	< 5	5–10	> 10
	Rationale for risk allele ^b (67% solid)	Debatable	-	Solid
	Overall quality			
Polygenic score (PS)	Based on (92.3% independent sample)	Overlapping GWAS	-	Independent GWAS
	Discovery sample size (M = 58,447)	< 10,000	10,000–25,000	> 25,000
	<i>p</i> -value threshold ^c (7.7% <i>p</i> < .0001)	<i>p</i> < .0001	-	<i>p</i> ≥ .0001
	Correspondence phenotypes ^d (84.6% weak)	Weak	Moderate	Strong
	Overall quality			

Averages or counts are given per criterion. Shading indicates that most studies (or the study average) fell into this quality category, with the darker shading indicating the average quality category per study type. For details per study, refer to Supplementary Tables SVa–SVc

^aInclusion of more genetic factors in the aggregate predictor was considered better. Cut-offs were based on commonly chosen numbers of variants for these studies

^bThe rationale for defining which haplotype or allele was the risk/protective allele was deemed less strong when it was based on the results of the main analyses in the same sample, rather than on theory or results from independent samples

^cThis threshold most commonly concerns the *p* value for the association between the SNPs and the phenotype in the original GWAS. The lower this value, the fewer SNPs are included in the PS. We considered PS including only a few SNPs as less strong than PS including more SNPs, although the exact optimal threshold depends on several other study characteristics (Chatterjee et al. 2013; Dudbridge 2013)

^dThe more similar the outcome variable is to the original GWAS phenotype on which the PS was based, the better the predictive value (Wray et al. 2014)

Polygenic score method

Studies using PS were the newest, on average 2 years old at the time of inclusion. Study designs were more often simple correlational designs. Sample size was $\hat{N} = 3001$ on average, with only two samples smaller than $N = 500$. Five out of 13 studies reported power calculations.

Control for confounding was most rigorous in this type of studies, with seven studies controlling statistically for both age and sex, and the rest using some procedure rendering statistical control less necessary. In PS studies it is possible to control statistically for ethnicity using ancestry-informative principal components. Eight of the ten studies used this procedure, with an additional two

using a more rudimentary approximation of this method and the final three using a reasonably homogenous sample. Of the ten studies where rGE could have played a role, one exerted statistical control for rGE and eight described the effects without controlling for them.

Half the studies constructed the PS based on results from larger GWASs (average discovery $\hat{N} = 58,447$, range $N = 31,266$ – $74,053$), while the others based it on results from GWASs with limited sample size ($\hat{N} = 3140$). One of the latter calculated the PS based on a GWAS in a sample that was genetically related to the target sample, which could have biased results (35). The similarity between the study outcome and the source GWAS phenotype was in most cases reasonably high. The specific score calculation method

differed somewhat across studies, with three preselecting a subset of SNPs to include in the score. Most studies used the PLINK program to calculate the PS, using pruning or clumping to remove variants that were in high LD. One study used LDpred to calculate PS while accounting for this LD (36). The p value thresholds for including SNPs in the PS varied widely from $p < 1 \times 10^{-8}$ (resulting in a score of only two variants) to $p = 1$ (retaining all SNPs in the score), and five studies tested multiple PS with different thresholds.

Overall, the phenotype measures were of limited quality, with most studies using short self-developed questionnaires.

Overall, the PS studies scored higher than the other study types on sample size, control for confounding, and the implementation of the polygenic method, but similar or worse on study design and phenotypical measures.

In general, study quality appeared somewhat higher for studies that found a significant G×E result than for studies that did not. This might have been driven by the higher quality of the PS studies, that could also have yielded more significant findings because of higher power. Study quality did not seem to influence what kind of G×E pattern was found.

Main results

p curve analysis

To get an indication of the overall evidence for G×E in substance use, a p curve analysis was conducted. The analysis was based on 82 p values (34 significant) derived from 28 studies (see Supplementary Table SIII). The other 11 studies did not report p values for the G×E term specifically (or for the simple effects in the case of cross-over interactions) or statistics by which these could be calculated. As can be seen in Fig. 2, there were more small p values than expected under the null hypothesis. The p curve is flatter than expected if studies had at least 70% power, such that there were not many more small p values than medium to large p values. This indicates that there was evidential value, but this was not very strong. Moreover, if only the first p value reported in each study was taken into account, results deteriorated, indicating that they were driven by a few studies that reported many small p values (data not shown). Results did not change if non-exact p values (e.g., $p < .05$) were excluded from analysis (results not shown). There was no clear evidence for p hacking, which would be indicated by a substantially higher proportion of p values just below 0.05.

G×E patterns

The G×E findings from each of the 39 studies were summarized in the last column of Tables 2, 3, and 4. Thirty studies reported at least 1 significant G×E finding and 9 did not.

Twenty-five of the significant G×E findings followed a pattern as depicted in Fig. 3a (22 studies) or b (3 studies). The pattern in panel a indicates that environmental risk enhances the effect of genetic risk, further increasing the chance of unfavorable outcomes. Or, likewise, a protective genetic predisposition might enhance the effects of a positive environment or counteract the effects of an adverse environment. Thus, in these cases, genetic and environmental factors *reinforce* each other's effects. In panel b the pattern is similar, only now a genetic factor that is a risk factor in one situation, is protective in the other situation, or likewise, an environmental exposure that is a risk factor for individuals with a certain genetic make-up is a protective factor for individuals with a different genetic make-up. Thus, genetic and environmental factors *reverse* each other's effects (cross-over interaction).

The colors in Tables 2, 3, and 4 correspond to specific patterns of results as summarized in Fig. 3. The light green color indicates that the study interpreted the G×E effects such that the genetic and environmental factors reinforce each other (study 1–3, 21, 26–31, 33). For example, in study 22 a genetic risk factor (associated with substance use) enhances the effect of an adverse environmental factor (high peer substance use), yielding a negative outcome (substance use disorder).

Studies marked in darker green are similar, but did not find (or report) a main effect of environment ('ME' in Fig. 3a) or a main effect of the genetic factor ('M' in Fig. 3a; study 4–5, 9, 15–16, 18–19, 36). The interpretation in these cases is that genetic risk *only* has an adverse effect in an adverse situation (study 9, 15, 16, 36) or when there is no intervention to counteract it (study 4, 5, 18, and 19). Likewise, the blue studies find that environmental risk only has an adverse effect in the absence of a protective genetic factor. The 3 studies finding this pattern are all haplotype studies where a specific combination of alleles protects for the effect of psychological trauma (studies 8, 10, 11).

Studies marked in orange (14, 23, and 38) showed that a genetic risk factor becomes a protective factor depending on the environment as depicted in Fig. 3b. For example, in study 14, certain haplotype combinations were risk factors for alcoholism in controls, but protective factors in traumatized individuals.

The yellow studies find patterns that correspond with neither Fig. 3a nor b. Study 20 and 21 find that *medium* levels of genetic risk predict adverse outcomes for high environmental risk. Study 25 reports that high genetic risk predicts *favorable* outcomes for an adverse environmental characteristic (low parental education). Study 29 found a PS for smoking heaviness to be related to less alcohol use for older cohorts, whereas it was related to more smoking in this group. Another cohort study (33) reported that genetic risk marginally predicted adverse outcomes more strongly for young cohorts, even though being in such a cohort is generally viewed as protective for substance use.

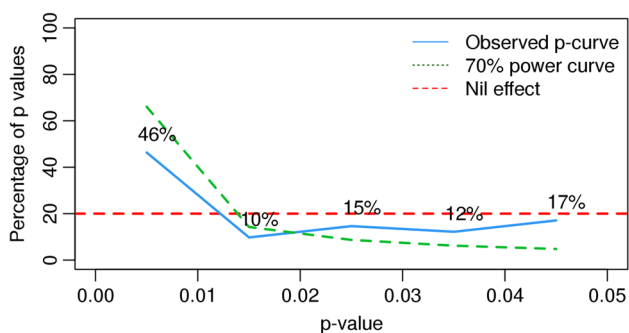


Fig. 2 Percentage (y axis) of reported *p* values in the studies that fell in the range specified on the *x* axis (*p* curve), against the percentage that would be expected under the null hypothesis (nil effect) and under the alternative hypothesis given a power level of 70% (70% power curve)

Patterns substances, genes, and environments

Patterns of G×E did not seem to differ depending on the substance under investigation. Fourteen of the 20 studies including alcohol outcomes, and 13 of the 16 studies including smoking outcomes found at least one significant G×E effect.

Results did not seem to differ depending on the kind of variants investigated. For example, studies looking at dopamine-related genes were not more likely to find significant G×E patterns than studies focusing on other candidate-genes. However, it was difficult to compare findings across gene-groups, as most studies used aggregates of genes from different groups.

Intervention studies and studies looking at trauma exposure seemed more likely to yield patterns corresponding to Fig. 3a or b, but not more likely to show significant results. The 16 studies focusing on common environmental exposures yielded more diverse patterns. For example, all 5 yellow outcomes fell in this category, investigating birth cohort, peer substance use, and parental education level.

Discussion

The aim of this review was to provide an overview of all available studies (*N* = 39) using measures of polygenic risk (haplotypes, candidate-gene scores, and polygenic scores) to investigate gene–environment interaction in substance use. There was some support for the existence of G×E in substance use, but the evidential value was weak.

Theoretical interpretation

Most G×E results followed the pattern as depicted in Fig. 3a. These patterns nicely fit in the diathesis-stress framework (Monroe and Simons 1991), stating that individuals who are at risk genetically show higher levels of some adverse outcome when they are exposed to a risk environment. Although not stated in the original model, the same seems to apply for individuals who have a protective genetic predisposition in that they have more positive outcomes in beneficial environments. It is important to point out that this would fit equally well in the differential susceptibility framework (Belsky and Pluess 2009), but it is rarely found (or reported) within studies that the same genetic factor has a positive effect in one situation and a negative effect in the other. Only three studies report such an effect (14, 23, and 38), providing direct evidence for a ‘genetic plasticity factor’ yielding differential susceptibility.

Many studies did not provide a strong theoretical framework for predicting one G×E pattern rather than another. It seems that a fitting theoretical explanation can be found regardless of the pattern that was discovered. For example, findings that environmental factors have a stronger effect at *medium* levels of genetic risk have been explained lending from a ‘social push’ model framework, stating that a risk factor is overruled at particularly low or high levels of another risk factor (Guo et al. 2015). Researchers may be tempted to place their findings in a theoretical framework a posteriori,

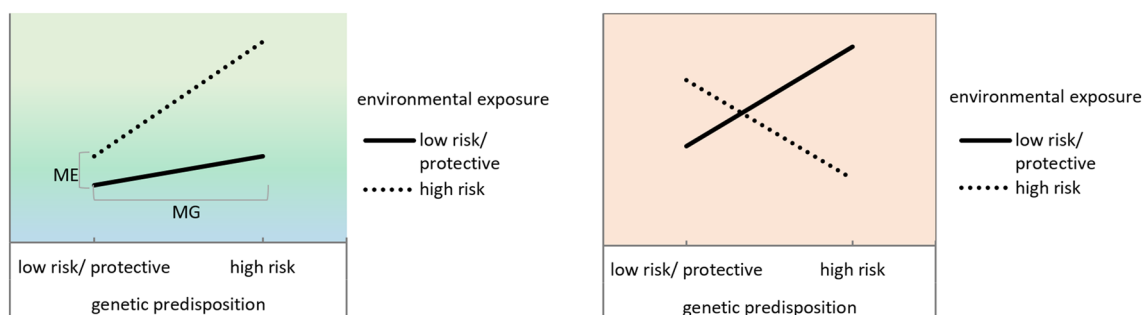


Fig. 3 General pattern of G×E. **a** Genetic factors and environmental factors reinforce each other (green–blue shades in Tables 2, 3, and 4, *N* = 21). ME represents the main effect of environmental expo-

sure, MG that of the genetic factor. **b** The effect of a genetic factor is reversed as a function of an environmental factor (or vice versa; orange findings in Tables 2, 3, and 4, *N* = 2)

rather than formulating hypotheses beforehand. Pre-registering hypotheses might be a good way to overcome these caveats.

As heterogeneity in outcomes, genetic predictors, and environmental factors was substantial, it was difficult to discern patterns in the results. These did not appear to depend on gene group, environmental factor, or substance investigated. However, patterns were hard to discern, as there were many different (combinations of) factors investigated.

Limitations of included studies

Study quality differed within and between genetic risk assessment methods, and was often limited. For all study types, power calculation was mostly omitted, and many studies are likely to have been underpowered (see Table SII). Interaction effects usually require more power to be detected than main effects, and given that main effects of genetic predictors are often small this is especially relevant in G×E studies (Duncan and Keller 2011). For example, if the effect size of a G×E effect would be $R^2 = 0.5\%$, in order to achieve 80% power sample size would need to be $N = 2185$ (assuming 3 predictors and $\alpha = .05$), and 31 out of 39 studies had smaller sample size than that (see Table SII). To put that in perspective, main effects of top SNPs in GWASs are often around 0.25%, and all measured SNPs together typically explain around 10% or less of the variance in substance use phenotypes (So et al. 2011). Also, control for gene–environment correlation was limited, and where it was tested, the test often entailed a simple correlation without controlling for the effects of covariates, or for interaction effects between the genetic predictor and covariates (Keller, 2014). This is problematic, as many environmental factors (such as parenting behaviors) are in fact influenced by genes themselves (Krapohl et al. 2017), and this covariation would decrease chances of detecting G×E and impede its interpretation (Rathouz et al. 2008a, b). Over all study types, heterogeneous designs, lack of replication studies, and inconsistent statistical reporting made assessment of publication bias impossible. In candidate-gene G×E studies such bias to underreporting negative results has been demonstrated (Duncan and Keller 2011). As the number of p values just below the significance threshold was not higher than the number of small p values, we concluded that p hacking did not seem to be an issue.

The haplotype method was limited as a measure of polygenic risk because many studies looked at a few variants in one gene, which is strictly speaking not ‘polygenic’ (but ‘polyvariant’) and will not capture much variation. The investigated genes were mostly plausible candidates for substance use because of their biological function. The benefits of this method compared to the traditional single candidate-gene method are modest. This might be reflected in the results, as more haplotype than other studies did not

find a G×E interaction or found results that are difficult to interpret (i.e., did not follow a pattern as depicted in Fig. 3).

Studies using candidate-gene score methods appeared of somewhat better quality than the haplotype studies. Sample sizes and the number of investigated variants still seem (too) low to detect small effects (Luan et al. 2001). A more fundamental drawback of the candidate-gene method in general is that the selection of variants and risk alleles by definition has to rely on a limited body of knowledge, that might or might not include information on the causally most important genetic variants (Zhu and Zhao 2007). As an example, previous research has shown that candidate-genes for schizophrenia did not predict schizophrenia better than candidate-genes for an unrelated phenotype (diabetes; Johnson et al. 2017). Indeed, few of the proposed candidates in haplotype or candidate-gene score studies have actually been identified in hypothesis-free GWASs (e.g., Liu et al. 2019; Pisman et al. 2018; Walters et al. 2018). This might have added to the finding that studies using these methods more often yielded unexpected patterns.

Technical advances and decreasing costs have made it possible to consider the whole genome for risk prediction, and studies using such PS seem to become more popular than those using haplotype and candidate-gene methods. PS studies yielded the highest quality ratings, with sample sizes more adequate to capture small effects, although study design and phenotype measurements were less strong. It is important to note that PS studies are not necessarily appropriate for testing differential vulnerability hypotheses, as SNPs that operate through that mechanism would not necessarily have large main effects likely to be detected in a GWAS (Fox and Beevers 2016). It is an interesting possibility that different variants are important for interaction effects than for main effects, and this might contribute to the fact that G×E studies show disappointing results in comparison with GWASs. Furthermore, even the qualitatively better studies reviewed here show only small effects.

Recommendations for future studies

Following from the limitations of the included studies, important recommendations for future research can be made. A roadmap for future research is summarized in Fig. 4. First, more attention should be given to hypothesis selection. Although addiction research would be advanced by a further expansion of the scope of research, direct replication attempts might be even more important at this stage (Duncan and Keller 2011). Replication and original studies alike should focus on formulating and pre-registering sharp predictions and give attention to the exact direction of the G×E effects (Belsky et al. 2013; Munafò et al. 2017).

Fig. 4 Road map for future studies with recommended steps for improving the stance of the substance use G×E literature



Second, high quality study methods should be used. Despite having limitations, the PS studies yielded more consistent results than the other two study types. GWASs with substantial sample sizes for alcohol ($N \approx 941,000$) and tobacco use ($N \approx 1,232,000$; Liu et al. 2019), and cannabis use (Pasman et al., 2018; $N \approx 184,000$) are increasingly available, enhancing the predictive power of PS. It is interesting to note that studies are emerging testing SNP by environment interactions in GWASs, making it possible to explore G×E in a hypothesis-free manner. This would circumvent the difficulty that SNPs captured in GWASs do not necessarily measure differential susceptibility. For instance, Polimanti et al. (2017) showed a SNP by trauma exposure interaction on the risk of alcohol misuse. As the multiple testing burden for this kind of design is substantial, large sample sizes are needed to test G×E on a genome-wide level. However, as these samples are becoming increasingly available, the merits of this method might be further explored. Other important characteristics of high quality study methods include using better phenotypical measures, using large discovery and target sample sizes, controlling

for covariates and taking into account possible rGE. Authors should report on rGE analysis (that controlled for covariates), and ideally the G×E analysis should control for the effects (for example using structural equation modeling).

Third, future studies should report more completely and transparently on statistics, such as effect size and achieved power level. Also, more attention should be given to null results, so that in future meta-analyses unbiased effect sizes can be estimated (see Fig. 4).

Strengths and limitations

This is the first review focusing on and comparing multiple polygenic methods for assessing G×E in multiple substance use outcomes. Patterns of results could be compared across different methods, outcomes, and predictors. The quality assessment provided insight in important lacunas in study methodology and gave some suggestion that study quality influences the patterns of results.

The heterogeneity of the included studies introduced some important constraints for the review. No meta-analysis could be attempted and we had to devise our own method to visualize study quality. As we tried to integrate all findings in one comprehensive interpretation, some detail was inevitably lost. Another limitation lies in the fact that some of the studies did not set out to test G×E, but rather included it as a secondary analysis. This might have contributed to the fact that details on methods and results could sometimes be retrieved only with difficulty. Also, it may have biased results, as studies in some cases seemed to test an interaction with a variable that proved to have a main effect, rather than a G×E effect predicted based on the literature.

Conclusion

The current review summarized literature investigating if environmental and polygenic factors interact in influencing alcohol, tobacco, and cannabis use phenotypes. There are important limitations to the literature, concerning overall study quality, failure to formulate directional hypotheses, inconsistent reports of statistics (effect sizes), and a great lack of replication studies. It is likely that some publication bias exists.

Because of these limitations, it is difficult to draw conclusions about the existence of G×E effects in substance use. Before any substantive claims can be made, it is crucial that some steps are undertaken, such as using more sophisticated methods and direct replication attempts of G×E findings. Although still weak, there is some evidence that polygenic G×E effects are a factor in the etiology of substance use, with PS being the best measure of polygenic risk. Studies suggest that environmental factors can influence the effect of genetic predisposition, either by enlarging its (positive or negative) effects, or by reversing those. Additional work is needed before firm conclusions can be drawn about the importance of G×E in the etiology of substance use.

G×E research has the potential to give crucial insight in biopsychosocial mechanisms underlying substance use that might be leveraged for clinical applications. For example, polygenic scores (Musci et al. 2015a, b) and even single genetic variants in the nicotinic receptor genes (Sarginson et al. 2011) can predict who will respond favorably to smoking interventions. In the future, well-conducted G×E studies have the potential to improve possibilities for clinical applications.

Funding JAP and JMV were supported by ERC starting Grant 284167: ‘Beyond the Genetics of Addiction’ (principal investigator JMV). KJHV is supported by the Foundation Volksbond Rotterdam.

Compliance with ethical standards

Conflict of interest Joëlle A. Pasman, Karin J. H. Verweij, and Jacqueline M. Vink declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Informed consent was obtained from all participants included in the study.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Belsky J, Pluess M (2009) Beyond diathesis stress: differential susceptibility to environmental influences. *Psychol Bull* 135(6):885–908
- Belsky J, Pluess M, Widaman KF (2013) Confirmatory and competitive evaluation of alternative gene-environment interaction hypotheses. *J Child Psychol Psychiatry* 54(10):1135–1143
- Berrettini WH, Wileyto EP, Epstein L, Restine S, Hawk L, Shields P, Niaura R, Lerman C (2007) Catechol-O-methyltransferase (COMT) gene variants predict response to bupropion therapy for tobacco dependence. *Biol Psychiatry* 61(1):111–118
- Boomsma D, Busjahn A, Peltonen L (2002) Classical twin studies and beyond. *Nat Rev Genet* 3(11):872
- Bountress K, Chassin L, Lemery-Chalfant K (2016) Parent and peer influences on emerging adult substance use disorder: a genetically informed study. *Dev Psychopathol* 29(1):1–22
- Boyce WT, Ellis BJ (2005) Biological sensitivity to context: I. An evolutionary-developmental theory of the origins and functions of stress reactivity. *Dev Psychopathol* 17(02):271–301
- Brody GH, Yf Chen, Beach SR (2013) Differential susceptibility to prevention: GABAergic, dopaminergic, and multilocus effects. *J Child Psychol Psychiatry* 54(8):863–871
- Chatterjee N, Wheeler B, Sampson J, Hartge P, Chanock SJ, Park J-H (2013) Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nat Genet* 45(4):400–405
- Chen L-S, Baker TB, Piper ME, Breslau N, Cannon DS, Doheny KF, Gogarten SM, Johnson EO, Saccone NL, Wang JC (2012) Interplay of genetic risk factors (CHRNA5-CHRNA3-CHRNA4) and cessation treatments in smoking cessation success. *Am J Psychiatry* 169(7):735–742
- Coley RL, Sims J, Carrano J (2017) Environmental risks outweigh dopaminergic genetic risks for alcohol use and abuse from adolescence through early adulthood. *Drug Alcohol Depend* 175:106–118
- David SP, Strong DR, Leventhal AM, Lancaster MA, McGeary JE, Munafò MR, Bergen AW, Swan GE, Benowitz NL, Tyndale RF (2013) Influence of a dopamine pathway additive genetic efficacy score on smoking cessation: results from two randomized clinical trials of bupropion. *Addiction* 108(12):2202–2211
- Dick DM (2011) An interdisciplinary approach to studying gene-environment interactions: from twin studies to gene identification and back. *Res Hum Dev* 8(3–4):211–226

- Do EK, Maes HH (2016) Genotype \times environment interaction in smoking behaviors: a systematic review. *Nicotine Tobacco Res* 19(4):387–400
- Domingue BW, Conley D, Fletcher J, Boardman JD (2016) Cohort effects in the genetic influence on smoking. *Behav Genet* 46(1):31–42
- Ducci F, Goldman D (2012) The genetic basis of addictive disorders. *Psychiatr Clin N Am* 35(2):495–519
- Ducci F, Enoch M, Hodgkinson C, Xu K, Catena M, Robin R, Goldman D (2008) Interaction between a functional MAOA locus and childhood sexual abuse predicts alcoholism and antisocial personality disorder in adult women. *Mol Psychiatry* 13(3):334–347
- Dudbridge F (2013) Power and predictive accuracy of polygenic scores. *PLoS Genet* 9(3):e1003348
- Duncan LE, Keller MC (2011) A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *Am J Psychiatry* 168(10):1041–1049
- Enoch M-A, Hodgkinson CA, Yuan Q, Shen P-H, Goldman D, Roy A (2010) The influence of GABRA2, childhood trauma, and their interaction on alcohol, heroin, and cocaine dependence. *Biol Psychiatry* 67(1):20–27
- European Monitoring Centre for Drugs and Drug Addiction (2017) European drug report 2017: Trends and developments. Publications Office of the European Union, Luxembourg
- Fox E, Beevers CG (2016) Differential sensitivity to the environment: contribution of cognitive biases and genes to psychological well-being. *Mol Psychiatry* 21(12):1657
- Guo G, Li Y, Wang HY, Cai TJ, Duncan GJ (2015) Peer influence, genetic propensity, and binge drinking: a natural experiment and a replication. *Am J Sociol* 121(3):914–954
- Handley ED, Rogosch FA, Cicchetti D (2015) Developmental pathways from child maltreatment to adolescent marijuana dependence: examining moderation by FK506 binding protein 5 gene (FKBP5). *Dev Psychopathol* 27(4pt2):1489–1502
- Handley ED, Rogosch FA, Cicchetti D (2017) From child maltreatment to emerging adult problem drinking: identification of a multilevel internalizing pathway among African American youth. *Dev Psychopathol* 29(5):1807–1821
- Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* 33:245–254
- Johnson EC, Border R, Melroy-Greif WE, de Leeuw CA, Ehringer MA, Keller MC (2017) No evidence that schizophrenia candidate genes are more associated with schizophrenia than noncandidate genes. *Biol Psychiatry* 82(10):702–708
- Kaufman J, Yang BZ, Douglas-Palumberi H, Crouse-Artus M, Lipschitz D, Krystal JH, Gelernter J (2007) Genetic and environmental predictors of early alcohol use. *Biol Psychiatry* 61(11):1228–1234
- Keller MC (2014) Gene \times environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. *Biol Psychiatry* 75(1):18–24
- Kendler KS, Schmitt E, Aggen SH, Prescott CA (2008) Genetic and environmental influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to middle adulthood. *Arch Gen Psychiatry* 65(6):9
- Kranzler HR, Feinn R, Nelson EC, Covault J, Anton RF, Farrer L, Gelernter J (2011) A CRHR1 haplotype moderates the effect of adverse childhood experiences on lifetime risk of major depressive episode in African-American women. *Am J Med Genet B* 156(8):960–968
- Krapohl E, Hannigan L, Pingault J-B, Patel H, Kadeva N, Curtis C, Breen G, Newhouse S, Eley T, O'Reilly P (2017) Widespread covariation of early environmental exposures and trait-associated polygenic variation. *PNAS* 114(44):11727–11732
- Lerer E, Kanyas K, Karni O, Ebstein R, Lerer B (2006) Why do young women smoke? II. Role of traumatic life experience, psychological characteristics and serotonergic genes. *Mol Psychiatry* 11(8):771–781
- Li JJ, Cho SB, Salvatore JE, Edenberg HJ, Agrawal A, Chorlian DB, Porjesz B, Hesselbrock V, Dick DM (2017) The impact of peer substance use and polygenic risk on trajectories of heavy episodic drinking across adolescence and emerging adulthood. *Alcohol Clin Exp Res* 41(1):65–75
- Liu M, Jiang Y, Wedow R et al (2019) Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* 51(2):237–244
- Luan J, Wong M, Day N, Wareham N (2001) Sample size determination for studies of gene-environment interaction. *Int J Epidemiol* 30(5):1035–1040
- McGeary JE, Knopik VS, Hayes JE, Palmer RH, Monti PM, Kalman D (2012) Predictors of relapse in a bupropion trial for smoking cessation in recently-abstinent alcoholics: preliminary results using an aggregate genetic risk score. *Subst Abuse* 6:107–114
- Meyers JL, Cerda M, Galea S, Keyes KM, Aiello AE, Uddin M, Wildman DE, Koenen KC (2013) Interaction between polygenic risk for cigarette use and environmental exposures in the Detroit Neighborhood Health Study. *Transl Psychiatry* 3:e290
- Mies GW, Verweij KJ, Treur JL, Ligthart L, Fedko IO, Hottenga JJ, Willemsen G, Bartels M, Boomsma DI, Vink JM (2018) Polygenic risk for alcohol consumption and its association with alcohol-related phenotypes: do stress and life satisfaction moderate these relationships? *Drug Alcohol Depend*. <https://doi.org/10.1016/j.drugalcdep.2017.10.018>
- Milaniak I, Watson B, Jaffee SR (2015) Gene-environment interplay and substance use: a review of recent findings. *Curr Addict Rep* 2(4):364–371
- Moher D, Liberati A, Tetzlaff J, Altman DG, Group P (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 6(7):e1000097
- Monroe SM, Simons AD (1991) Diathesis-stress theories in the context of life stress research: implications for the depressive disorders. *Psychol Bull* 110(3):406
- Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N, Simonsohn U, Wagenmakers E-J, Ware JJ, Ioannidis JPA (2017) A manifesto for reproducible science. *Nat Hum Behav* 1:0021
- Musci RJ, Masyn KE, Uhl G, Maher B, Kellam SG, Ialongo NS (2015a) Polygenic score \times intervention moderation: an application of discrete-time survival analysis to modeling the timing of first tobacco use among urban youth. *Dev Psychopathol* 27(01):111–122
- Musci RJ, Uhl G, Maher B, Ialongo NS (2015b) Testing gene \times environment moderation of tobacco and marijuana use trajectories in adolescence and young adulthood. *J Consult Clin Psychol* 83(5):866–874
- Musci RJ, Fairman B, Masyn KE, Uhl G, Maher B, Sisto DY, Kellam SG, Ialongo NS (2016) Polygenic score \times intervention moderation: an application of discrete-time survival analysis to model the timing of first marijuana use among urban youth. *Prev Sci* 19(1):1–9
- Nelson EC, Agrawal A, Pergadia ML, Wang JC, Whitfield JB, Saccone FS, Kern J, Grant JD, Schrage AJ, Rice JP, Montgomery GW, Heath AC, Goate AM, Martin NG, Madden PAF (2010) H2 haplotype at chromosome 17q21.31 protects against childhood sexual abuse-associated risk for alcohol consumption and dependence. *Addict Biol* 15(1):1–11
- Oroszi G, Anton RF, O'Malley S, Swift R, Pettinati H, Couper D, Yuan Q, Goldman D (2009) OPRM1 Asn40Asp predicts response to naltrexone treatment: a haplotype-based approach. *Alcohol Clin Exp Res* 33(3):383–393

- Pasman JA, Verweij KJ, Treur JL, Willemsen G, Hottenga J-J, Homburg JR, Spijker S, Otten R, Engels RC, de Geus EJ (2017) Hypo-dopaminergic polygenic risk for polysubstance use. *J Psychol Cogn* 2(2):138–148
- Pasman JA, Verweij KJH et al (2018) GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nat Neurosci* 21(9):1161–1170
- Polimanti R, Kaufman J, Zhao H, Kranzler HR, Ursano RJ, Kessler RC, Gelernter J, Stein MB (2017) A genome-wide gene-by-trauma interaction study of alcohol misuse in two independent cohorts identifies PRKG1 as a risk locus. *Mol Psychiatry* 23(1):154–160
- Polimanti R, Kaufman J, Zhao H, Kranzler H, Ursano R, Kessler R, Stein M, Gelernter J (2018) Trauma exposure interacts with the genetic risk of bipolar disorder in alcohol misuse of US soldiers. *Acta Psychiatr Scand* 137(2):148–156
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38(8):904–909
- Rathouz PJ, Van Hulle CA, Rodgers JL, Waldman ID, Lahey BB (2008a) Specification, testing, and interpretation of gene-by-measured-environment interaction models in the presence of gene–environment correlation. *Behav Genet* 38(3):301–315
- Rathouz PJ, Van Hulle CA, Rodgers JL, Waldman ID, Lahey BB (2008b) Specification, testing, and interpretation of gene-by-measured-environment interaction models in the presence of gene–environment correlation. *Behav Genet* 38(3):301–315
- Ray LA, Sehl M, Bujarski S, Hutchison K, Blaine S, Enoch MA (2013) The CRHR1 gene, trauma exposure, and alcoholism risk: a test of G x E effects. *Genes Brain Behav* 12(4):361–369
- Salvatore JE, Aliev F, Edwards AC, Evans DM, Macleod J, Hickman M, Lewis G, Kendler KS, Loukola A, Korhonen T, Latvala A, Rose RJ, Kaprio J, Dick DM (2014) Polygenic scores predict alcohol problems in an independent sample and show moderation by the environment. *Genes* 5(2):330–346
- Salvatore JE, Savage JE, Barr P, Wolen AR, Aliev F, Vuoksimaa E, Latvala A, Pulkkinen L, Rose RJ, Kaprio J (2018) Incorporating functional genomic information to enhance polygenic signal and identify variants involved in gene-by-environment interaction for young adult alcohol problems. *Alcohol Clin Exp Res* 42(2):413–423
- Sarginson JE, Killen JD, Lazzeroni LC, Fortmann SP, Ryan HS, Schatzberg AF, Murphy GM (2011) Markers in the 15q24 nicotinic receptor subunit gene cluster (CHRNA5-A3-B4) predict severity of nicotine addiction and response to smoking cessation therapy. *Am J Med Genet Neuropsychiat Genet* 156(3):275–284
- Schmitz L, Conley D (2016) The long-term consequences of Vietnam-era conscription and genotype on smoking behavior and health. *Behav Genet* 46(1):43–58
- Segman RH, Kanyas K, Karni O, Lerer E, Goltser-Dubner T, Pavlov V, Lerer B (2007) Why do young women smoke? IV. Role of genetic variation in the dopamine transporter and lifetime traumatic experience. *Am J Med Genet B* 144(4):533–540
- Simonsohn U, Nelson LD, Simmons JP (2014) P-curve: a key to the file-drawer. *J Exp Psychol Gen* 143(2):534
- So HC, Gui AH, Cherny SS, Sham PC (2011) Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet Epidemiol* 35(5):310–317
- Stogner JM, Gibson CL (2016) Genetic modification of the relationship between parental rejection and adolescent alcohol use. *Alcohol Alcohol* 51(4):442–449
- The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447(7145):661–678
- Treur JL, Verweij KJ, Abdellaoui A, Fedko IO, de Zeeuw EL, Ehli EA, Davies GE, Hottenga J-J, Willemsen G, Boomsma DI (2017) Testing familial transmission of smoking with two different research designs. *Nicotine Tobacco Res:ntx* 121
- Tyndale RF, Zhu AZ, George TP, Cinciripini P, Hawk LW Jr, Schnoll RA, Swan GE, Benowitz NL, Heitjan DF, Lerman C (2015) Lack of associations of CHRNA5-A3-B4 genetic variants with smoking cessation treatment outcomes in caucasian smokers despite associations with baseline smoking. *PLoS ONE* 10(5):e0128109
- U.S. Department of Health and Human Services (2016) Results from the 2016 National Survey on Drug Use and Health: Detailed Tables. <https://www.samhsa.gov/data/sites/default/files/NSDUH-DefTabs-2016/NSDUH-DefTabs-2016.htm#tab1-1B>
- Vink JM (2016) Zit middelengebruik en verslavingsgedrag in de familie? *Verslaving* 12(4):243–255
- Vrieze SI, McGue M, Iacono WG (2012) The interplay of genes and adolescent development in substance use disorders: leveraging findings from GWAS meta-analyses to test developmental hypotheses about nicotine consumption. *Hum Genet* 131(6):18
- Walters RK, Polimanti R et al (2018) Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat Neurosci* 21(12):1656–1669
- World Health Organization (2014) Global status report on alcohol and health. http://apps.who.int/iris/bitstream/10665/112736/1/9789240692763_eng.pdf
- World Health Organization (2016) Prevalence of tobacco smoking. <http://www.who.int/gho/tobacco/use/en/>
- Wray NR, Lee SH, Mehta D, Vinkhuyzen AA, Dudbridge F, Middeldorp CM (2014) Research review: polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatry* 55(10):1068–1087
- Zhu M, Zhao S (2007) Candidate gene identification approach: progress and challenges. *Int J Biol Sci* 3(7):420–427

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.