GROWUPGROWOLD | HUMAN DEVELOPMENT, BIRTH TO DEATH

Moffitt & Caspi | Genes Environment Health Behavior

Concept Paper Form

Provisional Paper Title: The Externalizing and Intergenerational Transmission (EXIT) Project: Identifying genetic and environmental pathways for the development of externalizing behavior using within-family genomic data
Proposing Author: Peter T. Tanksley (with support from Jasmin Wertz)
Author's Email: peter.tanksley@austin.utexas.edu
P.I. Sponsor: Click or tap here to enter text. (if the proposing author is a student or colleague of an original PI)
Today's Date: Click here to enter date

Objective of the Study:

The overall aim of this multi-cohort study is to follow-up results from a recent genome-wide association study of externalizing problems (Linnér & Mallard et al., in press) to examine the role of "direct" and "indirect" genetics effects on the development of externalizing problems. "Direct" genetic effects refer to associations between an individual's own genes and their externalizing problems. Indirect genetic effects (also referred to as "genetic nurture"; Kong et al., 2018) refers to associations between the genes of an individual's parents and the individual's externalizing problems, mediated, for example, by parental behavior. The proposed study will analyze data from multiple cohorts (including E-Risk) that contain genomic data of family members, to examine direct and indirect genetic effects. The data will be analyzed by Peter T. Tanksley (University of Texas at Austin) with support from Jasmin Wertz (Duke University). The study goals are:

Goal I To estimate the extent to which associations between a new polygenic score for externalizing problems (i.e., based on Linner & Mallard et al., in press) and externalizing problems reflect direct genetic effects vs. indirect genetic effects (vs. confounders).

Goal II To compare estimates of direct and indirect genetic effects across developmental periods to test whether the relative strength of direct genetic effects on externalizing problems changes over the early life course.

Goal III To integrate measures of family socioeconomic advantage/disadvantage to probe the extent to which indirect genetic effects on externalizing are entangled with socioeconomic stratification.

Data Analysis Methods:

For our analyses, we will only use the DZ twins because they share all contextual factors but remain distinguishable at the genetic level (i.e., unlike MZ twins). This allows for the use of a fixed effects framework that adjusts for familial confounding (i.e., indirect genetic effects of parents) while

leaving some variation at the genetic level to predict disparities in externalizing outcomes between DZ twins (i.e., direct genetic effects). For the E-Risk analysis, the externalizing polygenic scores of DZ twins will be used to predict their own externalizing behavior using a mixed-effects model with a random intercept representing heterogeneity between families (Selzam et al., 2019):

$$EXT_{ij} = \alpha_0 + \beta_{within}\left(PGS_{ij} - \overline{PGS_j}\right) + \beta_{Between}\left(\overline{PGS_j}\right) + \gamma_j + \varepsilon_{ij}$$

Where PGS_{ij} is the polygenic score value for individual *i* in family *j*, $\overline{PGS_j}$ is the family-specific average polygenic score value within family *j*, and Y_j is the family-specific random intercept. This approach decomposes the genetic association with externalizing problems into a within-family component, β_{within}, representing the direct genetic effect, versus a between-family component, $\beta_{Between}$, representing the indirect genetic effect. The baseline model will include sex and any cohort-specific technical covariates (e.g., batch, array) (Note: age-heterogeneous samples also include age and age-squared). Next, parent SES and parent externalizing problems will be added as covariates to test whether indirect genetic effects are mediated by these variables.

Results from the E-Risk Study will be aggregated with the other participating cohorts using a random-effects meta-analysis approach, following Demange et al., 2020. The participating cohorts include the National Longitudinal Study of Adolescent to Adult Health (Add Health), Collaborative Studies on Genetics of Alcoholism (COGA), the German Socioeconomic Panel Innovation Sample (SOEP-IS), the Millennium Cohort Study (MCS), Texas Twins Project (TTP), and the Quebec Newborn Twin Study (QNTS).

Measurement of Externalizing Behavior

Externalizing behavior is varied in its expression across the life course. For the sake of consistency across time and across cohorts, we adopt a two-part strategy to measure externalizing by: (1) construct a measure of total externalizing symptom burden for children/adolescents and (2) extract the first principal component of externalizing behaviors more broadly defined for adults.

Symptom burden score

The participating cohorts use a variety of validated instruments to measure externalizing (e.g., Strengths and Difficulties Questionnaire [Goodman, 1997], Child Behavioral Checklist [Achenbach & Edelbrock, 1991]); however, not all cohorts use the same measures. To make estimates more interpretable across cohorts, we will create total symptom burden scores by (1) summing the number of acknowledged externalizing symptoms across the validated measures used within a cohort and then (2) standardizing these symptom counts with z-scores (M=0; SD=1) within each cohort. In E-Risk, we would use the established CBCL total externalizing scores (as reported by parent/teacher) and z-standardize.

Principal components analysis

We will use principal components analysis (PCA) to assess externalizing behavior in adults (i.e., parents and adult offspring). Measurement of externalizing behavior is more direct in youths than in adults, making it necessary to leverage every source of information on adult externalizing

behaviors/traits available in each cohort. These may include elements of personality (e.g., [lack of] conscientiousness, [lack of] agreeableness), psychopathologies (e.g., ADHD symptoms), social behavior (e.g., rule-breaking, aggression, crime), risky-taking, and substance use (e.g., alcohol abuse, illegal drug use). This strategy mirrors the approach adopted by the original GWAS of externalizing behavior that included summary statistics from GWAS of ADHD (Demontis et al., 2018), problematic alcohol use (Sanchez-Roige et al., 2018; Walters et al., 2018), lifetime cannabis use (Pasman et al., 2018), age at first sexual intercourse, number of sexual partners, general risk tolerance (Linnér et al., 2019), and lifetime smoking initiation (Liu et al., 2019).

We will conduct PCA on the available information in each cohort and extract the first principal component (PC) as our measure of adult externalizing behavior. For adults with multiple observations, scores on each measure will be averaged across observations, and these average scores will be entered into the PCA. The resulting PC will thus be interpretable as a measure of time-stable adult externalizing behaviour.

Developmental Epochs

We will leverage the longitudinal nature of the participating cohorts to explore changes in the direct and indirect genetic effects of the externalizing PGI across development (Goal II). We accomplish this by dividing observations of externalizing behaviors within each cohort into four developmental epochs: (1) childhood (<5 years), (2) pre-adolescence (5-10 years), (3) adolescence (11-17 years), and (4) adulthood (18+ years). For the E-Risk cohort, this means that observations from Phases 5, 7, and 10 would be aggregated together as “pre-adolescence”, Phase 12 observations would be “adolescence”, and Phase 18 would be “adult”. To aggregate externalizing scores for the “pre-adolescence” externalizing score, we will average externalizing behavior scores within Phases 5, 7, and 10, and then these composite scores will be averaged across ages.

Polygenic Score for Externalizing Behavior

We used a unified analytic pipeline to construct the externalizing polygenic score in European ancestry individuals in each cohort. The pipeline relies on two software packages: PRS-CS (Ge et al., 2019), for adjusting original GWAS beta weights for linkage disequilibrium (LD), and Plink2 (Chang et al., 2015), for constructing the externalizing polygenic score from LD-adjusted beta weights.

Linkage disequilibrium

Prior to polygenic score construction, we performed LD-adjustment of the original GWAS beta weights, as modeling LD between SNPs is known to increase the signal-to-noise ratio in polygenic scores. We applied a recently developed method called “PRS-CS” (the October 20, 2019 software release) to adjust for LD. As the reference panel for estimating LD, we used the 1000 Genomes European reference files distributed with the software. Also, as the PRS-CS method is currently restricted to the ~1.3 million SNPs in the high-quality consensus genotype set defined by the HapMap 3 Consortium (Altshuler et al., 2010), for comparability, we only generated polygenic scores using HapMap 3 SNPs. All other parameters will be set to the default parameters of the PRC-CS software.

Polygenic scores for externalizing behavior

Plink2 will be used to compute polygenic score in individuals solely of recent European ancestries as estimated by their DNA in each cohort using the LD-adjusted beta weights. Polygenic scores are computed as the weighted sum of the effect-coded alleles for a given individual *i*:

$$PGS_i = \Sigma\, \beta_j(g_{ij})$$

where PGS_i is the polygenic score, β_j is the estimated additive effect of the effect-coded allele at SNP *j*, and g_{ij} is the genotype at SNP *j*.

Random Effects Meta-Analysis of Cohort-Specific Effects

Results from the E-Risk sample will be aggregated with other cohorts using a random-effects meta-analysis, following Demange et al., 2020. Our aim is for each cohort to yield the following estimates for each project goal:

Goal I 2 estimates (i.e., direct/indirect genetic effects).
Goal II 2 estimates × k number of epochs covered by the cohort.
Goal III 2 estimates × k number of epochs covered by the cohort × 3 estimates adjusting for (i) parental externalizing, (ii) parental socioeconomic status, and (iii) both together.

Variables Needed at which Ages:

PHASE	VARIABLE LABEL	VARIABLE DESCRIPTION (INFORMANT)
E-RISK PHASE CROSS-PHASE OR NO SPECIFIC PHASE	FAMILYID	FAMILY ID
	ATWINID	TWIN ID
	SAMPSEX	SEX OF CHILD
	SESW	COMPOSITE OF PARENTAL SES (MOTHER/FATHER)
	SESWQ35	TERTILES OF PARENTAL SES (SESW)
	RORDERP5	RANDOM TWIN ORDER
	Zygosity	Zygosity (please add most up-to-date)
E-RISK PHASE 5	TOTEXTE5	CBCL, AGGRESSION AND RULE-BREAKING SCALES (PARENT/TEACHER)
	ASBMM5	ANTISOCIAL BEHAVIOR (MOTHER)
	ASBFM5	ANTISOCIAL BEHAVIOR (FATHER)
		BIG FIVE: CONSCIENTIOUSNESS (MOTHER)
		BIG FIVE: CONSCIENTIOUSNESS (FATHER)
		BIG FIVE: AGREEABLENESS (MOTHER)
		BIG FIVE: AGREEABLENESS (FATHER)
E-RISK PHASE 7	TOTEXTE7	CBCL, AGGRESSION AND RULE-BREAKING SCALES (PARENT/TEACHER)
		PARENTAL ANTISOCIAL BEHAVIOR (MOTHER/FATHER)
E-RISK PHASE 10	TOTEXTE10	CBCL, AGGRESSION AND RULE-BREAKING SCALES (PARENT/TEACHER)
E-RISK PHASE 12	TOTEXTE12	CBCL, AGGRESSION AND RULE-BREAKING SCALES (PARENT/TEACHER)
E-RISK PHASE 18	RISKYSEX18	SEXUAL RISK TAKING (PARTICIPANT) – based on Wertz et al., JAACAP 2020
	CRIMCNTE18	OFFICIAL CRIME RECORDS
	ALCVOLE18	ALCOHOL DRINKING
	TBC	CONDUCT PROBLEMS (CO-INFORMANTS) – not derived yet?
	ZBFIC_COMBE18	CONSCIENTIOUSNESS (PARTICIPANT, CO-INFORMANTS)
	ZBFIA_COMBE18	AGREEABLENESS (PARTICIPANT, CO-INFORMANTS)
	CDSXE18	CONDUCT PROBLEMS (PARTICIPANT)

Significance of the Study (for theory, research methods or clinical practice):

Externalizing problems are a constellation of personality traits, behaviors, and mental health conditions characterized by difficulties with self-regulation and violations of social and/or legal norms (Barr & Dick, 2019). Externalizing problems are costly to both individuals and society (Richmond-Rakerd et al., 2020). In addition to being highly heritable (Krueger et al., 2002), externalizing problems are socially stratified: children raised in conditions of socioeconomic disadvantage show higher rates of externalizing behavior problems, and externalizing behavior exerts downward pressure on an individual's social mobility (Dodge et al., 1994; Miech et al., 1999). The development of externalizing problems thus involves entangled biological and social processes.

We recently conducted a large-scale GWAS of externalizing behavior in 1.5 million people and identified over 500 independent genetic loci that were significantly associated with a general factor of externalizing problems (Linnér & Mallard et al., 2021). A polygenic score constructed from these GWAS results was associated with an array of socially important phenotypes, including opioid and other substance use, employment histories, and contact with the criminal justice system. The variance explained by this polygenic score for externalizing behavior (Rsquared=8.9-10.5%) mirrored the variance explained by typical social science variables such as parental socioeconomic status, family income/structure, and neighborhood disadvantage/disorder.

Observed genetic associations with externalizing problems could reflect a number of processes, including (a) direct genetic effects; (b) indirect genetic effects (sometimes referred to as cial etiology of externalizing problems.

The direct genetic effect is the average expected change in the phenotype given a change in the individual's own genotype. "Direct" in this case refers to the fact that the genetic association originates in the individual's own genotype (rather than another person's genotype or confounding). Direct genetic effects might still be mediated and/or moderated by the physical or social environment and should not be interpreted as deterministic or as involving only strictly biological processes. A person's genotype might influence their externalizing problems in multiple ways, e.g. by influencing brain development in ways that make externalizing problems more likely; by affecting early-emerging behavioral risk factors for externalizing problems (such as lower cognitive and self-control skills) and so forth; or by making it more likely that a child selects a peer group with higher involvement in risky and thrill-seeking behaviors, and those peers socialize the child toward higher delinquency (Jencks, 1972; Scarr & McCartney, 1983).

In contrast, the indirect genetic effect is the portion of a genetic association that is due to the effects of environments provided by genetic relatives. It originates in the genotype of another individual, rather than in the genotype of the individual themselves (Koellinger & Harden, 2018; Kong et al., 2018). For example, parents who have a genetic risk for externalizing behaviors might be more likely to live in poverty, which in turn increases risk for offspring externalizing regardless of which genes those offspring inherit. Previous genetic studies of other social/behavioral phenotypes, such as educational attainment, have found that indirect genetic effects can account for up to half of observed genetic associations (Kong et al., 2018). Additionally, previous research in twins suggests that indirect genetic effects might be more consequential for externalizing at earlier stages of development compared to later in the life course (Branigan et al., 2013; Harden et al., 2015; Krueger et al., 2002).

Finally, in addition to direct and indirect effects, there are also confounding processes such as assortative mating and population structure that might affect genetic associations with externalizing problems (Young et al., 2019). Deconstructing the various processes that contribute to polygenic associations with externalizing problems can be accomplished using study designs that measure DNA in close family members, such as dizygotic twins or parent-offspring trios. Genomic differences between siblings or between parents and their children are the result of random assortment, thus genotype-phenotype associations observed when comparing these type of close family members capture direct genetic effects (Brumpton et al., 2019). At the same time, the association between parental genetic variants which an offspring does not inherit and the offspring phenotype cannot, by definition, be due to genetic inheritance and must be mediated by environmental processes (Kong et al., 2018).

The project is significant for theory and research, because it uses the first well-powered polygenic score for externalizing behavior. By combining this polygenic score with a within-family design, the EXIT project will be the first to jointly model observed social and genetic factors related to the manifestation of externalizing behavior across development. Studying these processes is significant because it will contribute to a better understanding of how genes and environments contribute to the development and the intergenerational transmission of externalizing problems.

References Cited:

Achenbach, Thomas M., and C. Edelbrock. "Child behavior checklist." *Burlington (Vt)* 7 (1991): 371-392.

Barr, P. B. & Dick, D. M. The Genetics of Externalizing Problems. 1–20 (2019) doi:10.1007/7854_2019_120.

Branigan, A. R., McCallum, K. J. & Freese, J. Variation in the heritability of educational attainment: An international meta-analysis. Soc. Forces 92, 109–140 (2013).

Brumpton, B. et al. Within-family studies for Mendelian randomization: avoiding dynastic, assortative mating, and population stratification biases. bioRxiv 602516 (2019) doi:10.1101/602516.

Demange, Perline A., Jouke Jan Hottenga, Abdel Abdellaoui, Margherita Malanchini, Benjamin W. Domingue, Eveline L. de Zeeuw, Kaili Rimfeld et al. "Parental influences on offspring education: indirect genetic effects of non-cognitive skills." biorxiv (2020).

Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat. Genet. 51, 63–75 (2019).

Dodge, K. A., Pettit, G. S. & Bates, J. E. Socialization Mediators of the Relation between Socioeconomic Status and Child Conduct Problems. Child Dev. 65, 649–665 (1994).

Goodman, Robert. "The Strengths and Difficulties Questionnaire: a research note." Journal of child psychology and psychiatry 38, no. 5 (1997): 581-586.

Harden, K. P. & Koellinger, P. D. Using genetics for social science. Nat. Hum. Behav. 4, 567–576 (2020).

Harden, K. P. et al. Developmental changes in genetic and environmental influences on rule-breaking and aggression: age and pubertal development. J. Child Psychol. Psychiatry 56, 1370–1379 (2015).

Jencks, C. Inequality: A reassessment of the effect of family and schooling in America. (Basic Books, 1972).

Karlsson Linnér, R. et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nat. Genet. 51, 245–257 (2019).

Koellinger, P. D. & Harden, K. P. Using nature to understand nurture. Science 359, 386–387 (2018).

Kong, A. et al. The nature of nurture: Effects of parental genotypes. Science 359, 424–428 (2018).

Krueger, R. F. et al. Etiologic connections among substance dependence, antisocial behavior and personality: Modeling

the externalizing spectrum. J. Abnorm. Psychol. 111, 411 (2002).

Linnér, R. K. et al. Multivariate genomic analysis of 1.5 million people identifies genes related to addiction, antisocial behavior, and health. bioRxiv 2020.10.16.342501 (2020) doi:10.1101/2020.10.16.342501.

Liu, M. et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat. Genet. 51, 237–244 (2019).

Miech, R. A., Caspi, A., Moffitt, T. E., Wright, B. R. E. & Silva, P. A. Low Socioeconomic Status and Mental Disorders: A Longitudinal Study of Selection and Causation during Young Adulthood. Am. J. Sociol. 104, 1096–1131 (1999).

Pasman, J. A. et al. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. Nat. Neurosci. 3 21, 1161–1170 (2018).

Richmond-Rakerd, L. S. et al. Clustering of health, crime and social-welfare inequality in 4 million citizens from two nations. Nat. Hum. Behav. 1–10 (2020) doi:10.1038/s41562-019-0810-4.

Sanchez-Roige, S. et al. Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts. 37 Am. J. Psychiatry 176, 107–118 (2018).

Scarr, S. & McCartney, K. How people make their own environments: a theory of genotype greater than environment effects. Child Dev. 54, 424–435 (1983).

Selzam, Saskia, Stuart J. Ritchie, Jean-Baptiste Pingault, Chandra A. Reynolds, Paul F. O'Reilly, and Robert Plomin. "Comparing within-and between-family polygenic score prediction." The American Journal of Human Genetics 105, no. 2 (2019): 351-363.

Walters, R. K. et al. Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. Nat. Neurosci. 21, 1656–1669 34 (2018).

Young, A. I., Benonisdottir, S., Przeworski, M. & Kong, A. Deconstructing the sources of genotype-phenotype associations in humans. Science 365, 1396–1400 (2019).

Young, Alexander I., Seyed Moeen Nehzati, Chanwook Lee, Stefania Benonisdottir, David Cesarini, Daniel J. Benjamin, Patrick Turley, and Augustine Kong. "Mendelian imputation of parental genotypes for genome-wide estimation of direct and indirect genetic effects." BioRxiv (2020).

Data Security Agreement

Provisional Paper Title: The Externalizing and Intergenerational Transmission (EXIT) Project: Identifying genetic and environmental pathways for the development of externalizing behavior using within-family genomic data
Proposing Author: Peter T. Tanksley (with support from Jasmin Wertz)
Today's Date: Click here to enter date

☒	I am current on Human Subjects Training (CITI (www.citiprogram.org) or equivalent)
☐	My project is covered by the Duke ethics committee OR I have /will obtain ethical approval from my home institution.
☒	I will treat all data as "restricted" and store in a secure fashion. My computer or laptop is: a) encrypted (recommended programs are FileVault2 for Macs, and Bitlocker for Windows machines) b) password-protected c) configured to lock-out after 15 minutes of inactivity AND d) has an antivirus client installed as well as being patched regularly.
☒	I will not "sync" the data to a mobile device.
☒	In the event that my laptop with data on it is lost, stolen or hacked, I will immediately contact Moffitt or Caspi.
☒	I will not share the data with anyone, including my students or other collaborators not specifically listed on this concept paper.
☒	I will not post data online or submit the data file to a journal for them to post. *Some journals are now requesting the data file as part of the manuscript submission process. Study participants have not given informed consent for unrestricted open access, so we have a managed-access process. Speak to Temi or Avshalom for strategies for achieving compliance with data-sharing policies of journals.*
☒	I will delete all data files from my computer after the project is complete. Collaborators and trainees may not take a data file away from the office. This data remains the property of the Study and cannot be used for further analyses without an approved concept paper for new analyses.
☒	I have read the Data Use Guidelines and agree to follow the instructions.

Signature: Peter T. Tanksley