Polygenic Scores in Add Health

David B. Braudt braudt@unc.edu





Polygenic Scores (PGSs): An Overview

 PGSs capture the combined additive genetic influence of SNPs across the entire genome on a specific trait/behavior in a single measure



Molecular genetic data

GWAS summary statistics





Polygenic Scores (PGSs): An Overview

 PGSs capture the combined additive genetic influence of SNPs across the entire genome on a specific trait/behavior in a single measure

$$PGS_i = \sum_{j=1}^{\kappa} \beta_j SNP_{ij}$$
 standarized
within ancestry
groups $\mu_{PGS} = 0$ and $sd = 1$





Polygenic Scores (PGSs): An Overview

 PGSs capture the combined additive genetic influence of SNPs across the entire genome on a specific trait/behavior in a single measure

$$PGS_i = \sum_{j=1}^{\kappa} \beta_j SNP_{ij}$$
 standarized
within ancestry
groups $\mu_{PGS} = 0$ and $sd = 1$

- Whole genome PGSs
 - Include genetic association from across the entire genome, but eliminate the possibility to testing hypotheses related to specific biological pathways (Belsky and Israel, 2014)





- Four possible uses
 - Nature net of Nurture
 - Nurture net of Nature
 - How Nurture modifies the effect(s) of Nature
 - How Nature modifies the effect(s) of Nurture





- Four possible uses
 - Nature net of Nurture
 - Nurture net of Nature
 - How Nurture modifies the effect(s) of Nature
 - How Nature modifies the effect(s) of Nurture





• Four possible uses

- Nature net of Nurture
- Nurture net of Nature (Control for genetic influences... Endogeneity problem)
- How Nurture modifies the effect(s) of Nature
- How Nature modifies the effect(s) of Nurture





• Four possible uses

- Nature net of Nurture
- Nurture net of Nature
- How Nature modifies the effect(s) of Nurture







• Four possible uses

- Nature net of Nurture
- Nurture net of Nature
- How Nurture modifies the effect(s) of Nature







Research Article

The Genetics of Success: How Single-Nucleotide Polymorphisms Associated With Educational Attainment Relate to Life-Course Development

Daniel W. Belsky^{1,2}, Terrie E. Moffitt^{3,4,5,6}, David L. Corcoran⁵, Benjamin Domingue⁷, HonaLee Harrington³, Sean Hogan⁸, Renate Houts³, Sandhya Ramrakha⁸, Karen Sugden³, Benjamin S. Williams³, Richie Poulton⁸, and Avshalom Caspi^{3,4,5,6}



Psychological Science 2016, Vol. 27(7) 957–972 © The Author(s) 2016 Reprints and permissions: sagepub.com/journalsPermissions.nav DOI: 10.1177/0956797616643070 pss.sagepub.com SAGE





Research Article

The Genetics of Success: How Single-Nucleotide Polymorphisms Associated With Educational Attainment Relate to Life-Course Development

Abstract

Daniel W. Belsky^{1,2}, Te Benjamin Domingue⁷, Renate Houts³, Sandhy Benjamin S. Williams³, Avshalom Caspi^{3,4,5,6}

A previous genome-wide association study (GWAS) of more than 100,000 individuals identified molecular-genetic predictors of educational attainment. We undertook in-depth life-course investigation of the polygenic score derived from this GWAS using the four-decade Dunedin Study (*N* = 918). There were five main findings. First, polygenic scores predicted adult economic outcomes even after accounting for educational attainments. Second, genes and environments were correlated: Children with higher polygenic scores were born into better-off homes. Third, children's polygenic scores predicted their adult outcomes even when analyses accounted for their social-class origins; social-mobility analysis showed that children with higher polygenic scores were more upwardly mobile than children with lower scores. Fourth, polygenic scores predicted behavior across the life course, from early acquisition of speech and reading skills through geographic mobility and mate choice and on to financial planning for retirement. Fifth, polygenic-score associations were mediated by psychological characteristics, including intelligence, self-control, and interpersonal skill. Effect sizes were small. Factors connecting DNA sequence with life outcomes may provide targets for interventions to promote population-wide positive development.





Psychological Science 2016, Viol. 27(7) 957–972 O The Author(5) 2016 Repetito and permissions: sagepub.com/journals/remissions.nav DOI: 10.1177/0956797616643070 pss.sagepub.com

Nature net of Nurture

Research Article

The Genetics of Success: How Single-Nucleotide Polymorphisms Associated With Educational Attainment Relate to Life-Course Development

Abstract

Daniel W. Belsky^{1,2}, Te Benjamin Domingue⁷, Renate Houts³, Sandhy Benjamin S. Williams³, Avshalom Caspi^{3,4,5,6}

A previous genome-wide association study (GWAS) of more than 100,000 individuals identified molecular-genetic predictors of educational attainment. We undertook in-depth life-course investigation of the polygenic score derived from this GWAS using the four-decade Dunedin Study (*N* = 918). There were five main findings. First, polygenic scores predicted adult economic outcomes even after accounting for educational attainments. Second, genes and environments were correlated: Children with higher polygenic scores were born into better-off homes. Third, children's polygenic scores predicted their adult outcomes even when analyses accounted for their social-class origins; social-mobility analysis showed that children with higher polygenic scores were more upwardly mobile than children with lower scores Fourth, polygenic scores predicted behavior across the life course, from early acquisition of speech and reading skills through geographic mobility and mate choice and on to financial planning for retirement. Fifth, polygenic-score associations were mediated by psychological characteristics, including intelligence, self-control, and interpersonal skill. Effect sizes were small. Factors connecting DNA sequence with life outcomes may provide targets for interventions to promote population-wide positive development.



Psychological Science 2016, Vol. 27(7) 957-972

0 The Author(s) 2016 Reprints and permissions

SAGE

ka ar publiciony (examals/Permissions, nav

DOI: 10.1177/0956797636643070

The National Longitudinal Study of Adolescent to Adult Health

Gene-Environment

Interactions



Research Article

The Genetics of Success: How Single-Nucleotide Polymorphisms Associated With Educational Attainment Relate to Life-Course Development

Abstract

Daniel W. Belsky^{1,2}, Te Benjamin Domingue⁷, Renate Houts³, Sandhy Benjamin S. Williams³, Avshalom Caspi^{3,4,5,6}

A previous genome-wide association study (GWAS) of more than 100,000 individuals identified molecular-genetic predictors of educational attainment. We undertook in-depth life-course investigation of the polygenic score derived from this GWAS using the four-decade Dunedin Study (*N* = 918). There were five main findings. First, polygenic scores predicted adult economic outcomes even after accounting for educational attainments. Second, genes and environments were correlated: Children with higher polygenic scores were born into better-off homes. Third, children's polygenic scores predicted their adult outcomes even when analyses accounted for their social-class origins; social-mobility analysis showed that children with higher polygenic scores were more upwardly mobile than children with lower scores. Fourth, polygenic scores predicted behavior across the life course, from early acquisition of speech and reading skills through geographic mobility and mate choice and on to financial planning for retirement. Fifth, polygenic-score associations were mediated by psychological characteristics, including intelligence, self-control, and interpersonal skill. Effect sizes were small. Factors connecting DNA sequence with life outcomes may provide targets for interventions to promote population-wide positive development.





Psychological Science 2016, Vol. 27(7) 957–972 O The Author(s) 2016 Reprints and permissions sagepub.com/joarnalsPermissions.nav 1501: 10.1177/0956797616663070 pss.sagepub.com

Gene-Environment Correlations

Add Health Polygenic Scores – Release 1

Cardiovascular Health

- Coronary Artery Disease
- Myocardial Infarction
- Plasma Cortisol
- Low-density Lipoprotein Cholesterol
- High-density Lipoprotein Cholesterol
- Total Cholesterol
- Triglycerides

Anthropomorphic Traits

- Body-Mass Index
- Waist Circumference
- Waist-to-Hip Ratio
- Height

Reproductive Traits

- Age at Menarche
- Age at Menopause
- Number of Children Ever Born
- Age at First Birth





Add Health Polygenic Scores – Release 1

Tobacco use

- Current/Ever Smoker
- Number of Cigarettes per day

Mental Health / Personality

- Bipolar Disorder
- ADHD
- Major Depressive Disorder
- Schizophrenia
- Mental Health Cross Disorder
- Extraversion



Cognitive Health / Education

- Alzheimer's Disease
- Educational Attainment



PGSs and GWAS Genetic Ancestry

Phenotype	GWAS Ancestry Group(s)
Coronary Artery Disease	European
Myocardial Infarction	European, South Asian, East Asian
Plasma Cortisol	European
Low-denisty Lipoprotein Cholesterol	European
High-denisty Lipoprotein Cholesterol	European
Total Cholesterol	European
Triglycerides	European
Type II Diabetes (2012)	European
Type II Diabetes (2018)	European, East Asian, South Asian, Mexican/Mexican-American
BMI	European
Waist Circumference	European
Waist-to-Hip Ratio	European
Height	European
Age at Menarche	European
Age at Menopause	European
Number of Children Ever Born	European
Age at First Birth	European
Ever/Current Smoker	European
Number of Cigarettes per day	European
Extraversion	European
Attention-deficit/hyperactivity Disorder (2010)	European
Attention-deficit/hyperactivity Disorder (2017)	European, Chinese
Bipolar Disorder	European
Major Depressive Disorder (2013)	European
Major Depressive Disorder (2018)	European
Schizophrenia	European, East Asian
Mental Health Cross Disorder	European
Alzheimer's Disease	European
Educational Attainment (2016)	European
Educational Attainment (2018)	European

PGSs for individuals not of the same ancestry group(s) as the GWAS from which summary statistics are retrieved may be less predictive (Martin et al. 2017; Ware et al. 2017).





Race/Ethnicity VS Genetic Ancestry

- While genetic ancestry and self-identified race/ethnicity are strongly correlated (0.89 in Add Health), they are two separate constructs.
 - Race and ethnicity are social constructs based on a multitude of factors, of which ancestry may be included depending on historical and societal differences in racialization (Omi and Winant 1994).
 - Genetic ancestry is defined using the first two principal components of the genotyped data.





Population Structure and Genetics



PMID:18758442















CENTER





Self-Identified Race/Ethnicity	European	African	East Asian	Hispanic	Total	
Non-Hispanic White	5,644	5	0	105	5,754	
Non-Hispanic Black	0	1,939	0	1	1,940	
Native American	14	2	0	6	22	
Asian	0	1	422	26	449	
Hispanic	70	27	14	850	961	_
Total Sample Size	5,728	1,974	436	988	9,126	

Ancestry





Using Add Health PGSs

- Accounting for ancestry
 - Separate analyses by ancestry
 - Ancestry-specific principal components
- Ancestry-specific principal components
 - Randomized in sets of five
 - 1-5, 6-10, 11-15, 16-20





SOME RESULTS





BMI (Waves I – IV)

BMI from Clinical Visit

All Ancestry Groups

1

POPULATION

CENTER

BMI from Self-Report All Ancestry Groups

1.0-1.0-+1 sd PGS (BMI) +1 sd PGS (BMI) -----1 sd PGS (BMI) -1 sd PGS (BMI) ----BMI from Self-Reported Weight & Height 0.5 0.5 BMI 0.0 0.0 Ŧ -0.5 -0.5 -1.0 -1.0 2010 1995 2000 2005 2010 1995 2000 2005 Year Year

Add Health The National Longitudinal Study of Adolescent to Adult Health

BMI (Waves I – IV)







BMI (Waves I – IV)

BMI from Clinical Visit

All Ancestry Groups

BMI from Self-Report

The

ational Longitudinal Study

All Ancestry Groups



POPULATION

CENTER

Depression (Waves I – IV)

All Ancestry Groups





Depression (Waves I – IV)







Word Recall (Wave 4)

All Ancestry Groups





Word Recall (Wave 4)







PGS Contracts

	dd Ho	alth	Social Rehavioral an	d Biological Linkages	Across the Life Course			Search Site	Search
The National Longitu	idinal Study of Adolescent	to Adult Health	Social, Behavioral, and Biological Linkages Across the Life Course				only in current section Follow @Add_Health		
Home	About	People	Study Design	Documentation	Contracts	FAQ	Publications	Contact Us	Participants
Security Plans Deductive Disclosure New Data Releases	re New Data Releases Wave IV: Ambient Air Pollutants Data (Released: July 20, 2018)								
		Ambient	AirPollutants_OrderFo	m					
 Particulate Matter: PM2.5, PM10, SO4, NO3, NH4, organic carbon (OC), elemental carbon (EC), Fe, Al, Si, Ti, Ca, Mg, K, Mn, Na, and Cl Gases: O3, NO2, HNO3, HONO, H2O2, CO, and SO2; 2007 only: acrolein, acetaldehyde, benzene, butadiene, ethanol, formaldehyde, and naphthalene 									
Parents Phase 2 Data (Released: July 20, 2018)									
	Parent Study_Order Form								
The parent data files contain social, demographic, behavioral, and health data collected in 2015-2017 on a probability sample of Add Health parents who were originally interviewed in 1995. Data for 2,013 Wave I parents, representing 2,247 Add Health sample members, are available. Additionally, 988 current spouse/partner interviews are available.									
Polygenic Scores (PGS) (Released: July 20, 2018)									
PGS_Order Form									
		Thirty constructed polygenic risk scores (PGS) are available for Add Health respondents who provided archival saliva samples for genetic testing at Wave IV (N = 9,129). Scores are available for coronary artery disease, myocardial infarction, plasma cortisol, LDL cholesterol, HDL cholesterol, total cholesterol, triglycerides, type II diabetes (2), BMI, waist circumference, waist-to-hip ratio, height, age at menarche, age at menopause, number of children, age at first birth, ever/current smoker, number of cigarettes per day, extraversion, attention deficit disorder (2), bipolar disorder, major depressive disorder (2), schizophrenia, mental health cross disorder, Alzheimer's disease, and educational attainment (2).							





PGS Contracts

http://www.cpc.unc.edu/projects/addhealth/contracts/data-releases





Thank you!

braudt@unc.edu

Funding Acknowledgements: This presentation was funded by grant R01 HD073342 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). Data for this study come from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from NICHD, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwise for assistance in the original design. Additional support was received from the Population Research Training grant (T32 HD007168) and the Population Research Infrastructure Program (P2C HD050924) awarded to the Carolina Population Center at The University of North Carolina at Chapel Hill by NICHD.



