Genome-wide Polygenic Scores for Multiple Psychiatric and Common Traits Identify Preadolescent Youth with Risk for Suicide

Short Title: Genes, Early Life Stress, and Youth Suicidality

Yoonjung Yoonie Joo, PhD,1 Seo-Yoon Moon,2* Hee-Hwan Wang,1* Hyeonjin Kim, MA,1 Eun-Ji Lee,1 Jonathan Posner, MD,4 Woo-Young Ahn, PhD,1 Incheol Choi, PhD,1 Jae-Won Kim, MD, PhD,5 Jiook Cha, PhD1,6

1. Department of Psychology, Seoul National University
2. College of Liberal Studies, Seoul National University
3. Department of Psychology, Yonsei University
4. Department of Psychiatry, Columbia University Medical Center, New York, New York
5. Department of Child and Adolescent Psychiatry, Seoul National University Hospital
6. AI Institute, Seoul National University
(* denotes equal contribution)

Corresponding author:

Jiook Cha, Ph.D.
Gwanak-ro 1, Gwanak-gu, Seoul, Republic of Korea, 08826
Tel: 82-2-880-8618, Email: connectome@snu.ac.kr

Word counts: 2,020 (excluding abstract, online methods, references, figure legends)

Keywords:

Psychiatric Genetics; Genome-wide Polygenic Score; Childhood Suicidality; Early Life Stress
Abstract

Suicide is a leading cause of death in youth worldwide, but identifying which youth are at high risk for suicide remains challenging. We constructed genome-wide polygenic scores (GPSs) from 24 psychiatric disorders and common traits from 8,212 US preadolescent children ages 9 to 10 and investigated their associations and predictive utility with suicidality (suicidal ideation and attempt). We identified three GPSs significantly associated with youth suicidality: ADHD ($P=2.83\times10^{-4}$; odds ratio=1.12), general happiness with a belief that life is meaningful ($P=1.30\times10^{-3}$; odds ratio=0.89) and autism spectrum disorder (ASD) ($P=1.81\times10^{-3}$; odds ratio=1.14). We also found a significant gene-by-environment interaction such that the GPS of ASD in the context of early life stress substantially increased suicidal ideation ($P=1.39\times10^{-2}$, odds ratio=1.11). Machine learning models showed, in predicting suicidal ideation, a receiver operators characteristics-area under the curve (ROC-AUC) of 0.72, and, in suicidal attempts, a ROC-AUC of 0.765. By providing the first quantitative account of the polygenic and environmental factors of suicidality in a large, representative population of preadolescent youth, this study shows the potential utility of the GPSs in investigating youth suicidality for early screening, intervention, and prevention.
Introduction

Every 10 minutes someone in the U.S. dies by suicide. In youth, suicide has been among the leading causes of death; the suicide-rate curve has not been bent in decades.\textsuperscript{1,2} Identifying preadolescent youth with high risk for suicide is challenging and existing approaches show poor predictive validity.\textsuperscript{3} Twin and family studies have suggested that genetic factors play an important role in suicidal behavior\textsuperscript{4,5}, with the estimated heritability ranging from 30% to 55%\textsuperscript{4,6}. The literature supports that an inherited genetic factor may account for suicidal risk, and furthermore, the genetic predisposition may be aggravated by psychiatric illnesses and environmental factors such as stress\textsuperscript{7}. No literature has addressed this in children yet.

The overall genetic architecture of suicidal behaviors is assumed to be polygenic\textsuperscript{8,9}, involving the cumulative effects of numerous single-nucleotide polymorphisms (SNPs), each often with small effects. Recent Genome-wide Association Studies (GWAS) revealed several loci linked to completed suicide\textsuperscript{10,11}, suicide attempt\textsuperscript{12,13}, and suicidal ideation\textsuperscript{14,15,16,17}. But none of the loci have been replicated across studies. Moreover, GWAS do not determine how much the genetic information contributes to identification of children at risk for suicide. As a promising alternative, genome-wide polygenic scores (GPS) aim to integrate numerous effects of genome-wide SNPs\textsuperscript{18,19} to predict the likelihood of a complex trait in individuals,\textsuperscript{20} potentially enabling personalized medicine.\textsuperscript{22,23,24}

Genetic influences on or predisposition of suicidal behavior also interacts with environmental factors (GxE interaction).\textsuperscript{9} For example, early life stress (ELS) contributes to suicidality (suicidal ideation and/or attempt)\textsuperscript{25,26} potentially through the epigenetic mechanisms\textsuperscript{27,28}. Testing whether and to what extents ELS and genetic factors together act synergistically on youth suicide will offer a much-needed insight into the biological pathway of suicide and an actionable target for intervention. No literature has yet to report this. The predictive utility of genetic factors had never been examined while adjusting for potential
environmental confounders, such as early life stress or sociodemographic characteristics. Determining these factors and their interaction may help elucidate early screening of risk for suicide. We test the extent to which GPSs\textsuperscript{22,23} for common traits and psychiatric disorders, and their interaction with ELS is linked to the risk for suicide in young children.
Results

We used data from the Adolescent Brain and Cognitive Development (ABCD) study, a nationwide multisite prospective, longitudinal study. The enrolled sample includes 11,827 preadolescent children ages of 9 to 10 years old in the US across 21 sites recruited between 2015 and 2019. Among the samples, 1,656 case and 10,171 controls were found for suicidal ideation, and 124 cases and 11,703 controls were found for suicide attempt (Figure 1). Preadolescent children with suicidal ideation or attempts showed similar sociodemographic, behavioral, clinical characteristics to those without suicidal ideation or attempts (Table 1).

For GPS generation, we selected 24 psychiatric and common traits that are known to be related to suicidality including personality, cognitive, psychological traits, and psychiatric disorders that are known to be broadly related to suicidality: general happiness, 29,30 insomnia, 31 depression, 32 risk behaviors, 33 risk tolerance, educational attainment, 34,35 cognitive performance, 33,34 snoring, 36 worry, 37 IQ, 34 cannabis usage, 38,39 drink per week, 40 smoker, 41 Attention deficit hyperactivity disorder (ADHD), 42 Autism spectrum disorder (ASD), 43 major depressive disorder (MDD), 44 schizophrenia, 45 bipolar disorder, 46 post-traumatic stress disorder (PTSD). 47 The GPS for each individual was computed as a sum of their SNPs adjusting for first 10 principal components, with each SNP being weighted by the effect from each GWAS summary statistic. 48 Since we wanted to fully incorporate the effects of genome-wide SNPs, we used no thresholding of p-value significance and included all the available SNPs with GWAS p-value < 1.

Out of the 24 GPSs, in suicidal ideation, a greater GPS of ADHD significantly correlated with a greater likelihood (P=2.83×10^{-4}, odds ratio (OR)=1.12; FDR significance). Likewise, the ADHD GPS significantly correlated with active (P=9.75×10^{-4}, OR=1.14) and passive suicidal ideation (P=6.63×10^{-4}, OR=1.13), respectively (Table 2, Figure 2). In active suicidal ideation, a greater GPS of ASD correlated with a greater likelihood of suicidal ideation (P=1.81×10^{-3}, OR=1.14). Conversely, in passive suicidal ideation, a less GPS of
general happiness correlated with a greater likelihood of suicidal ideation \( (P=1.30 \times 10^{-3}, \text{OR} = 0.89) \). Correlations among the tested GPSs were tested and none of their correlations seemed to affect the results (Supplementary Figure 1).

Of those three GPSs showing significant associations with suicidality, we further tested an interaction with ELS. We found a significant ASD-GPS-by-ELS interaction on active suicidal ideation: a greater ASD GPS in the presence of ELS correlated with a synergistic increase in these likelihood of suicidality \( (P=1.39 \times 10^{-2}, \text{OR}=1.11; \text{FDR corrected for the three analyses}) \). Effects were adjusted for the same covariates with the regression analysis. No direct associations were found significant between ELS and suicidality \( (P_s > 0.24) \). When stratified by sex, no significant associations were found.

Additionally, we performed the same analysis using only healthy controls without any KSAD records \( (n=3,244) \). The signals of the ADHD GPS (strongest association with suicidal ideation, \( P=4.42 \times 10^{-4}, \text{OR}=1.15 \)) and ASD GPS (with suicidal ideation, \( P=8.37 \times 10^{-4}, \text{OR}=1.14 \)) remained significant and even became stronger in terms of effect size compared to the original analysis including every controls (Supplementary Table 3(a)). The sex-stratified analysis with healthy controls identified a significant association between active suicidal ideation and the ADHD GPS \( (P=4.49 \times 10^{-4}, \text{OR}=1.31) \) in the female population surviving the FDR significance (Supplementary Table 3(b)).

We tested whether the multiple GPSs contributed to the improved prediction of youth suicidality using machine learning. The following input features were used: 24 GPSs, socio-demographic information (sex, age, marital status, parental education, study site), psychological information (Child behavior checklist, CBCL), cognitive phenotypes (NIH toolbox), family environment factors (Youth Family Environment Scale), and ELS. In predicting suicidal ideation, the cross-validated and optimized stacked ensemble model showed an ROC-AUC of 0.720, accuracy of 0.756, sensitivity of 0.837, specificity of 0.058, positive predictive value of 0.538, negative predictive value of 0.214 in the balanced held-out test set \( (N=476; \text{bootstrap samples}) \) (Figure 3). Adding the GPS to the phenotype model
resulted in a 2.4% boost in ROC-AUC. In predicting suicidal attempts, the cross validated and optimized stacked ensemble model showed an ROC-AUC of 0.765, accuracy of 0.750, sensitivity of 0.583, specificity of 0.25, positive predictive value of 0.70, negative predictive value of 0.167 in the balanced held-out test set (N= 32; bootstrap samples) (Figure 3). Adding the GPS to the phenotype model resulted in a 3.2% boost in ROC-AUC. However, adding ELS or ELS-by-happiness GPS interaction did not show statistically meaningful improvement. The stacked ensemble models always outperformed the benchmark Gradient Boosting Machine model.
Discussion

We tested the utility of the GPS in identifying preadolescent youth at risk for suicide. Using large-scale, representative samples of youth, we found novel associations between youth suicidality and GPS for ADHD, ASD -- positive associations -- and general happiness - - a negative association. We next found significant gene-by-environment interactions between the GPS of ASD and ELS, a known risk factor for youth suicidality, together acting synergistically on youth suicidality. In our data-driven predictive modeling, together with the self-reported questionnaires for psychopathology (e.g., CBCL), intelligence, family environment, and socio-demographic variables, inclusion of the multiple GPSs permitted to classify preadolescent youth with suicidality with moderate accuracy. In sum, this study sheds light on the genetic approach to youth suicidality for better understanding of the etiologic pathway and for prevention and intervention.

Using the multi-GPS method, we discovered the association of youth suicidality with the genetic profiles of psychiatric disorders, i.e., ADHD and ASD, particularly relevant to childhood psychopathology, and of a common trait, i.e., general happiness. For ADHD and ASD, a greater polygenic score correlated with a greater likelihood of suicidal ideation, whereas for general happiness, a smaller polygenic score correlates with a greater likelihood of passive suicide ideation. This is in line with the literature, suggesting that ADHD is associated with suicidal attempts in youth. Our GPS results may provide genomic evidence for the link between ADHD and youth suicidality. The explained variance (McFadden’s pseudo-$R^2$) of suicidal phenotypes by ADHD GPS was approximately 1.5% in children. This estimation is higher than the results from past PRS studies of suicidality, which showed maximum variance explained by 0.13-0.20% with self-harm behaviors or up to 0.30-0.70% of the phenotypic variance for suicide attempt explained by depression-based GPS.

The GPS of ASD not only correlated with suicidal ideation, but also showed a significant interaction with ELS. These results corroborate previous findings of preadolescent
youth with ASD being at an increased risk for suicidality.\textsuperscript{54,55} One study shows that youth with ASD are 28 times more likely to endorse suicidal thoughts or behaviors than their unaffected peers.\textsuperscript{56} This strong link between risk of suicidality and elevated autistic traits could be explained by specific behavioral attributes of both phenotypes, such as poor socialization and problem-solving skill, or increased levels of anxiety.\textsuperscript{55} Although literature has reported several social risk factors underlying the ASD-suicidality link,\textsuperscript{56,57} the genetic and environmental contribution to the overlap of ASD and suicidal behaviors remained unknown.\textsuperscript{58} Our results fill this gap by presenting the shared genetic factors for ASD and suicidality, also in environmental (childhood experiences) dimensions. The results thus further suggest that protective, as opposed to adverse, childhood experiences are particularly important for those with genetic risk of ASD. This might be a potential actionable target for intervention in that population.

In line with our findings, previous literature has reported high suicidal risk in individuals who in addition to ASD also have ADHD.\textsuperscript{58} As 20-50\% of individuals with ASD are known to have comorbid ADHD,\textsuperscript{59} our significant results of ASD and ADHD GPSs associated with suicidality suggest the genetic liability of psychiatric comorbidity for high suicidal risk in children.

The association of a smaller GPS of general happiness (believing that her or his life is meaningful) and a greater likelihood of youth suicidal ideation is novel. Of several measures of subjective well-being, only the belief of own meaningful life is significantly linked to youth suicidality. Prior studies report a negative correlation between subjective happiness and suicide\textsuperscript{29,30}. Our results corroborate the link by showing genetic underpinning of the relationship.

Our study contributes in several ways to our genomic understanding of suicidal phenotypes and provides a basis for utilizing multiple GPSs for suicide prediction. The combination of the non-suicidal GPSs, behavioral, psychological scales, and ELS allowed accurate prediction of suicidality with a ROC-AUC up to 0.720 for suicidal ideation and 0.765
for suicidal attempts on the balanced test set. Several previous studies have tested different methods for suicide risk prediction and showed substantial classification ability having AUC ranging from 0.74 to 0.88, using known sociodemographic and psychiatric risk factors. Although successful, some of these models included long-term measurement of risk indices such as 12-month or lifelong risk factors, which are not readily measured in naturalistic clinical practices. Our models showed relatively lower performances than previous studies, but our methodology is promising in that they incorporated short-term or cross-sectional factors that could be accessible and readily measured in clinical settings, on top of the genetic factors. Despite the acceptable accuracy and sensitivity, however, there is much room for improvement regarding specificity and negative predictive value of the model. Despite the potential utility of the GPS in youth suicide research discovered here, further methodological improvement is needed to enhance their utility in early screening. First, in generating the GPS, optimizing hyper-parameters (e.g., p-value cutoff or linkage disequilibrium criteria) may lead to an improved predictive power. On top of the genetic factors, future research should identify more environmental risk factors, e.g., extending to social factors or sociomarkers. Also, due to the difficulty in obtaining committed suicide as a outcome measure in preadolescent youth (e.g., less than 100 in the US), we alternatively used suicidal ideation and attempts as a proxy phenotype of suicidal behavior. Despite the high correlation between suicidal behaviors and greater risk for suicide attempt, we acknowledge this limitation in that not all suicidal ideation or attempts lead to committed suicide.

By providing the first quantitative account of the polygenic and environmental factors of youth suicidality in a large, representative population, this study shows the importance and potential utility of the GPS approach in preadolescent youth suicidality in early screening. These findings motivate the future investigation of more effective screening and intervention strategies for youth suicidality.
Online Methods

Study design and participants

The study sample includes 11,827 preadolescent children ages of 9 to 10 years old in the US across 21 sites recruited between 2015 and 2019 from the Adolescent Brain and Cognitive Development (ABCD) study. Among the samples, 1,656 case and 10,171 controls were found for suicidal ideation, and 124 cases and 11,703 controls were found for suicide attempt (Table 1). Ethical approval for the study was obtained from Seoul National University IRB. Full details of the study, measures, and samples can be found elsewhere.69

Genotype data

Saliva samples of the participants were collected and genotyped at Rutgers University Cell and DNA Repository using Affymetrix SmokeScreen Array consisting of 733,293 single nucleotide polymorphisms (SNPs). After removing the SNP with genotype call rate <95%, sample call rate <95%, and minor allele frequency (MAF) <1%, raw genotypes were imputed toward 1000G Phase5 reference panel using the Michigan Imputation Server and phased with Eagle v2.4. Additional quality control (QC) process removed SNPs with genotype call rate <95%, Hardy-Weinberg Equilibrium p-value <1E-06, sample missingness >5%, and MAF <1%. We also removed samples with extreme heterozygosity having F coefficient bigger than 3 standard deviation of the population mean. We conducted Principal Component Analysis (PCA) on genetic variants that are pruned for variants in linkage disequilibrium with an r² > 0.25 in a 200kb window. Since ABCD participants reside on a continuum of genetic ancestry as American population (1000Genome super population), rather than distinct population groups, and none of them belongs to either African or East Asian populations, we did not exclude any genetic outliers based on PCA biplot (Supplementary Figure 2). Proportional identify-by-descent analysis was performed with the same SNP pruning using PLINK70 and we removed moderately related individuals (pihat > 0.18). After QC, genotype data were available for 8,496 children.
Construction of Genomewide Polygenic Scores (GPS)

For GPS generation, we selected 24 psychiatric and common traits that are known to be related to suicidality including personality, cognitive, psychological traits, and psychiatric disorders that are known to be broadly related to suicidality: general happiness, \(^{29,30}\) insomnia, \(^{31}\) depression, \(^{32}\) risk behaviors, \(^{33}\) risk tolerance, educational attainment, \(^{34,35}\) cognitive performance, \(^{33,34}\) snoring, \(^{36}\) worry, \(^{37}\) IQ, \(^{34}\) cannabis usage, \(^{38,39}\) drink per week, \(^{40}\) smoker, \(^{41}\) Attention deficit hyperactivity disorder (ADHD), \(^{42}\) Autism spectrum disorder (ASD), \(^{43}\) major depressive disorder (MDD), \(^{44}\) schizophrenia, \(^{45}\) bipolar disorder, \(^{46}\) post-traumatic stress disorder (PTSD). \(^{47}\) For general happiness, four GPS were built and tested for the study: the participants were asked with different questionnaires about their level of subjective well-being, such as (1) two different GWAS of “how happy are you in general”, (2) “how happy are you with your health in general”, and (3) “to what extent do you feel your life to be meaningful”. All the GWAS summary statistics of the aforementioned traits are publically available and have been collected for GPS generation.

We performed clumping and pruning of SNPs using PRSice2\(^ {48}\) with a clumping window of 500 kb, clumping \(r^2\) of 0.2, and no thresholding of p-value significance on the summary statistics since we wanted to fully incorporate the effects of all the SNPs. The GPS for each individual was then computed as a sum of their SNPs adjusting for first 10 PCs, with each SNP being weighted by the effect in the discovery samples. \(^ {48}\)

Outcomes

Suicidal ideation and attempt were derived from the computerized version of Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS).\(^ {71,72}\) Passive suicide ideation is wanting to be dead and active ideation is considering suicide with specific methods or plans. Among parent and child reports of KSADS, we used one reporting more severe symptoms or diagnoses. We adjusted for demographic covariates including sex, age, study site, years of
parental education, and parental marital status. The following variables were additionally considered for inputs to the classification models. For psychopathology, ABCD Parent Child Behavior Checklist (CBCL); for intelligence, NIH toolbox; for family environment, Youth Family Environment Scale; and for ELS, physical and sexual abuse, household challenges such as family substance abuse, family mental illness, family criminals, parental separation or divorce, and violently treated mothers and emotional and physical neglect (a total composite score was used).

Statistical Analysis and Machine Learning Prediction

After QC and GPS construction, the complete dataset of phenotypic outcomes, GPS, and covariate data were available for 8,212 children and used for statistical analysis. (Figure 1) Associations of the GPSs with suicidal variables within the ABCD children were tested using logistic regression with the following covariates: sex, age, site for sample collection, and maternal education level. We also performed the same analysis after excluding any control individuals with one or more KSAD records (n=3,108 healthy controls remained).

To test the extents to which the GPSs are useful to predict the suicidal risk in children, we trained stacked ensemble models or super learners based on the several machine learning models, such as lightGBM, xgboost, general linear model (Driverless AI version 1.8, H2O.ai Inc., CA, USA) and, as a benchmark to the stacked ensemble model, Gradient Boosting Machine (Scikit-learn 0.21; https://scikit-learn.org/). The input variables included the GPSs, sociodemographic, self-reported, and behavioral variables (e.g., psychopathology, cognitive, early life stress). We rationalized that the multitude of the genetic profile (multiple GPSs) would account for the multi-dimensional risk for suicide. For model training and evaluation, we split the data into 80% and 20% in which samples were balanced (bootstrapping without replacement). Within the 80% training data, we conducted five-fold stratified cross validation and optimized hyperparameters in cross validation sets.
Author contributions

Study concept and design: Y.J., J.C.
Acquisition, analysis, or interpretation of data: Y.J., H.K., S.M., H.W.
Drafting of the manuscript: Y.J., S.M., H.W., J.C.
Critical revision of the manuscript for important intellectual content: Y.J., J.K., I.C., J.C., W.-Y.A.
Statistical analysis: Y.J., S.M., H.W.
Obtained funding: I.C., J.C.
Study supervision: J.C.

Data and Code Availability

Codes and data are freely available for reproducibility
(https://github.com/Transconnectome/ConnectomeLab/blob/master/suicide).

Competing Interests Declaration

None of the authors have significant competing financial, professional, or personal interests that might have influenced the performance or presentation of the work described in the manuscript.

Acknowledgement

This work was supported by the New Faculty Startup Fund from Seoul National University, Research grant from the Center for Happiness via the Center for Social Sciences at Seoul National University at Seoul National University, and the H2O.ai Academic Program at H2O inc., CA, USA.
References.

1. Cunningham, R. M., Walton, M. A. & Carter, P. M. The Major Causes of Death in Children and Adolescents in the United States. *New England Journal of Medicine* vol. 379 2468–2475 (2018).
2. Website. National Institute of Mental Health. Mental health information. Statistics. Suicide. https://www.nimh.nih.gov/health/statistics/suicide.shtml.
3. Belsher, B. E. *et al.* Prediction Models for Suicide Attempts and Deaths: A Systematic Review and Simulation. *JAMA Psychiatry* 76, 642–651 (2019).
4. Brent, D. A., Bridge, J., Johnson, B. A. & Connolly, J. Suicidal behavior runs in families. A controlled family study of adolescent suicide victims. *Arch. Gen. Psychiatry* 53, 1145–1152 (1996).
5. Roy, A. & Segal, N. L. Suicidal behavior in twins: a replication. *J. Affect. Disord.* 66, 71–74 (2001).
6. Voracek, M. & Loibl, L. M. Genetics of suicide: a systematic review of twin studies. *Wien. Klin. Wochenschr.* 119, 463–475 (2007).
7. Mann, J. J., Waternaux, C., Haas, G. L. & Malone, K. M. Toward a clinical model of suicidal behavior in psychiatric patients. *Am. J. Psychiatry* 156, 181–189 (1999).
8. Sokolowski, M., Wasserman, J. & Wasserman, D. Polygenic associations of neurodevelopmental genes in suicide attempt. *Mol. Psychiatry* 21, 1381–1390 (2016).
9. Mullins, N. *et al.* Genetic relationships between suicide attempts, suicidal ideation and major psychiatric disorders: a genome-wide association and polygenic scoring study. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 165B, 428–437 (2014).
10. Docherty, A. R. *et al.* Genome-Wide Association Study of Suicide Death and Polygenic Prediction of Clinical Antecedents. *American Journal of Psychiatry* vol. 177 917–927 (2020).
11. Galfalvy, H. *et al.* A pilot genome wide association and gene expression array study of suicide with and without major depression. *World J. Biol. Psychiatry* 14, 574–582 (2013).
12. Willour, V. L. *et al.* A genome-wide association study of attempted suicide. *Molecular Psychiatry* vol. 17 433–444 (2012).
13. Perlis, R. H. *et al.* Genome-Wide Association Study of Suicide Attempts in Mood Disorder Patients. *American Journal of Psychiatry* vol. 167 1499–1507 (2010).
14. Mullins, N. *et al.* GWAS of Suicide Attempt in Psychiatric Disorders and Association With Major Depression Polygenic Risk Scores. *Am. J. Psychiatry* 176, 651–660 (2019).
15. Otsuka, I. *et al.* Genome-wide association studies identify polygenic effects for completed suicide in the Japanese population. *Neuropsychopharmacology* 44, 2119–2124 (2019).
16. Laje, G. *et al.* Genome-wide association study of suicidal ideation emerging during citalopram treatment of depressed outpatients. *Pharmacogenetics and Genomics* vol. 19 666–674 (2009).
17. Perroud, N. *et al.* Genome-wide association study of increasing suicidal ideation during antidepressant treatment in the GENDEP project. *Pharmacogenomics J.* 12, 68–77 (2012).
18. Consortium, T. I. S. & The International Schizophrenia Consortium. Common polygenic
variation contributes to risk of schizophrenia and bipolar disorder. *Nature* vol. 460 748–752 (2009).

19. Consortium, S. W. G. of T. P. G. & Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* vol. 511 421–427 (2014).

20. Wray, N. R., Goddard, M. E. & Visscher, P. M. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Research* vol. 17 1520–1528 (2007).

21. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* 9, e1003348 (2013).

22. Hunter, D. J. & Longo, D. L. The Precision of Evidence Needed to Practice ‘Precision Medicine’. *New England Journal of Medicine* vol. 380 2472–2474 (2019).

23. Sugrue, L. P. & Desikan, R. S. What Are Polygenic Scores and Why Are They Important? JAMA 321, 1820–1821 (2019).

24. Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* 19, 581–590 (2018).

25. Turecki, G., Ota, V. K., Belangero, S. I., Jackowski, A. & Kaufman, J. Early life adversity, genomic plasticity, and psychopathology. *Lancet Psychiatry* 1, 461–466 (2014).

26. Turecki, G. & Brent, D. A. Suicide and suicidal behaviour. *The Lancet* vol. 387 1227–1239 (2016).

27. Lewis, C. R. & Foster Olive, M. Early-life stress interactions with the epigenome. *Behavioural Pharmacology* vol. 25 341–351 (2014).

28. Cruceanu, C., Matosin, N. & Binder, E. B. Interactions of early-life stress with the genome and epigenome: from prenatal stress to psychiatric disorders. *Current Opinion in Behavioral Sciences* vol. 14 167–171 (2017).

29. Pompili, M. et al. The Interplay Between Suicide Risk, Cognitive Vulnerability, Subjective Happiness and Depression Among Students. *Current Psychology* vol. 35 450–458 (2016).

30. Hsu, C.-Y., Chang, S.-S. & Yip, P. S. F. Subjective wellbeing, suicide and socioeconomic factors: an ecological analysis in Hong Kong. *Epidemiology and Psychiatric Sciences* vol. 28 112–130 (2019).

31. Hamilton, J. L. & Buysse, D. J. Reducing Suicidality Through Insomnia Treatment: Critical Next Steps in Suicide Prevention. *American Journal of Psychiatry* vol. 176 897–899 (2019).

32. Gabrielle A. Carlson, D. P. C. Suicidal Behavior and Depression in Children and Adolescents. *J. Am. Acad. Child Psychiatry* 21, 361–368 (1982).

33. King, R. A. et al. Psychosocial and risk behavior correlates of youth suicide attempts and suicidal ideation. *J. Am. Acad. Child Adolesc. Psychiatry* 40, 837–846 (2001).

34. Keilp, J. G. et al. Neuropsychological dysfunction in depressed suicide attempters. *Am. J. Psychiatry* 158, 735–741 (2001).

35. Rosoff, D. B., Kaminsky, Z. A., McIntosh, A. M., Davey Smith, G. & Lohoff, F. W. Educational attainment reduces the risk of suicide attempt among individuals with and without psychiatric disorders independent of cognition: a bidirectional and multivariable Mendelian randomization study with more than 815,000 participants. *Transl. Psychiatry* 10, 388 (2020).
36. Becker, S. P., Dvorsky, M. R., Holdaway, A. S. & Luebbe, A. M. Sleep problems and suicidal behaviors in college students. J. Psychiatr. Res. 99, 122–128 (2018).
37. Gorday, J. Y., Rogers, M. L. & Joiner, T. E. Examining characteristics of worry in relation to depression, anxiety, and suicidal ideation and attempts. J. Psychiatr. Res. 107, 97–103 (2018).
38. Price, C., Hemmingsson, T., Lewis, G., Zammit, S. & Allebeck, P. Cannabis and suicide: longitudinal study. Br. J. Psychiatry 195, 492–497 (2009).
39. Orri, M. et al. A genetically informed study on the association of cannabis, alcohol, and tobacco smoking with suicide attempt. Molecular Psychiatry (2020) doi:10.1038/s41380-020-0785-6.
40. Borges, G. et al. A meta-analysis of acute use of alcohol and the risk of suicide attempt. Psychol. Med. 47, 949–957 (2017).
41. Green, M., Turner, S. & Sareen, J. Smoking and suicide: biological and social evidence and causal mechanisms. J. Epidemiol. Community Health 71, 839–840 (2017).
42. Beauchaine, T. P., Ben-David, I. & Bos, M. ADHD, financial distress, and suicide in adulthood: A population study. Sci Adv 6, (2020).
43. Cassidy, S. & Rodgers, J. Understanding and prevention of suicide in autism. Lancet Psychiatry 4, e11 (2017).
44. Angst, J., Angst, F. & Stassen, H. H. Suicide risk in patients with major depressive disorder. J. Clin. Psychiatry 60 Suppl 2, 57–62; discussion 75–6, 113–6 (1999).
45. Palmer, B. A., Shane Pankratz, V. & Bostwick, J. M. The Lifetime Risk of Suicide in Schizophrenia. Archives of General Psychiatry vol. 62 247 (2005).
46. Hawton, K., Sutton, L., Haw, C., Sinclair, J. & Harriss, L. Suicide and attempted suicide in bipolar disorder: a systematic review of risk factors. J. Clin. Psychiatry 66, 693–704 (2005).
47. Gradus, J. L. et al. Posttraumatic Stress Disorder and Completed Suicide. American Journal of Epidemiology vol. 171 721–727 (2010).
48. Choi, S. W., Mak, T. S.-H. & O’Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. Nat. Protoc. 15, 2759–2772 (2020).
49. Chronis-Tuscano, A. et al. Very early predictors of adolescent depression and suicide attempts in children with attention-deficit/hyperactivity disorder. Arch. Gen. Psychiatry 67, 1044–1051 (2010).
50. McFadden, D. Conditional Logit Analysis of Qualitative Choice Behavior. (1973).
51. Campos, A. I. et al. Genetic aetiology of self-harm ideation and behaviour. Sci. Rep. 10, 9713 (2020).
52. Levey, D. F. et al. Genetic associations with suicide attempt severity and genetic overlap with major depression. Transl. Psychiatry 9, 22 (2019).
53. Shen, H. et al. Polygenic prediction and GWAS of depression, PTSD, and suicidal ideation/self-harm in a Peruvian cohort. Neuropsychopharmacology vol. 45 1595–1602 (2020).
54. Howe, S. J., Hewitt, K., Baraskewich, J., Cassidy, S. & McMorris, C. A. Suicidality Among Children and Youth With and Without Autism Spectrum Disorder: A Systematic Review of Existing Risk Assessment Tools. J. Autism Dev. Disord. 50, 3462–3476 (2020).
55. Chen, Y.-Y., Chen, Y.-L. & Gau, S. S.-F. Suicidality in Children with Elevated Autistic
56. Mayes, S. D., Gorman, A. A., Hillwig-Garcia, J. & Syed, E. Suicide ideation and attempts in children with autism. Research in Autism Spectrum Disorders vol. 7 109–119 (2013).

57. McDonnell, C. G. et al. An Exploratory Analysis of Predictors of Youth Suicide-Related Behaviors in Autism Spectrum Disorder: Implications for Prevention Science. J. Autism Dev. Disord. 50, 3531–3544 (2020).

58. Hirvikoski, T. et al. Individual risk and familial liability for suicide attempt and suicide in autism: a population-based study. Psychol. Med. 50, 1463–1474 (2020).

59. Ljung, T., Chen, Q., Lichtenstein, P. & Larsson, H. Common etiological factors of attention-deficit/hyperactivity disorder and suicidal behavior: a population-based study in Sweden. JAMA Psychiatry 71, 958–964 (2014).

60. Borges, G. et al. Twelve-month prevalence of and risk factors for suicide attempts in the World Health Organization World Mental Health Surveys. J. Clin. Psychiatry 71, 1617–1628 (2010).

61. Borges, G. et al. A risk index for 12-month suicide attempts in the National Comorbidity Survey Replication (NCS-R). Psychol. Med. 36, 1747–1757 (2006).

62. Mann, J. J. et al. Classification Trees Distinguish Suicide Attempters in Major Psychiatric Disorders. The Journal of Clinical Psychiatry vol. 69 23–31 (2008).

63. Glenn, C. R. & Nock, M. K. Improving the short-term prediction of suicidal behavior. Am. J. Prev. Med. 47, S176–80 (2014).

64. Shin, E. K. et al. Association of Maternal Social Relationships With Cognitive Development in Early Childhood. JAMA Netw Open 2, e186963 (2019).

65. Shin, E. K., Mahajan, R., Akbilgic, O. & Shaban-Nejad, A. Sociomarkers and biomarkers: predictive modeling in identifying pediatric asthma patients at risk of hospital revisits. NPJ Digit Med 1, 50 (2018).

66. Cash, S. J. & Bridge, J. A. Epidemiology of youth suicide and suicidal behavior. Current Opinion in Pediatrics vol. 21 613–619 (2009).

67. Pigeon, W. R., Bishop, T. M. & Titus, C. E. The Relationship Between Sleep Disturbance, Suicidal Ideation, Suicide Attempts, and Suicide Among Adults: A Systematic Review. Psychiatric Annals vol. 46 177–186 (2016).

68. Law, K. C., Jin, H. M. & Anestis, M. D. The intensity of suicidal ideation at the worst point and its association with suicide attempts. Psychiatry Res. 269, 524–528 (2018).

69. Karcher, N. R. & Barch, D. M. The ABCD study: understanding the development of risk for mental and physical health outcomes. Neuropsychopharmacology (2020) doi:10.1038/s41386-020-0736-6.

70. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7 (2015).

71. Kaufman, J. et al. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. J. Am. Acad. Child Adolesc. Psychiatry 36, 980–988 (1997).

72. Kaufman, J., Townsend, L. D. & Kobak, K. The Computerized Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS): Development and Administration Guidelines. Journal of the American Academy of Child & Adolescent Psychiatry vol. 56 S357 (2017).
73. Luciana, M. et al. Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery. Dev. Cogn. Neurosci. 32, 67–79 (2018).

74. Laan, M. J. van der, van der Laan, M. J., Polley, E. C. & Hubbard, A. E. Super Learner. Statistical Applications in Genetics and Molecular Biology vol. 6 (2007).

75. H2O Driverless AI - Open Source Leader in AI and ML. https://www.h2o.ai/products/h2o-driverless-ai/.
Table 1. Sociodemographic and clinical outcomes of the study participants from the Adolescent Brain and Cognitive Development (ABCD) study. The completed dataset of phenotypic outcomes, genotype, and covariates data available for 8,212 preadolescent children.

|                             | Youth with suicidality (n=1,161) | Control Youth (N=7,051) |
|-----------------------------|----------------------------------|-------------------------|
|                             | Count or Mean | Percentage (%) or SD | Count or Mean | Percentage (%) or SD |
| Female                      | 451               | 38.8%                   | 3413               | 48.4%                   |
| Average Age (in months)     | 118.8             | 7.54                    | 119.0             | 7.45                    |
| Average Income (bracket in 1~10 scales) | 7.07               | 2.38                    | 7.26               | 2.41                    |
| Marital status of the family (Currently married) | 733               | 63.1%                   | 4987               | 70.7%                   |
| Average Mother's EA         | 16.62             | 2.63                    | 16.61             | 2.78                    |
Table 2. Significant associations of GPSs of 24 psychiatric and common traits with youth suicidality, surviving FDR significance < 0.05.

| Outcome                        | GPS predictor | OR  | 95% OR       | P-value          | N of cases |
|--------------------------------|---------------|-----|--------------|------------------|------------|
|                                |               |     |              | (FDR corrected)  |            |
| Suicidal Ideation (Active + Passive) | ADHD        | 1.123 | (1.06 - 1.20) | 2.83E-04 (2.27E-02) | 1155       |
| Passive Suicidal Ideation      | ADHD         | 1.126 | (1.05 - 1.20) | 6.63E-04 (3.11E-02) | 946        |
| Active Suicidal Ideation       | ADHD         | 1.15  | (1.06 - 1.25) | 9.75E-04 (3.11E-02) | 611        |
| Passive Suicidal Ideation      | GENERAL HAPPINESS MEANINGFUL | 0.895 | (0.84 - 0.96) | 1.30E-03 (3.11E-02) | 946        |
| Active Suicidal Ideation       | ASD          | 1.141 | (1.05 - 1.24) | 1.81E-03 (3.48E-02) | 611        |
Figure 1. Consort flow diagram for the study. The study initially assessed 11,827 preadolescent children ages of 9 to 10 years old recruited from the Adolescent Brain and Cognitive Development (ABCD) study. The completed dataset of phenotypic outcomes, genotype, and covariates data available for 8,212 preadolescent children and used for the analysis.
Figure 2. Manhattan plot showing associations of multiple genome-wide polygenic scores (GPS) of cognitive, psychological, personality traits, and psychiatric disorders with childhood suicidality. Red line indicates FDR corrected p-value of 0.05. Target outcomes of the GPSs are color-coded. An effect size of each GPS is denoted with the circle size.
Figure 3. Prediction performance of the machine learning models based on GPS and cognitive, psychological, behavioral, environmental, familial variables. Receiver-Operator Characteristics curves of the models predicting suicidal ideation (A) and suicide attempt (B).
Supplementary Materials

Genome-wide Polygenic Scores for Multiple Psychiatric and Common Traits Identify Preadolescent Youth with Risk for Suicide

Yoonjung Yoonie Joo, PhD, Seo-Yoon Moon,* Hee-Hwan Wang,* Hyeonjin Kim, MA, Eun-Ji Lee, Jonathan Posner, MD, Woo-Young Ahn, PhD, Incheol Choi, PhD, Jae-Won Kim, MD, PhD, Jiook Cha, PhD

Contents:

Supplementary Table 1
Supplementary Table 2
Supplementary Table 3
Supplementary Figure 1
Supplementary Figure 2
Supplementary References
**Supplementary Table 1.** Performance of suicidality classification models in the imbalanced validation set (in suicidal ideation, 956 cases and 7,030 controls; in suicide attempt, 64 cases and 8,366 controls)

| Model                           | Base Model¹ | Base+Self-Reported Questionnaire² | Base+Self-Reported Questionnaire+24 GPSs |
|---------------------------------|-------------|-----------------------------------|-----------------------------------------|
| Target outcome                  |             |                                   |                                         |
| Suicidal ideation               | Sensitivity | 0.15                              | 0.06                                    | 0.07                                    |
|                                 | Specificity | 0.09                              | 0.4                                    | 0.29                                    |
|                                 | PPV         | 0.14                              | 0.28                                   | 0.27                                    |
|                                 | NPV         | 0.11                              | 0.09                                   | 0.08                                    |
|                                 | Accuracy    | 0.880                             | 0.722                                  | 0.686                                   |
|                                 | ROC-AUC     | 0.566                             | 0.883                                  | 0.674                                   |
| Suicide attempt                 | Sensitivity | 0.002                             | 0.002                                  | 0.001                                   |
|                                 | Specificity | 0.03                              | 0.13                                   | 0.21                                    |
|                                 | PPV         | 0.01                              | 0.05                                   | 0.06                                    |
|                                 | NPV         | 0.007                             | 0.006                                  | 0.006                                   |
|                                 | Accuracy    | 0.962                             | 0.796                                  | 0.822                                   |
|                                 | ROC-AUC     | 0.629                             | 0.992                                  | 0.992                                   |

1. age, sex, years of parental education, parental marital status, study site, BMI
2. CBCL, child behavior checklist; ELS, early life stress; FES, family environment scale
*NPV, negative predictive value; PPV, positive predictive value; ROC-AUC, area under curve of receiver operating characteristic; Sensitivity, true positive rate; Specificity, true negative rate;
**Supplementary Table 2.** Performance of suicidality classification models in the balanced test set (in suicidal ideation, 238 cases and 238 controls; in suicide attempt, 16 cases and 16 controls)

| Model | Base Model¹ | Base+Self-Reported Questionnaire² | Base+Self-Reported Questionnaire+24 GPSs |
|-------|-------------|----------------------------------|------------------------------------------|
|       | Target outcome | Suicidal ideation | Suicide attempt | Suicidal ideation | Suicide attempt | Suicidal ideation | Suicide attempt |
| Sensitivity | 0.96 | 0.53 | 0.837 | 0.583 |
| Specificity | 1 | 0.54 | 0.058 | 0.25 |
| PPV | 0.5 | 0.66 | 0.538 | 0.7 |
| NPV | 0.15 | 0.32 | 0.214 | 0.167 |
| Accuracy | 0.565 | 0.679 | 0.756 | 0.75 |
| AUC | 0.571 | 0.695 | 0.72 | 0.766 |

1. age, sex, years of parental education, parental marital status, study site,
2. CBCL, child behavior check list; ELS, early life stress; FES, family environment scale

*NPV, negative predictive value; PPV, positive predictive value; ROC-AUC, area under curve of receiver operating characteristic; Sensitivity, true positive rate; Specificity, true negative rate;
**Supplementary Table 3.** Significant association of GPS with Childhood Suicidality (FDR significance < 0.05) in the ABCD children after excluding any control individuals with one or more KSAD records.

(a) **All sexes (n=3,244)**

| Outcome          | GPS Predictor | OR   | 95% OR (bottom) | 95% OR (top) | P-value (raw) | N of cases |
|------------------|---------------|------|-----------------|--------------|---------------|------------|
| Suicidal ideation | ADHD          | 1.15 | 1.06            | 1.24         | 4.42E-04      | 1149       |
| Suicidal ideation Active | ASD          | 1.17 | 1.07            | 1.29         | 5.57E-04      | 608        |
| Suicidal ideation | ASD          | 1.14 | 1.05            | 1.22         | 8.37E-04      | 1149       |
| Suicidal ideation Passive | ADHD       | 1.14 | 1.05            | 1.24         | 1.26E-03      | 942        |
| Suicidal ideation Active | ADHD       | 1.16 | 1.05            | 1.27         | 2.45E-03      | 608        |
| Suicidal ideation Passive | SMOKER     | 1.13 | 1.04            | 1.22         | 2.56E-03      | 942        |

(b) **Female only (n=1,562)**

| Outcome          | GPS Predictor | OR   | 95% OR (bottom) | 95% OR (top) | P-value (raw) | N of cases |
|------------------|---------------|------|-----------------|--------------|---------------|------------|
| Suicidal ideation Active | ADHD       | 1.31 | 1.13            | 1.52         | 4.49E-04      | 220        |
Supplementary Figure 1. Correlogram of the GPSs of 24 psychiatric and common traits.
Supplementary Figure 2. Biplot from Principal Component Analysis on the combined genotype data of 1000Genome reference panel\(^1\) and the ABCD study participants (colored blue). Super-populations are defined broadly as: AFR African, AMR admixed American, EAS East Asian, EUR European, SAS South Asian. To facilitate the understanding of genetic ancestry of the study participants, PCA was performed with 1000Genome phase3 reference panel and plotted with the first two principal components. Since ABCD participants reside on a continuum of genetic ancestry as American population (AMR), rather than distinct population groups, and none of them belongs to either African (AFR) or East Asian (EAS) populations, we did not exclude any genetic outliers based on the plot.
Supplementary References

1. “A global reference for human genetic variation” Nature 526 68-74 2015